Università degli Studi di Milano - Bicocca

FACOLTÀ DI SCIENZE STATISTICHE
Corso di Dottorato di Ricerca in Statistica XXV ciclo

Statistical Analysis of Brain Networks

Candidata:
Sara Sala

Matricola 734382

Tutor interno:
Prof. Piero Quatto

Tutor esterni:
Prof. Massimo Filippi
Prof. Theodore Kypraios

Anno Accademico 2011 – 2012

Acknowledgements

I would like to express my gratitude to Professor Massimo Filippi and his collaborators for giving me the opportunity to attend his Institute and to use his interesting data for my applications.
I would like to thank Professor Theodore Kypraios for giving me the possibility to attend his department in Nottingham. Our discussions have been of great help for new ideas.
I would like to show my gratitude to Professor Piero Quatto for his continuous encouragement and invaluable suggestions during this work. Without his help, this work would not have been completed.
My gratitude also goes to my colleagues, Isabella, Viviana, for sharing doubts, problems during my doctoral studies.
I would like to show my gratitude to Maddalena, Romina and Stefano because they supported me mentally during the course of this work.
My special appreciation goes to my parents and my brothers!

Contents

List of Figures

List of Tables

Introduction

Recent developments in the complex networks analysis, based largely on graph theory, have been used to study the brain network organization. The growth of network science has been driven by the fact that the behavior of brain systems is shaped by interactions among their constituent elements (Bullmore and Sporns, 2009).

The brain is a complex system that can be represented by a graph. A graph is a mathematical representation constituted by a set of nodes and a set of links (Kolaczyk, 2009),which can be useful to study the connectivity of the brain. Nodes in the brain can be identified dividing its volume in regions of interest and links can be identified calculating a measure of dependence between pairs of regions whose activation signal, measured by functional magnetic resonance imaging (fMRI) techniques, represents the strength of the connection between regions. The study of brain connectivity through graph theory and network analysis can be also applied to other brain mapping techniques, for instance diffusion magnetic resonance imaging (MRI), which investigate structural/anatomical brain connectivity.

A graph can be synthesized by the so-called adjacency matrix, which, in its simplest form, is an undirected, binary, and symmetric matrix, whose entries are set to one if a link exists between a pair of brain areas and zero otherwise. The adjacency matrix is particularly useful because allows the calculation of several measures which summarize global and local characteristics of functional brain connectivity, such as centrality, efficiency, density and *small worldness* property.

In Chapter 1 of this work, the basic terminology and notation for graphs are discussed and a network analysis is applied to a particular case study in

order to investigate the characteristics of brain networks. We consider the global measures, such as the clustering coefficient, the characteristic path length and the global efficiency, and the local measures, such as centrality measures and local efficiency, in order to represent global and local dynamics and changes between networks. This is achieved by studying with resting state (rs) fMRI data of healthy subjects and patients with neurodegenerative diseases.

The goal of Chapter 2 is to illustrate an original methodology to construct the adjacency matrix. Its entries, containing the information about the existence of links, are identified by testing the correlation between the time series that characterized the dynamic behavior of the nodes. This involves the problem of multiple comparisons in order to control the error rates. The method based on the estimation of positive false discovery rate ($pFDR$), proposed by Storey (2002) has been used. A similar measure involving false negatives (type II errors), called the positive false nondiscovery rate ($pFNR$) is then considered, proposing new point and interval estimators for $pFNR$ and a method for balancing the two types of error. This approach is demonstrated using both simulations and fMRI data, and providing finite sample as well as large sample results for $pFDR$ and $pFNR$ estimators. Furthermore a ranking of the most central nodes in the networks is proposed using q-values, the $pFDR$ analog of the p-values. The differences on the inter-regional connectivity between cases and controls are studied. Finally original methods for constructing group differences networks and group-based representative networks are proposed.

In Chapter 3 network models are discussed. Network models based on a *small-world* property have been mostly utilized to describe different brain networks. In order to gain deeper insights into the complex neurobiological interaction, exponential random graph models (ERGMs) are applied to assess several network properties simultaneously and to compare case/control brain networks. As an interesting proposal for future works, ERGMs can be applied in medical prediction problems, including diagnostic outcomes.

Chapter 1

Brain Network Analysis

1.1 The brain connectivity

Brain mapping techniques, such as electroencephalography (EEG), magnetoencephalography (MEG), functional magnetic resonance imaging (fMRI) and diffusion tensor MRI, produce large datasets of functional or structural connection patterns (Rubinov and Sporns, 2010).
Structural brain networks can be described as a graph composed by nodes (denoting neural elements, i.e. neurons or brain regions) that are connected by links, representing physical connections (synapses or axonal projections) (Bullmore and Sporns, 2009).
In addition to the anatomical connectivity, it is important to consider also the functional connectivity within the brain. The investigation of the inter-regional functional connections might allow to understand how this architecture supports neurophysiological dynamics. In their simplest form, functional brain networks can be represented by undirected graphs, in which nodes denote brain areas, and links represent correlations between the time series of functional activity of brain regions.
Measuring functional connectivity by means of a correlation coefficient is simple, but it does not indicate the causality of the connection, or whether the functional connection between two nodes is direct. A further step is therefore the representation of brain functional connections by means of effective

connectivity (e.g., by using dynamic causal models or granger causality). Measures of effective connectivity between regions can be used to generate a directed graph, and the estimation of the directionality of functional connections could be useful to show how information flows through the network, and to correctly understand the connection's dynamics (Smith *et al.*, 2011). The nature of nodes and links in individual brain networks is determined by combinations of brain mapping methods, anatomical parcellation schemes and measures of connectivity. There are many ways to define network nodes. In the case of electrophysiological data, the simplest approach is to consider each recorded channel as a node. In the case of fMRI, nodes are often defined as spatial regions of interest (ROIs) obtained from brain atlases (Tzourio-Mazoyer *et al.*, 2002). The choice of a parcellation scheme determines the neurobiological interpretation of network topology. Nodes should ideally represent homogeneous brain regions with patterns of anatomical or functional connections. Parcellation schemes should divide heterogeneously connected brain regions into single nodes completely covering the surface of the cortex, or of the entire brain, without spatially overlap. Networks constructed using distinct parcellation schemes may significantly differ in their properties and cannot, in general, be quantitatively compared. Specifically, structural and functional networks may only be compared if these networks share the same parcellation (Rubinov and Sporns, 2010).

In the process of construction of functional brain networks, once the nodes are defined, the next step is the estimate of a measure of association between nodes. This is usually done by using suitable time-courses associated with nodes. The more similar the time-courses are between any given pair of nodes, the more likely it is that there is a functional connection between those nodes. Many different methods are being used in the literature as measure of dependence between nodes (Smith *et al.*, 2011). As previously mentioned, correlation-based approaches can be quite successful. However, correlation between two time-series does not imply causality (by definition, it tells anything about the direction of information flow) and directionality (there may be a third node between the two under consideration without a direct connection existing between the two). Therefore, the use of correlation

coefficients often doesn't allow the distinction between true correlations and spurious correlations, and doesn't allow to correctly interpret the nature of links. One simple method to overcome this limitation is the estimation of functional connections with partial correlation coefficients. For example, in a three-node network, this method works by taking each pair of time-series in turn, and regressing out the third from each of the two time-series in question, before estimating the correlation between the two (Smith *et al.*, 2011). However, the estimation of partial correlation coefficient might be problematic and unstable for networks with several nodes (e.g., 90 nodes, as in the case of a typical brain parcellation of a fMRI experiment). Therefore, the majority of literature studies use methods (such as full correlation) giving no directional information at all (Bullmore and Sporns, 2009).
In addition to the type of connectivity (anatomical, functional or effective) links can be also differentiated on the basis of their weight (Rubinov and Sporns, 2010). Binary links denote the presence or absence of connections, while weighted links also contain information about connection strength. Weights in anatomical networks may represent size, density, or coherence of anatomical tracts, while weights in functional and effective networks may represent respective magnitudes of interactions (Latora and Marchiori, 2003). Binary networks are in most cases simpler to characterize and have a more easily defined null model for statistical comparison. On the other hand, weighted characterization usually focuses on different and complementary aspects of network organization and may be useful in filtering the influence of weak and potentially non-significant links. These links tend to obscure the topology of strong and significant connections and as a result are often discarded, by applying an absolute, or a proportional weight threshold. Threshold values are often arbitrarily determined, and networks should ideally be characterized across a broad range of thresholds. With the use of a threshold, however, all negative connections are removed from the networks, causing a loss of information about the role of negative weights in global network organization.

1.2 Basic Definitions and Concepts: the Graph Theory

As we have seen in the previous section, the brain is as a complex system that can be modeled as a network, where the vertices (or nodes) are brain regions and the edges (or links) represent the interactions between them. An accurate description of the architecture and a characterization of the properties of the network can be important to understand the dynamic of the system (Latora and Marchiori, 2003). Network analysis (also called graph theory) describes complex systems by quantifying properties of network representation.

In this section we discuss the basic terminology and notation for graphs.

Networks can be represented as a graph $G = (V, E)$, a mathematical structure consisting of a set V of vertices and a set E of edges, where elements of E are unordered pairs $\{i, j\}$ of distinct vertices $i, j \in V$. Adjacency is a basic notion when describing connectivity. Two vertices i, j are said to be neighbors or adjacent if joined by a common end point in V (Kolaczyk, 2009). The cardinality of V is usually denoted by n, the cardinality of E by m. The two vertices joined by an edge (e) are called its end-vertices.

Graphs can be undirected or directed. In directed graphs, each directed edge (arc) has an origin (tail) and a destination (head). An edge with origin $i \in V$ and destination $j \in V$ is represented by an ordered pair (i, j). In both undirected and directed graphs, we may allow the edge set E to contain the same edge several times, i.e., E can be a multiset. If an edge occurs several times in E, the copies of that edge are called parallel edges. Graphs with parallel edges are also called multigraphs.

A graph is called simple, if each of its edges is contained in E only once, i.e., if the graph does not have parallel edges.

An edge joining a vertex to itself, i.e., an edge whose end-vertices are identical, is called a loop. A graph is called loop-free if it has no loops.

Often it is useful to associate numerical values (weights) with the edges of a graph $G = (V, E)$. Edge weights can be represented as a function $\omega : E \to \mathbb{R}$ that assigns each edge a weight $\omega(e)$. Depending on the context, edge weights

can describe various properties such as cost, capacity, strength of interaction, or similarity. An unweighted graph $G = (V, E)$ is equivalent to a weighted graph with unit edge weights $\omega(e) = 1 \forall e \in E$.

The degree of a vertex i in an undirected graph $G = (V, E)$, denoted by $Deg(i)$, is the number of edges in E that have i as an end-vertex. If G is a multigraph, parallel edges are counted according to their multiplicity in E. The set of edges that have i as an end-vertex is denoted by $\Gamma(i)$. The set of neighbors of i is denoted by N_i. In a directed graph $G = (V, E)$, the out-degree of $i \in V$, denoted by $Deg^{out}(i)$, is the number of edges in E that have origin i. The in-degree of $i \in V$, denoted by $Deg^{in}(i)$, is the number of edges with destination i. For weighted graphs, all these notions are generalized by summing over edge weights rather than taking their number. The maximum and minimum degree of an undirected graph are denoted by $\Delta(G)$ and $\delta(G)$, respectively. The average degree is denoted by $Deg(G) = \frac{1}{|V|} \sum_{i \in V} Deg(i)$.

A graph $G' = (V', E')$ is a subgraph of the graph $G = (V, E)$ if $V' \subseteq V$ and $E' \subseteq E$. It is a (vertex-) induced subgraph if E' contains all edges $e \in E$ that join vertices in V'. The induced subgraph of G with vertex set $V' \subseteq V$ is denoted by $G[V']$. The (edge-)induced subgraph with edge set $E' \subseteq E$, denoted by $G[E']$, is the subgraph $G' = (V', E')$ of G, where V' is the set of all vertices in V that are end-vertices of at least one edge in E'. Particular cases of subgraphs are called dyads and triads: a dyad represents a pair of vertices and the possible edges between them, a triad consists of three vertices and possible edges between them (Wasserman, 1994).

If C is a proper subset of V, then $G - C$ denotes the graph obtained from G by deleting all vertices in C and their incident edges. If F is a subset of E, $G - F$ denotes the graph obtained from G by deleting all edges in F.

A walk from x_0 to x_k in a graph is an alternating sequence:

$$x_0, e_1, x_1, e_2, \ldots, x_{(k-1)}, e_k, x_k$$

of vertices and edges, where $e_i = x_{(i-1)}, x_i$ in the undirected case and $e_i = (x_{(i-1)}, x_i)$ in the directed case. The length of the walk is defined as the number of edges on the walk. The walk is called a path, if $e_i = e_j$ for $i = j$,

and a path is a simple path if $x_i = x_j$ for $i = j$. A path with $x_0 = x_k$ is a cycle. A cycle is a simple cycle if $x_i = x_j$ for $0 \leq i < j \leq k - 1$. A geodesic between vertex i and j is a shortest path between these vertices (Jackson, 2008). An important thing to keep in mind in many applications of networks is which nodes can reach which other nodes through paths in the network. This plays a critical role in things like contagion, learning and the diffusion of the information through a network. In particular, in the brain network that could be important in the study of the connectivity of the brain regions. Looking at the path relationships naturally partitions a network into different connected subgraphs that are commonly referred to as components (Rubinov and Sporns, 2010). An undirected graph $G = (V, E)$ is connected if every vertex can be reached from every other vertex, i.e., if there is a path from every vertex to every other vertex. A graph consisting of a single vertex is also taken to be connected. Graphs that are not connected are called disconnected. For a given undirected graph, a connected component of G is an induced subgraph $G' = (V', E')$ that is connected and maximal.

A directed graph $G = (V, E)$ is strongly connected if there is a directed path from every vertex to every other vertex. A strongly connected component of a directed graph G is an induced subgraph that is strongly connected and maximal.

A graph that is connected and is acyclic is called a tree. Trees contain the minimum number of edges necessary to be connected and they do not contain cycles, they are particular important for several characteristics. First, trees are minimally connected graphs since every line in the graph is a bridge; the removal of any one edge causes the graph to be disconnected. Second, the number of edges in a tree equals the number of vertices minus one. Adding another edge the graph becomes a cycle, so is no longer a tree. Finally, there is only one path between any two vertices in a tree. A graph that is disconnected and contains no cycles is called a forest, where each component is a tree (Wasserman, 1994).

There are equivalent ways of representing a graph. Instead of describing a graph by listing of all the links or edges that are in the graph, it is sometimes easier to view G as an $n \times n$ matrix. Indeed the fundamental connectivity

of a graph may be captured in a $n \times n$ binary symmetric (for the undirected case) matrix A with entries

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

where the integers $1, \ldots, N$ generically denote the elements of V. In other words, A is non-zero for entries whose row-column indices correspond to vertices joined by an edge, and zero for those that are not. This matrix, called the adjacency matrix, is also useful for operations on A that yield additional information concerning G. For example, the row sum $A_{i.} = \sum_j A_{ij}$ is equal to degree $Deg(i)$ of vertex i. By symmetry $A_{i.} = A_{.i}$.
Furthermore:

Theorem.
If A is the $n \times n$ adjacency matrix of a graph G, then $(A^L)_{ij}$ is the number of walks of length L from the node i to the node j in G.

Proof. A simple proof is by induction on L. For $L = 1$ the result follows from the definition of adjacency matrix, since a walk of length 1 is just an edge. The induction hypothesis assumes that $(A^L)_{ij}$ is the number of walks of length L from i to j in G. Hence, there are $(A^L)_{ih}A_{hj}$ walks of length $L+1$ from i to j in which h is the penultimate node. The result then follows directly from $(A^{L+1})_{ij} = \sum_h (A^L)_{ih}A_{hj}$.

□

In addition, there are many interesting relations involving the eigenvalues of A. For example, G is regular if and only if the maximum degree $\Delta(G)$ is an eigenvalues of A.
An adjacency matrix may also be created for digraphs (directed graphs)

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

where E is a direct edge from i to j. A is no longer symmetric in this case, but it contains similarly useful information. For example $A_{i.} = Deg^{out}(i)$ and $A_{.j} = Deg^{in}(j)$.
Another useful matrix capturing the structure in G is the incidence matrix B, an $n \times m$ binary matrix with entries

$$B_{ij} = \begin{cases} 1, & \text{if vertex } i \text{ is incident to edge } j \\ 0, & \text{otherwise.} \end{cases}$$

There is a relation between A and B, indeed if we extend B to be a signed incidence matrix $\bar{B}$ where the entries 1 are given plus or minus signs indicating an arbitrarily assigned orientation of the corresponding edge. That is the role of either head or tail is assigned to each vertex in an edge j (like in digraph), and let $\bar{B}_{ij} = 1$ if vertex i is incident to edge j as a tail, and -1, if as a head. It can be show that $\bar{B}\bar{B}' = D - A$ where $D = \text{diag}[(d_i)_{(i \in V)}]$ is diagonal matrix containing the degree sequence. The $n \times n$ matrix $L = D - A$ is called the Laplacian of the graph for the analogy to the Laplacian in the context of the multivariable calculus. Indeed, for a vector $\mathbf{x} \in \mathbb{R}^n$

$$\mathbf{x}' L\mathbf{x} = \sum_{(i,j) \in E} (x_i - x_j)^2.$$

The closer this value is to zero, the more similar are the elements of $\mathbf{x}$ at adjacent vertices in V. Hence the Laplacian provides some sense of the smoothness of functions on a graph G, with respect to the connectivity of G. Furthermore the eigenvalues and eigenvectors are important to know the structure of the graph. Since L can be show to be a positive semi-definite matrix, the eigenvalues are all non-negative and because $L\mathbf{1} = \mathbf{0}$, where $\mathbf{1}$ and $\mathbf{0}$ are n vectors of ones and zeroes, the smallest eigenvalue λ_1 of L is equal to zero. The second smallest eigenvalue, λ_2, is typically nontrivial and the most important of all the eigenvalues. For example, the larger λ_2 is, the more connected G is, and the more difficult is to separate G into disconnected subgraphs by eliminating some number of edges (Kolaczyk, 2009).

1.3 Local and Global Network Measures

In the study of brain connectivity, it is interesting to investigate the structure and the characteristics of the corresponding brain network. For example, neuronal dynamics can be represented by triplets of nodes with a particular pattern of lings among them (i.e., triads); questions about the movement of information throughout brain regions can be posed in terms of paths on the network, and flows along those paths. The importance of brain regions in the network may be captured by measures of how central the corresponding node is. The search for clusters and analogous types of unspecified groups within the brain may be addressed as a graph partitioning problem (Kolaczyk, 2009).

1.3.1 Measures of Centrality

An aspect to be investigated about a vertex is its importance in the network. Measures of centrality are designed to quantify such notion of importance. We will assume that G is an undirected graph.

As discussed above, the degree in a graph $G = (V, E)$ counts the number of edges in E incident upon the vertex i. Therefore, the degree provides a basic quantification of the extent to which i is connected to other vertices within the graph. The degree can be considered as the most common measures of centrality.

In the context of functional brain connectivity, a node can be considered central in the network if it is connected with the other nodes. The degree of a node i is the number of links ($e \in E$) connected to the node i, its value therefore reflects the importance of a node in the network, and it can be defined as:

$$Deg(i) = \sum_{j \in N} A_{ij},$$

where N is the total number of nodes in the network.

Another popular class of centrality measures takes into account that the importance of a vertex is not only related to the number of its connections, but also to its location with respect to the paths in the network. Vertices that sit on many paths are likely more critical to the communication process (Wasser-

man, 1994). Betweenness centrality measures are aimed at summarizing the extent to which a vertex is located between other pairs of vertices. The most commonly used definition of betweenness centrality has been introduced by Freeman (1977):

$$Bet(i) = \sum_{j \neq h \neq i \epsilon V} \frac{D_{jh}(i)}{D_{jh}}$$

where $D_{jh}(i)$ is the total number of shortest paths between j and h that pass through i, and $D_{jh}$is the number of shortest paths between j and h. In the case that shortest paths are unique, $Bet(i)$ just counts the number of shortest paths going through i. This centrality measure can be normalized to the unit interval through division by a factor of $(n-1)(n-2)/2$. Calculation of the collection of all betweenness centralities $Bet(i)$ requires the calculation of the lengths of shortest paths among all pairs of vertices, by means of the Theorem in section 1.3, and the computation of the equation above for each vertex.

1.3.2 Measures of Efficiency

Let N be the total number of nodes in the network. Global efficiency of the network is defined as:

$$Eff(G) = \frac{1}{N} \sum_{i} \frac{\sum_{i \neq j} D_{ij}^{-1}}{N-1}$$

The local efficiency is

$$Eff(i) = \frac{1}{N} \sum_{i} \frac{\sum_{i \neq j,h} A_{ij} A_{ih} [D_{jh}^{-1}]}{Deg(i)[Deg(i)-1]}$$

where D_{jh} is the length of the shortest path between j and h (Latora and Marchiori, 2001).

1.3.3 Measures of Density and Cohesion

The description of a network structure as coherent is based on the notion that a coherent subset of nodes should be locally dense in the graph (Ko-

laczyk, 2009).

Given a network graph G, define f_d to be the fraction of vertices $i \in V$ with degree $Deg(i) = d$. The collection $\{f_d\}_{(d \geq 0)}$ is called the degree distribution of G, and is simply the histogram formed from the degree sequence, with bins of size one, centered on the non-negative integers. The degree distribution provides a natural summary of the connectivity in the graph. The mean network degree $Deg(G) = \frac{1}{N} \sum_{i \in V} Deg(i)$ is most commonly used as a measure of density, or the total wiring cost of the network.

From a graph-theoretic point of view, another concept that can be employed to describe network density is that of 'clique': a complete subgraph H of G. Cliques are subsets of vertices that are fully cohesive, in the sense that all vertices within the subset are connected by edges. A common case of practical interest, particularly in social network analysis, is that of 3-cliques (i.e., triangles). In practice, large cliques are relatively rare, as they necessarily require that G itself be fairly dense. This approach to network cohesion proceed by first stating a prespecified notion of locally dense structure and then looking to see whether it is anywhere satisfied in a graph G. Alternatively, we can define a measure of local density and then characterize the extent to which subsets of vertices are dense, according to this measure. Such measures are commonly based on ratios of the number of edges among a subset of vertices to the total number of possible edges. For example, in a graph G with no self-loops and no multiple edges, the density of a subgraph $H = (V_H, E_H)$ is

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

The value of $\text{den}(H)$ will lie between zero and one and provides a measure of how close H is to being a clique. Note that $\text{den}(H)$ is just a rescaling of the average degree seen above. There is freedom in the choice of subgraph H. Taking $H = G$ yields the density of the overall graph G. Conversely, taking $H = N_i$ to be the set of neighbors of a vertex $i \in V$, and the edges between them, yields a measure of density in the immediate neighborhood of i. Watts and Strogatz (1998) propose $\text{den}(N_i)$ as a summary of the extent

to which there is clustering of edges local to i and propose that the average of $\text{den}(N_i)$ over all vertices can be used as a clustering coefficient for the overall graph. These measures of clustering can be expressed alternatively in terms of the density of triangles among connected triples. A triangle is a complete subgraph of order three. A connected triple is a subgraph of three vertices connected by two edges (also sometimes called a 2-star). Intuitively, a measure of the frequency with which connected triples close to form triangles will provide some indication of the extent to which edges are clustered in the graph. Let $\tau_\Delta(i)$ denote the number of triangles in G into which $i \in V$ falls, and $\tau_3(i)$ the number of connected triples in G for which the two edges are both incident to i. Note that $\tau_3(i) = \binom{Deg(i)}{2}$. The Watts-Strogatz local clustering coefficient $\text{den}(H_i))$ can be re-expressed as $cl(i) = \frac{\tau_\Delta(i)}{\tau_3(i)}$, for those vertices i with $\tau_3(i) > 0$. The corresponding clustering coefficient for G is

$$cl(G) = \frac{1}{|V'|} = \sum_{i \in V'} cl(i)$$

where $V' \subseteq V$ is the set of vertices i with $Deg(i) \geq 2$.

Clustering coefficients have become a standard quantity used in the analysis of network structure. Their values have typically been found to be quite large in real-world networks, in comparison to what expected in random graph models.

Measures based on path concept estimate the ease with which brain regions communicate. Paths are sequences of distinct nodes, the average shortest path length between all pairs of nodes in the network is known as path length of the network. Let the shortest path length (distance), between nodes i and j, be

$$D_{ij} = \sum_{A_{ij} \in g} A_{ij},$$

where g is a shortest path (geodesic) between node i and j, that is equivalent to

$$D_{ij} = \min\{d > 0 : (A^d)_{ij} > 0\},$$

by Theorem in section 1.3. The characteristic path length of the network

can be defined as

$$pl(G) = \frac{1}{N}\sum_i pl(i)$$

where $pl(i) = \frac{1}{N}\sum_i \frac{\sum_{i\neq j} D_{ij}}{N-1}$.

1.3.4 Small-worldness

In contrast to a local perspective, and the search for small-scale subsets of cohesive vertices, it may useful to also investigate networks at a global perspective. Basic questions of interest are whether a given graph separates into distinct sub-graphs, and how information flows within the graph. A graph G is said to be connected if every vertex is reachable from every other (i.e., if for any two vertices, there exists a walk between the two). A connected component of a graph is a maximally connected subgraph. Such a component is called a giant component. A characteristic of giant components of many real-world networks is the so-called small-world property. The initial concept and terminology for this term is due to the experiment of Milgram (1967) in the late 1960's, and his assertion that people are only separated by roughly six acquaintances. If the population of the world and acquaintances among its members is represented as a social network graph, despite its enormous size, the number of hops along shortest paths between its vertices would be quite small.
Formally the average distance between distinct vertices

$$\bar{l} = \frac{1}{N(N+1)/2}\sum_{i\neq j} D_{ij}$$

scales as $O(\log(N))$ or less.
Watts and Strogatz (1998) discover that this property is valid for many networks. These authors also found that often small average distance is accompanied by a high clustering coefficient, and they proposed a random graph model for such a network: the 'Small-world' network. The small-world property has important implications for problems of network communication. Small-world networks are formally defined as networks that are significantly

more clustered than random networks, yet have approximately the same characteristic path length as random networks[1] (Watts and Strogatz, 1998). Random networks have to preserve the same number of nodes, edges, and degree distribution as the investigated networks (Maslov and Sneppen, 2002). A useful way to investigate the small-world property of a real network is to compute the normalized clustering coefficient $cl_{norm} = \frac{cl_{real}}{cl_{rand}}$ and the normalized characteristic path length $pl_{norm} = \frac{pl_{real}}{pl_{rand}}$ (Watts and Strogatz, 1998), where cl_{real} and pl_{real} are cl and pl of a real brain networks, and cl_{rand} and pl_{rand} represent the corresponding measures calculated on a great number of matched random networks. By definition, a small-world network had to respect the following criteria: $cl_{norm} > 1$ and $pl_{norm} \approx 1$ (Watts and Strogatz, 1998).

1.4 Applications of Graph Theory to Neurophysiological and Neuroimaging data

In the previous sections we introduced the main concepts of the graph theory, describing how brain connectivity can be investigated using network analysis, and how measures derived from Theorem of section 1.3, are useful to describe the characteristics of the network itself.

Over the last decade network analysis has been applied to biological research fields such as immunology, genetics and neuroscience and these applications. The growing interest of neuroscientists on network analysis has been driven by the realization that the behaviour of complex systems is shaped by the interactions among their constituent elements (Bullmore and Sporns, 2009). The application of graph theory concepts to neurophysiological and neuroimaging data is a promising way to characterize brain connectivity (De Haan *et al.*, 2009). Network analysis is a tool able to represent the

[1]The significance of network measures should be established by comparison with the same measures calculated on random networks. Random networks have simple random or ordered topologies but preserve basic characteristics of the original network. The most commonly used random network has a random topology but shares the size, density and binary degree distribution of the original network (Maslov and Sneppen, 2002).

local and the global structural changes, as well as the dynamics of the brain. The complex organization of connectivity in the human brain has still to be completely investigated, nonetheless measures based on graph theory (Latora and Marchiori, 2001)(Watts and Strogatz, 1998) have provided a new way to understand the relation between brain structure and function (Bullmore and Sporns, 2009).
In this section, network analysis will be applied to study how brain functional connectivity changes with pathology. This will be achieved by studying with graph theory fMRI data of subjects coming from different populations, including healthy subjects and patients with neurodegenerative diseases. In this context, network analysis might contribute to detect abnormalities of network connectivity in different brain disorders and to improve our understanding of the pathobiological mechanisms responsible of the development of irreversible disability. In this case, the role of network analysis is to provide new measures to quantify differences between patient groups. In Figure 1.1, the example of two networks, one (a) referred to an healthy subject and the other (b) referred to a patient, is provided. The difference between the two networks can be noted even at a visual inspection. In the next sections, we will quantify this difference by using metrics of graph theory, and we will interpret their meaning in a biological context.

1.4.1 Subjects and data acquisition

Aim of the presented applicative study is to explore the difference of brain functional connectivity between healthy subjects and patients with a diagnosis of behavioural variant of frontotemporal dementia (bvFTD). The bvFTD is a clinical syndrome caused by the degeneration of the frontal and temporal lobes of the brain and it is the second-most common dementia after Alzheimer's disease. bvFTD affects social conduct, social inhibitions, and personality, with patients becoming lethargic and apathetic (or, oppositely, dishinibited) and unable to perform skills that require complex planning or sequencing. The analysis of functional connectivity in these patients is useful to define if there is a correspondence between clinical symptoms and func-

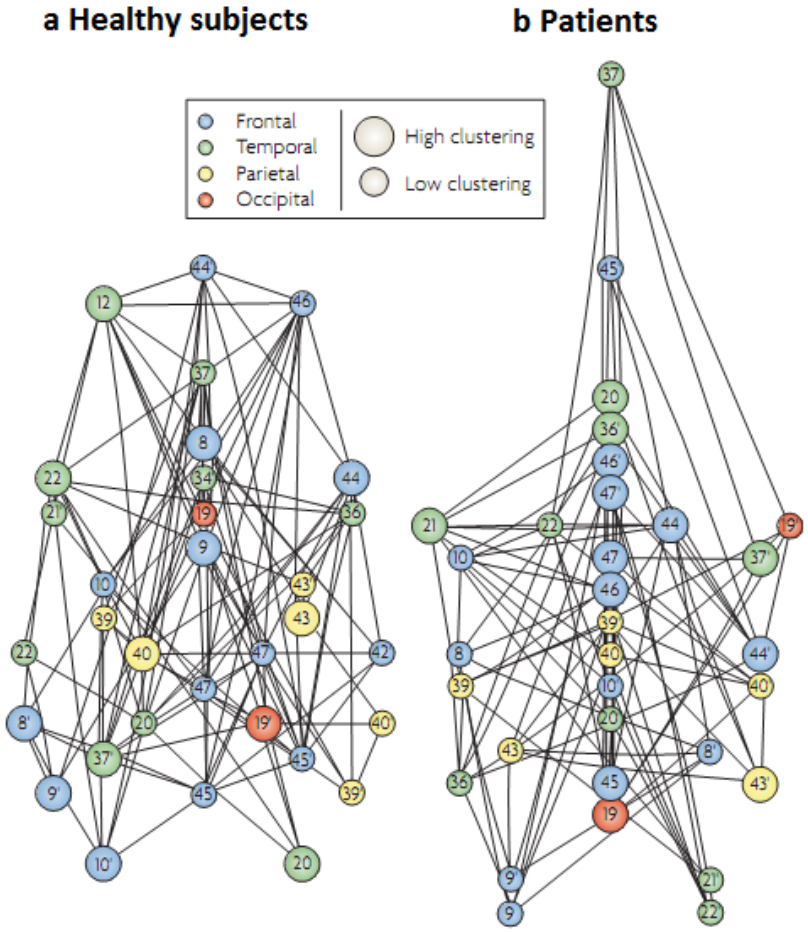

Figure 1.1: Disease-related disorganization of brain networks

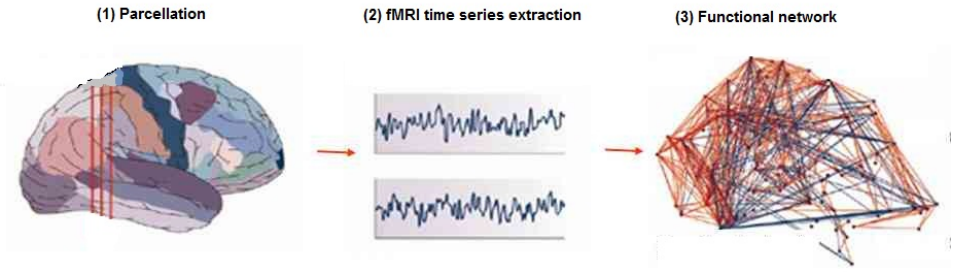

Figure 1.2: Construction of the functional brain network

tional disconnection of regions of the frontal and temporal lobe. To this aim, we measure brain network connectivity with fMRI at resting state, i.e., with subjects lying still in the scanner and thinking nothing in particular.
In this study, we recruite eighteen right-handed patients with a diagnosis of bvFTD (mean age=61.9 years; range=45-79 years, 10/8 males/females) and fifty right-handed age and sex-matched healthy controls (mean age=61.10 years, range=46-79 years; 23/27 males/females). Resting state fMRI images are pre-processed to correct for minor head movements and normalized to a default template using a standard affine transformation. Then, data are band-pass filtered to partially remove low-frequency drifts and physiological high-frequency noise. No spatial smoothing is applied, in order to avoid spurious correlations between neighboring voxels, as previously suggested by Sanz-Arigita *et al.* (2010).

1.4.2 Construction of brain functional networks

In a graphical representation of a brain network, a node corresponds to a brain region while a link corresponds to the functional interaction between two brain regions (Rubinov and Sporns, 2010). To construct brain functional networks we follow the steps illustrated in Figure 1.2.
First we employ an automated anatomical labeling (AAL) atlas (Tzourio-Mazoyer *et al.*, 2002) to parcellate the brain into 90 cortical regions of interest (ROIs) (1). Second time series are extracted from each ROI by averaging the

signal of all voxels within that region (2). Bivariate correlations between each ROIs pair are obtained by calculating the Pearson's correlation coefficient ρ_{ij} between ROIs time courses i and j. Such correlation coefficients represent the functional connectivity strength between brain regions. Correlation matrices [2] of size $\{90 \times 90\}$, obtained from all study subjects, are thresholded deriving an adjacency matrix A with entries:

$$A_{ij} = \begin{cases} 1, & \text{if } \rho_{ij} \geq \tau \\ 0, & \text{otherwise} \end{cases}$$

A is computed putting the A_{ij} elements to zero if $\rho_{ij} < \tau$ and unity if $\rho_{ij} \geq \tau$, where $0 < \tau < 1$ is a suitable threshold. Finally, through this thresholding, unweighted graphs are obtained (3), with the nodes representing brain regions and edges/links representing functional relationships between brain regions (Rubinov and Sporns, 2010).

A crucial issue is the choice of the threshold used to generate the adjacency matrix from the correlation matrix: different thresholds generate graphs of different average degree, this means that, at a given value of τ, a single-subject graph can have more or less significant links than graphs of other subjects. Previous studies faced this problem by avoiding to use a fixed τ for all study subjects, but they rather fixed the average degree $Deg(G)$ instead to fix τ, i.e. they forced the total number of existing connections to be the same for all study subjects (Tian *et al.*, 2011), (Yao *et al.*, 2010), (Van Den Heuvel and Hulshoff Pol, 2010), (Stam *et al.*, 2009). However, this approach is not always straightforward, since fixing a pre-defined $Deg(G)$ may involve the risk of modifying network structure (Van Wijk *et al.*, 2010). Therefore, for this application, we choose to construct our graphs by fixing the same τ in all study subjects (Van Den Heuvel *et al.*, 2009), (Salvador *et al.*, 2005), (Meunier *et al.*, 2009). In Chapter 3 we will show another case study for which it is advisable to fix the average degree.

Because there is no unique way to choose τ, here, we examine several possible

[2] Note that the elements of the principal diagonal of the correlation matrix are not considered (loop-free graph)(Kolaczyk, 2009).

network configurations for a range of values from 0 to 0.9, and we explore the consistency of the results over this range (Sanz-Arigita *et al.*, 2010). In the next chapter, we will offer an alternative approach to the investigation of network properties over a range of thresholds, and the problem of the choice of τ will be dealt by multiple testing approach, in order to offer a statistical evaluation of the error committed in constructing the adjacency matrix.
In this first applicative section, we establish to explore network characteristics over the range of threshold that yielded fully connected graphs ($0 < \tau < 0.20$, with increments of 0.01). The condition of fully-connectedness is necessary to calculate all network measures defined in the previous sections of this chapter, because the path length $pl(G)$ of a sparse graph is infinite.

1.4.3 Descriptive network analysis

Global and local network characteristics are explored over the above defined range of plausible thresholds by assessing the principal measures of graph theory, which are able to detect various aspects of functional connectivity, e.g. quantify importance of brain regions, characterize patterns of local anatomical circuitry, test resilience of networks to insult and so on. Values of network measures from all elements (i.e., nodes and links) of a graph form a distribution, which provides a more global description of the network. This distribution is most commonly characterized by its mean, although other features, such as the median or the distribution shape, may be more important if the distribution is non-homogeneous.

Global network analysis

In order to characterize the global organization of brain networks, the average degree, the global efficiency, the clustering coefficient and the characteristic path length are adopted. As defined above, the degree $Deg(i)$ of a single node i is the number of links connected to that node. The average network degree $Deg(G)$ is defined as the average $Deg(i)$ of all network nodes, and is used as a measure of density of the network (Rubinov and Sporns, 2010). The clustering coefficient $cl(i)$ of a node i is the number of links existing

between its nearest neighbors; in other words, it represents the fraction of the node's neighbors that are also neighbors of each other (Watts and Strogatz, 1998). The mean network clustering coefficient $cl(G)$ over a network, therefore, represents how strong it is locally interconnected. The characteristic path length $pl(G)$ is defined as the average shortest path length between all pairs of nodes in the network (Watts and Strogatz, 1998). Since $pl(G)$ represents the average number of nodes that must be traversed to go from one node to any other node, it is usually taken as a measure of functional integration. The global efficiency $Eff(G)$ is defined as the average inverse shortest path length (Latora and Marchiori, 2001).

Global network architecture is then quantified in terms of small-worldness, as defined in Section 1.3.4 (Watts and Strogatz, 1998). To this aim, we computed the normalized clustering coefficient cl_{norm} and the normalized characteristic path length pl_{norm}. A Kolmogorov-Smirnov test is performed on all global network parameters, in order to verify that data are normally distributed. In our case all global network metrics follow a Gaussian distribution. To test whether small-worldness property is present or absent in the networks of healthy controls and bvFTD patients, separately, we use a one-sample T test, the aim of this analysis is to verify if cl_{norm} is significantly higher than the asymptotic value of 1 and, contemporary, pl_{norm} is not significantly higher than the asymptotic value of 1, which are the conditions of a graph with small-worldness property Supekar *et al.* (2008).

In order to explore the difference of brain functional connectivity between healthy subjects and bvFTD patients, global network parameters are compared between the two groups, first a F test is used to verify that the two samples have equal variances, then the difference in means of the network parameters is tested using a two-sample T test for non-paired data. (Kolaczyk, 2009). Plots of the mean values of global network parameters as a function of the correlation threshold, τ, for controls and bvFTD patients are reported in Figure 1.3. Most of graph theoretical metrics are found to be significantly altered in bvFTD patients with respect to healthy controls. In particular, $Deg(G)$, $pl(G)$ and $Eff(g)$ were significantly different between bvFTD patients and controls when $0 < \tau < 0.10$ (p-value ranging from 0.04 to

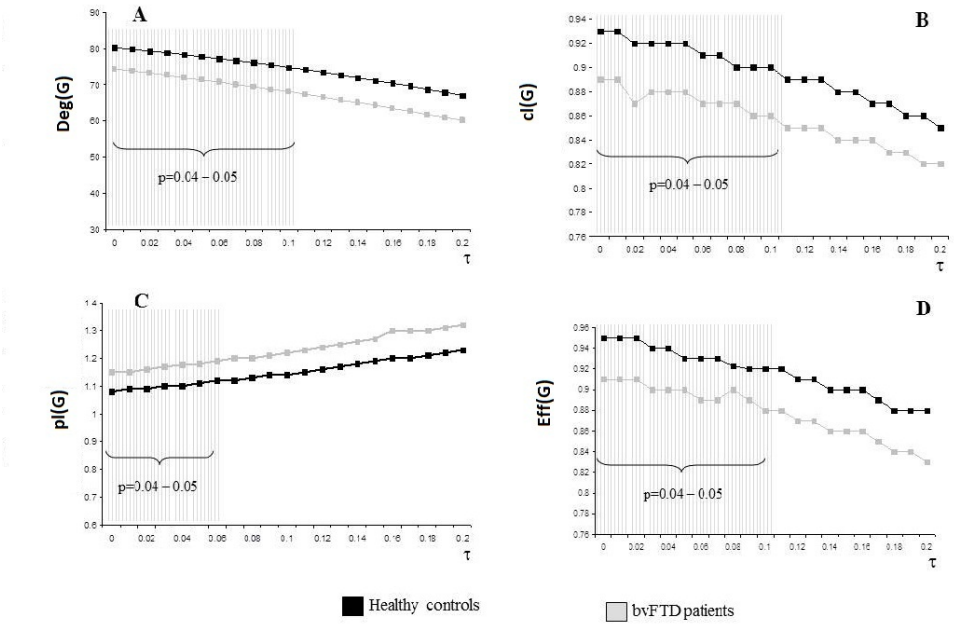

Figure 1.3: Global measures results: degree in plot A, clustering coefficient in plot B, path length in plot C, global efficiency in plot D

0.05), and $cl(G)$ was significantly lower in bvFTD patients than in healthy controls when $0 < \tau < 0.06$ (p-value ranging from 0.04 to 0.05). Even when statistical significance is not reached, there is a trend very close to significance (p-value ranging from 0.056 to 0.08).
In both controls and bvFTD patients, small-worldness properties are verified: cl_{norm} is significantly higher than 1 at all considered thresholds in both groups (p-value ranging from 0.001 to 0.01 in healthy controls, and from 0.008 to 0.001 in bvFTD patients). Conversely, pl_{norm} is not significantly different from 1 for any threshold in both groups (p-value ranging from 0.06 to 0.5).

Local network analysis

Local network measures are calculated in order to identify which brain regions could be considered 'hubs', i.e., regions interacting with several other brain areas, thus facilitating functional integration. Hubs are identified on the basis of their centrality in the network. To this aim, we calculate two metrics of centrality: the degree and the betweenness centrality. The degree $Deg(i)$, as defined above, is one of the most common measures of centrality: nodes with high $Deg(i)$ are functionally interacting with many other network nodes. The betweenness centrality $Bet(i)$ of a node is defined as the fraction of all shortest paths in the network that pass through that node (Rubinov and Sporns, 2010). Therefore, bridging nodes that connect disparate parts of the network have an high $Bet(i)$. These nodal measures (i.e., $Deg(i)$ and $Bet(i)$) were calculated for every τ in the range yielding fully connected graphs (i.e., between 0 and 0.20, as previously mentioned). Then, a nodal parameters X_{nod} integrated over all considered thresholds, is calculated as

$$X_{nod}(i) = \sum_{k=0.01}^{0.20} X(i, k\Delta\tau)\Delta\tau$$

where X is one of the local measures of interest and $\Delta\tau = 0.01$. A brain region i is defined as an 'hub' when $X_{nod}(i)$ of the nodal metrics is at least one standard deviation (SD) greater than the average of the parameters over the network (Tian *et al.*, 2011). In the Figure 1.4 the 'hubs' for the

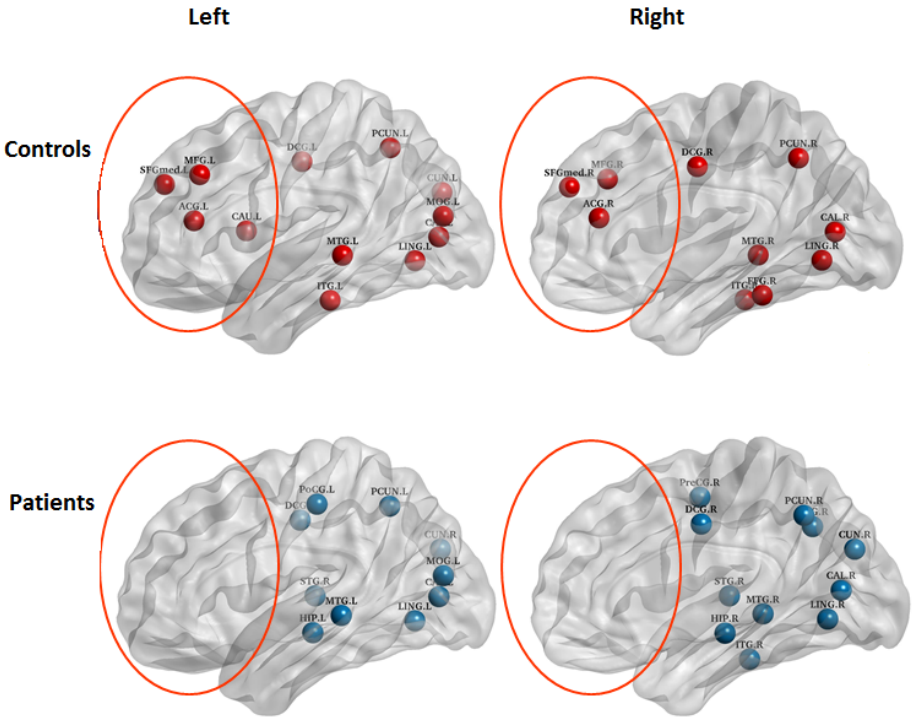

Figure 1.4: Cortical hubs in left and right hemisphere relative to controls and patients networks

right and left hemisphere are shown for healthy controls (red) and bvFTD patients (blue). As it can be easily noted, bvFTD patients lose their hubs in the frontal region, confirming with a non-invasive technique that functional disconnection of frontal regions is underlying bvFTD clinical symptoms.

1.4.4 Discussion of the results

This applicative study confirms that complex network analysis can be used to explore connectivity relationships in individual subjects and to explore connectivity differences between subject groups. By analyzing fMRI fluctuations during resting state condition, we have demonstrated that bvFTD condition induced changes in global and local brain functional connectivity in line with the expected neuropathological alterations. Global functional

network organization was altered in bvFTD patients, suggesting a loss of efficiency in information exchange between brain areas. Moreover, the local network analysis reveals that the altered brain regions are mainly located in the frontal lobe (including the anterior cingulate cortex), temporal lobe and basal ganglia that are closely associated with the expected neuropathology of bvFTD.

1.5 Closing Remarks

In this first chapter we described the brain as a complex system modeled as a network where the vertices are brain regions and edges represent the interactions between them. An accurate description of the architecture and a characterization of the network properties can be important to understand the functional and structural connectivity of the brain. In this contest the network analysis describes complex systems by quantifying properties of network representation, such as centrality, efficiency, density and small worldness characteristics.

Here the basic terminology and notation for graphs were discussed, then this theory has been applied to a particular case study in order to investigate the structure and the characteristics of brain networks and to represent the local and the global structural changes, as well as their dynamics. In this applicative study, network analysis has been used also to study how brain functional connectivity changes with pathology. This was achieved by studying with graph theory rs-fMRI data of subjects coming from different populations, including healthy subjects and patients with neurodegenerative diseases. In this context, network analysis might contribute to detect abnormalities of network connectivity in different brain disorders and to improve our understanding of the mechanisms responsible of the development of irreversible disability.

Rs-fMRI correlations are defined by the temporal relationships of distinct brain areas observed in the absence of experimental demands. To generate a representation of an rs-fMRI brain network, nodes should first be defined by parcellating the brain into meaningful objects of interest (e.g., brain areas).

Building the adjacency matrix (i.e., defining the edges) requires a quantification of the pairwise relationship between each of the nodes in the network (e.g., the presence of a significant correlation between two areas resting-state time courses). In each of these steps a different choice of techniques is required.
In our study the brain network was clustered employing an automated anatomical labeling (AAL) method, but it could be interesting to explore differences in node definition using other methods (e.g anatomic parcels, random divisions of anatomic parcels, voxels).
Another choice taken during the brain network construction concerns the definition of an edge. In a graph, an edge represents the pairwise relationship or interaction between two independent nodes. We have focused our discussion on measures of rs-fMRI whereby time courses of activity were extracted from each node, pairwise correlations between the time courses of each node pair were calculated using the Pearson's correlation coefficient to quantify the strength of relationships. Several alternatives can be chosen to estimate the strength of functional relationships. For example, pairwise relationships may be quantified according to spectral coherence, correlations of wavelets derived from the variance decomposition of the resting-state time series, or measurements that estimate directional information. As an alternative to the full correlation as a measure of strength of relationships, a partial correlation coefficient can be used, in this case the resulting network will be directed. In Smith *et al.* (2011) a set of simulations with a model for fMRI time series is discussed in order to determine which dependence measures are the most sensitive in detecting network functional connections.
Finally, we constructed binary adjacency matrices according to a threshold. It can be interesting to note that network edges are easy to define and can be represented by the presence or absence of the relationship of interest, but variable or weighted relationships can also be studied using a weighted graphs, considering the information on differences in the strength of the relationships across pairs of nodes. In our study we decided to consider undirected and unweighted networks. Care must be taken when thresholding an adjacency matrix to remove edges. Although thresholding can remove weak

relationships that may not be physiologically relevant, excessive thresholding for the purposes of computing graph properties or constructing comparable graphs across subjects or cohorts can result in the creation of a matrix that, while sparse, misrepresents the underlying connectivity (Wig *et al.*, 2011). In Chapter 2 we will deal with the problem of choosing the threshold and we will propose a statistical solution.

As we have briefly reported in this section, the emerging field of complex brain networks raises a number of interesting questions and provides some of the first quantitative insights into general topological principles of brain network organization. The growth in the statistical mechanics of complex networks, and the use of graph theoretical analysis, suggests that this approach will play an increasingly important part to study the neurological connectivity.

Chapter 2

Statistical Assessing of Adjacency Matrices

As we have seen in the previous chapter, complex network analysis aims to characterize brain networks with global and local measures (Rubinov and Sporns, 2010). A network is a mathematical representation of a real-world complex system (brain) and it is defined by a collection of nodes (brain regions) and links (measure of dependence) between pairs of nodes. To simplify analysis, networks are often reduced to a binary (undirected) form, through an adjacency matrix (Kolaczyk, 2009). The correlation matrices are typically converted to a graph by considering a threshold τ, to create an adjacency matrix. If the correlation coefficient between a pair of nodes exceeds τ, an edge is said to exist between their representing vertices; otherwise no edge exists between them. In previous works there was no unique way to choose τ and the possible network configurations were examined by constructing graphs for a range of values of τ within which the consistency of the network characteristics was explored (Sanz-Arigita *et al.*, 2010). The aim of this section is to propose an original method to derive the adjacency matrix using a multiple test on correlation coefficient and to provide a statistical evaluation of the error committed in constructing the adjacency matrix, considering the problem of multiple comparisons.

2.1 Test on correlation coefficient and multiple comparisons problem

The proposed method to construct the adjacency matrix works as follows. The brain connectivity is defined as temporal correlations between spatially distinct brain regions investigated during task-free (resting-state) conditions. Time series of t temporal instants were extracted from each region by averaging the signal of all voxels within it. Bivariate correlations between each ROIs pair were obtained by calculating the Pearson's correlation coefficient between node time courses. The correlation matrix is computed by compiling all pairwise correlation ρ_{ij} between node i and node j. The binary adjacency matrix is constructed according to a test on correlation coefficient, one for each entry of the correlation matrix, instead of applying an arbitrary threshold ($\hat{\rho_{ij}} > \tau$), with: $H_0 : \rho_{ij} \leq 0$ vs $H_1 : \rho_{ij} > 0$, where H_0 and H_1 are null and alternative hypothesis respectively. The entry of the adjacency matrix will be 1 if H_0 is rejected and zero otherwise. Note that with the assumption of normality the exact probability density of the Pearson correlation coefficient is a T-Student on $t - 2$ degrees of freedom and when t is large this can be approximated with a Normal distribution (Landenna *et al.*, 1997). In the case of study t corresponds to the 200 temporal instants associated to every single node.

Let us to introduce, briefly, the basic paradigm for single hypothesis. We wish to test a null hypothesis H_0 versus an alternative H_1, based on a statistic X. For a given rejection region Γ, we reject H_0 when $X \in \Gamma$, we accept H_0 when $X \notin \Gamma$, otherwise. A type I error occurs when $X \in \Gamma$, but H_0 is really true. A type II error occurs when $X \notin \Gamma$ but H_1 is really true. To choose Γ, the type I error is set to a level α and all rejection regions that have a type I error less or equal to α and the one that has the lowest type II error is chosen.

We may notice that there is a relation between the threshold on correlation coefficient and the threshold on p-value of the considered test. If τ is the threshold on correlation coefficient, then the rejection region $\hat{\rho} \geq \tau$ can be

written as:

$$\text{p-value} \leq \gamma \tag{2.1}$$

where

$$\gamma = \Phi\left(-\sqrt{\frac{(t-2)\tau^2}{1-\tau^2}}\right) \tag{2.2}$$

and Φ is the Standard Normal distribution function.
Test on correlation matrix, described before, involves multiple-hypothesis problem. In this case the situation is more complicated because it becomes unclear how we should measure the overall error rate. In this work we decided to control the positive false discovery rate ($pFDR$), following Storey (2002), which proposes to fix the rejection region and, then, to estimate its corresponding error rate. We will see that this approach offers increased applicability, accuracy and power. We will see later that this method involves the punctual estimation of both the FDR and the $pFDR$. Besides we consider a similar measure involving false negatives (type II errors), which is called the positive false nondiscovery rate ($pFNR$) (Genovese *et al.*, 2002). In particular, we propose a point estimation for $pFNR$ and a method to balance the two types of error. In addition we propose an interval estimation for $pFDR$ and $pFNR$, based on the Bootstrap method.
In the following section we will deal with the problem of multiple comparisons and with the method based on the estimation of $pFDR$ and $pFNR$ and we will describe how it can be useful in our context.

2.2 From FWER to FDR

When comparing several medical treatments, the numbers of treatments is typically fairly small and it is possible to disregard the problem of multiple comparisons. But, instead, recent applications of multiple comparison theory occurs, for example, in microarrays, where thousands or even tens of thousands of genes are tested simultaneously. Each microarray corresponds to one unit and in these experiments the sample size (the number of such units) is typically of a much smaller order of magnitude (in the tens) than

m	1	2	5	10	50	100
Pr(at least one false rejection)	0.05	0.10	0.23	0.40	0.92	0.99

Table 2.1: The probability of one or more false rejections when all of the hypotheses H_0^i are true at level $\alpha = 0.05$

the number of comparisons being tested (Lehmann and Romano, 2005). The same problem also occurs testing correlation coefficients in a correlation matrix, we have to test $\frac{N(N-1)}{2}$ hypotheses, corresponding to the elements of the upper triangular matrix (where N is the number of nodes).
Let us now consider the general problem of simultaneously testing m hypotheses $H_0^i (i = 1, ..., m)$. We shall assume that tests for the individual hypotheses are available and the problem is how to combine them into a multiple test procedure. The easiest approach is to simply test each hypothesis at level α. However, with such a procedure the probability of one or more false rejections rapidly increases with m. Table 2.1 shows the probability of one or more false rejections when all of the hypotheses $H_0^1, ..., H_0^m$ are true, when the test statistics used for testing $H_0^1, ..., H_0^m$ are independent, and when the level at which each of the m hypotheses is tested is $\alpha = 0.05$. In this sense the claim that the procedure controls the probability of false rejections at level 0.05 is very misleading. We shall therefore replace the usual condition for testing a single hypothesis, that the probability of a false rejection not exceed α, by the requirement, when testing several hypotheses, that the probability of one or more false rejections, not exceed a given level. This probability is called the family-wise error rate ($FWER$), where the term family refers to the collection of hypotheses H_0^i that is being considered for joint testing. Once the family has been defined, we shall require that

$$FWER \leq \alpha. \tag{2.3}$$

So:

$$FWER = Pr(\cup \Gamma_i | \cap H_0^i) \leq \sum_i^m Pr(\Gamma_i | \cap H_0^i) \tag{2.4}$$

the last step follows from the Boole's inequality,

$$\sum_{i}^{m} Pr(\Gamma_i | \cap H_0^i) = \sum Pr(\Gamma_i | H_0^i) = m\alpha \tag{2.5}$$

so if we want to control $m\alpha \leq \epsilon$ we have to choose $\alpha = \frac{\epsilon}{m}$ that is known as the Bonferroni procedure and, if m is high, this is a very conservative control.
When the tests are independent:

$$FWER = Pr(\cup \Gamma_i | \cap H_0^i) = 1 - Pr(\cap \bar{\Gamma}_i | \cap H_0^i). \tag{2.6}$$

Using the hypothesis of independence:

$$Pr(\cap \bar{\Gamma}_i | \cap H_0^i) = \prod_{i}^{m} Pr(\bar{\Gamma}_i | \cap H_0^i) = \prod_{i}^{m} [1 - Pr(\Gamma_i | \cap H_0^i)] = \prod_{i}^{m} [1 - Pr(\Gamma_i | H_0^i)]; \tag{2.7}$$

So:

$$FWER = 1 - \prod_{i}^{m} [1 - \alpha] = 1 - [1 - \alpha]^m. \tag{2.8}$$

We can notice that:

$$\forall \epsilon \in (0,1) \quad \exists \quad M \quad t.c. \quad \forall m \geq M \quad FWER \geq \epsilon \tag{2.9}$$

that is: when m increases there is no way to control $FWER$ (it tends to one). In particular, when the number of true hypotheses is large, we shall be nearly certain to reject some of them.
When the number of tests is in the tens or hundreds of thousands, control of the $FWER$ at conventional levels becomes too conservative. A radical weakening of the $FWER$ was proposed by Benjamini and Hochberg (1995), who suggested the following. For a given multiple testing decision rule, let R be the total number of rejections and let V be the number of false rejections, i.e., the number of rejections among the R rejections corresponding to true null hypotheses (see Table 2.2). Define Q to be $\frac{V}{R}$ if $R \neq 0$ and to be 0 if $R = 0$.Thus Q is the proportion of rejected hypotheses that are rejected erroneously. When none of the hypotheses are rejected, both numerator and denominator of that proportion are 0, and Q is then claimed to be 0. The

Hypothesis	Accept null	Reject null	Total
Null true	U	V	m_0
Alternative true	T	S	m_1
	W	R	m

Table 2.2: Possible outcomes from m hypothesis tests

false discovery rate (FDR) is defined by

$$FDR = E(Q). \tag{2.10}$$

With this notation the $FWER$ becomes $Pr(V \geq 1)$. When all hypotheses are true, $FDR = FWER$. In general, $FDR \leq FWER$ and typically this inequality is strict, so that the FDR is more liberal (in the sense of permitting more rejections) than the $FWER$ (Lehmann and Romano, 2005). Benjamini and Hochberg (1995) provided a sequential p-value method to control FDR: using the observed data, it estimates the rejection region so that $FDR \leq \alpha$ for some pre-chosen α. The product of a sequential p-value method is an estimate $\hat{k}$ that tells us to reject $P_{(1)}, P_{(2)}, \ldots, P_{(\hat{k})}$ where $P_{(1)} \leq P_{(2)} \leq, \ldots, \leq P_{(m)}$ are the ordered observed p-values. If

$$\hat{k} = \max_k \{k : p_{(k)} \leq \alpha \frac{k}{m}\} \tag{2.11}$$

is calculated then rejecting the null hypotheses corresponding to $P_{(1)} \leq, \ldots, \leq P_{(\hat{k})}$ provides $FDR = \frac{m_0}{m}\alpha \leq \alpha$. If there is no p-value satisfying this inequality, then no hypothesis test is called significant. The FDR offers less stringent control over Type I errors than the $FWER$, and is therefore usually more powerful. Storey (2003) underlines that, in reality, this process involves estimation and that the more variable the estimate of $\hat{k}$ is, the worse the procedure will work, so it is important to explore the reliability of $\hat{k}$ case by case. Another weakness of this approach to false discovery rates is that the error rate is controlled for all values of m_0 (the number of true null hypotheses) simultaneously without using any information in the data about it. In the method proposed by Storey (2003), this information is used yielding a less stringent procedure and more power, while maintaining strong control.

Often, the power of the multiple-hypothesis testing method decreases with increasing m. Especially when the tests are independent. This should not be so, the larger m is, the more information we have about m_0 that should be used.

2.3 The Positive False Discovery Rate (pFDR)

Storey (2002) notes that (2.10) can be also written as follows [1]:

$$FDR = E\Big(\frac{V}{R}|R > 0\Big)Pr(R > 0), \tag{2.12}$$

that is the expected proportion of false positive findings among all rejected hypotheses times the probability of making at least one rejection. This quantity has to be considered because in most cases there is a positive probability that $R = 0$, so V/R is not well-defined. Storey (2002), instead, proposes to use the positive false discovery rate ($pFDR$), defined as follows:

$$pFDR = E\Big(\frac{V}{R}|R > 0\Big), \tag{2.13}$$

where the expectation of Q is conditioned on the event that positive findings have occurred, so this quantity does't consider the case where no test is significant. Furthermore, we can notice, from (2.12) and (2.13), that $FDR \leq pFDR$.

There are two approaches that can be taken. The first is to fix the acceptable rate α beforehand and estimate a significance threshold to obtain this rate conservatively on average. The second is to fix the significance threshold and provide a conservative estimate of the rate over that threshold. When taking the first approach, we are forced to use the FDR since the $pFDR$ cannot be controlled in this sense. The $pFDR$ can be conservatively estimated in the second approach (Storey, 2002).

[1] $E\Big(\frac{Y}{\max\{X,1\}}\Big) = E\Big(\frac{Y}{X}|X > 0\Big)Pr(X > 0)$, which holds when $Y = 0$ if $X = 0$ and $X, Y \geq 0$

2.3.1 Inference for FDR and pFDR

For simplicity, we suppose to test m hypotheses $H_0^1, H_0^2, \ldots, H_0^m$ with independent test statistics $T_1, T_2, \ldots, T_m$.

Theorem 1.

Suppose that:

- the same rejection region, Γ, is used for each test;
- the hypotheses are simple and represented as identical distributed and independent Bernoulli random variables with $Pr(H_0^i \text{ is true}) = \pi_0$ and $Pr(H_0^i \text{ is false}) = \pi_1$;

which make the tests identical. Then we can write:

$$\begin{aligned} pFDR(\Gamma) &= Pr(H_0 \text{ is true}|\Gamma) \\ &= \frac{\pi_0 Pr(\Gamma|H_0 \text{ is true})}{Pr(\Gamma)} \\ &= \frac{\pi_0\{\text{Type I error of } \Gamma\}}{\pi_0\{\text{Type I error of } \Gamma\}+\pi_1\{\text{Power of } \Gamma\}}. \end{aligned} \tag{2.14}$$

For a proof of this theorem see Storey (2003). In addition, Storey and Tibshirani (2001) shows how to estimate the pFDR in more general situations. Theorem 1 shows that $pFDR$ increases with increasing Type I errors and decreases with increasing power. With (2.14) we can write $pFDR$ in a form that does not depend on m. For large m the assumption that H_0^i are random makes little difference (Storey, 2003).
In terms of p-values we can write the result of Theorem 1 as:

$$pFDR(\gamma) = \frac{\pi_0 Pr(P \leq \gamma|H_0 \text{ is true})}{Pr(P \leq \gamma)} = \frac{\pi_0 \gamma}{Pr(P \leq \gamma)}, \tag{2.15}$$

where P is the random p-value from any test. We can notice that

$$Pr(P > \lambda, H_0) = Pr(P > \lambda|H_0)P(H_0) = [1 - Pr(P \leq \lambda|H_0)]P(H_0) = (1-\lambda)\pi_0,$$

for some well chosen λ

$$Pr(P > \lambda, H_0) \cong \frac{\#\{p_i > \lambda\}}{m};$$

therefore a conservative estimate of π_0 is given by

$$\widehat{\pi_0}(\lambda) = \frac{\#\{p_i > \lambda\}}{(1-\lambda)m} = \frac{W(\lambda)}{(1-\lambda)m} \tag{2.16}$$

where $p_1, \ldots, p_m$ are the observed p-values and $W(\lambda) = \#\{p_i > \lambda\}$. The choice of the optimal value of the tuning parameter λ will be the topic of the next section.
A natural estimate of $Pr(P \leq \gamma)$ is

$$\widehat{Pr}(P \leq \gamma) = \frac{\#\{p_i \leq \gamma\}}{m} = \frac{R(\gamma)}{m}, \tag{2.17}$$

where $R(\gamma) = \#\{p_i \leq \gamma\}$. Therefore , a natural estimate of (2.15), for fixed λ is

$$\widehat{Q}_\lambda(\gamma) = \frac{\widehat{\pi}_0(\lambda)\gamma}{\widehat{Pr}(P \leq \gamma)} = \frac{W(\lambda)\gamma}{(1-\lambda)R(\gamma)}. \tag{2.18}$$

In finite sample contest, we must make two adjustments to estimate $pFDR$. When $R(\gamma) = 0$, the estimate would be undefined, so we replace $R(\gamma)$ with $\max\{R(\gamma), 1\}$. Since

$$E\Big(\frac{Y}{\max\{X,1\}}\Big) \geq E\Big(\frac{Y}{X}|X>0\Big)Pr(X>0) \quad \text{for} X, Y \geq 0$$

we obtain

$$E(\hat{Q}|R(\gamma)>0) \leq E\left[\frac{W(\lambda)\gamma}{(1-\lambda)\max\{R(\gamma),1\}Pr(R(\gamma)>0)}\right]. \tag{2.19}$$

Since that $Pr(R > 0) \geq Pr(V \geq 1) = FWER$ and, when the tests are independent $FWER = 1 - (1-\gamma)^m$, then $1 - (1-\gamma)^m$ is a lower bound for $Pr(R(\gamma) > 0)$ and $\gamma/[1-(1-\gamma)^m]$ is a conservative estimate of the type I error, conditional that $R(\gamma) > 0$. Therefore, Storey (2002) estimates $pFDR$

as:

$$\widehat{pFDR}_\lambda(\gamma) = \frac{\hat{\pi}_0(\lambda)\gamma}{\widehat{Pr}(P \le \gamma)[1-(1-\gamma)^m]} = \frac{W(\lambda)\gamma}{(1-\lambda)\max\{R(\gamma),1\}[1-(1-\gamma)^m]}. \tag{2.20}$$

Since FDR is not conditioned on at least one rejection occurring, $\widehat{FDR}$ can be written:

$$\widehat{FDR}_\lambda(\gamma) = \frac{\hat{\pi}_0(\lambda)\gamma}{\widehat{Pr}(P \le \gamma)} = \frac{W(\lambda)\gamma}{(1-\lambda)\max\{R(\gamma),1\}}. \tag{2.21}$$

Large Sample Results

Storey (2002) proves that $pFDR$ and FDR are asymptotically equivalent for a fixed rejection region. For m identical simple hypothesis tests, let

$$g(\lambda) = Pr(P \le \lambda | H_0 \text{ is false}) \tag{2.22}$$

be the power as a function of type I error λ. Note that $g(\cdot)$ is the cumulative density function of the alternative p-values. If the alternative hypothesis is composite, then $g(\lambda)$ must be defined as an appropriate mixture. We assume that $g(0) = 0$, $g(1) = 1$ and $g(\lambda) \ge \lambda$ for $0 < \lambda < 1$ (unbiasedness condition). In the following the asymptotic properties are discussed.

Theorem 2.

With probability 1:

$$\lim_{m\to\infty}\{\widehat{pFDR}_\lambda(\gamma)\} = f(\lambda)pFDR(\gamma) \ge pFDR(\gamma), \tag{2.23}$$

where

$$f(\lambda) = \frac{\pi_0 + \pi_1[1-g(\lambda)]/(1-\lambda)}{\pi_0} \ge 1. \tag{2.24}$$

Proof. By the strong law of large numbers, with probability 1:

$$\frac{W(\lambda)}{m} \to Pr(P > \lambda) =$$
$$= Pr(P > \lambda|H_0)Pr(H_0) + Pr(P > \lambda|H_1)Pr(H_1) =$$
$$= (1-\lambda)\pi_0 + [1-g(\lambda)]\,\pi_1$$

and so, with probability 1:

$$\begin{aligned} \hat{\pi_0}(\lambda) = \frac{W(\lambda)}{(1-\lambda)m} \to \pi_0 + \frac{1-g(\lambda)}{1-\lambda}\pi_1 = \\ = \pi_0\left[1 + \frac{1-g(\lambda)}{1-\lambda}\frac{\pi_1}{\pi 0}\right] = \pi_0 f(\lambda) \geq \pi_0 \end{aligned} \tag{2.25}$$

this proves that $\hat{\pi_0}$ is an asymptotically conservative estimate for π_0.
Then, by the strong law of large numbers, with probability 1:

$$\frac{R(\gamma)}{m} \to Pr(P \leq \gamma) =$$
$$= Pr(P \leq \gamma|H_0)Pr(H_0) + Pr(P \leq \gamma|H_1)Pr(H_1) = \gamma\pi_0 + (1-\beta)\,\pi_1. \tag{2.26}$$

We may note that $\frac{R(\gamma)}{m}$ is an asymptotically conservative estimate of the power $1-\beta$.
Besides

$$\begin{aligned} \widehat{pFDR}_\lambda(\gamma) = \\ = \frac{\gamma\widehat{\pi_0}(\lambda)}{\frac{R(\gamma)}{m}[1-(1-\gamma)^m]} \to \frac{\gamma\pi_0 f(\lambda)}{Pr(P\leq\gamma)} = \\ = pFDR(\gamma)f(\lambda) \geq pFDR(\gamma), \end{aligned} \tag{2.27}$$

almost surely.

□

In particular we can notice that $\frac{\gamma\pi_0 f(\lambda)}{Pr(P\leq\gamma)} \cong pFDR(\gamma)$ when $g(\lambda) \cong 1$, that is the $\widehat{pFDR}$ converges almost surely to $pFDR$ when the test is powerful.
Moreover, we observe that

$$\begin{aligned} \frac{R(\lambda)}{m} \to Pr(P \leq \lambda) = \\ = Pr(P \leq \lambda|H_0)Pr(H_0) + Pr(P \leq \lambda|H_1)Pr(H_1) = \\ = \lambda\pi_0 + g(\lambda)\pi_1 \end{aligned} \tag{2.28}$$

almost surely. So $\frac{R(\lambda)}{m}$ is an asymptotically conservative estimate of $g(\lambda)$ and we can select a value of the tuning parameter λ such that $g(\lambda)$ is near to one as we want.
Under the assumptions of Theorem 1, $\widehat{Q}_\lambda(\gamma)$ is the maximum likelihood estimate of

$$f(\lambda)pFDR(\gamma). \tag{2.29}$$

For a proof of this property see Storey (2003). This quantity is slightly greater than $pFDR(\gamma)$ for powerful tests. In situations where g is unknown, this estimate is optimal in that the bias can usually be made arbitrarily small, while obtaining the smallest asymptotic variance for an estimator of that bias. $\widehat{pFDR}_\lambda(\gamma)$ is asymptotically equivalent to $\widehat{Q}_\lambda(\gamma)$, so it has the same large sample properties.

Finite Sample Results

For finite sample considerations, Storey (2002) reports the main finite property of $\widehat{pFDR}$ and$\widehat{FDR}$:

$$E[pFDR_\lambda(\gamma)] \geq pFDR(\gamma) \text{ and } E[FDR_\lambda(\gamma)] \geq FDR(\gamma), \tag{2.30}$$

for all λ and π_0.

2.3.2 The q-value

The q-value represents the $pFDR$ analogue of the p-value and gives a hypothesis testing error measure for each observed statistic with respect to $pFDR$. The p-value, in a single hypothesis test, is the minimum type I error rate that can occur when we reject a hypothesis, using a statistic T with value t for the set of rejection regions Γ. In a multiple testing contest, usually, the p-values of several statistics, $T_1, \ldots, T_m$, are adjusted to control $FWER$. The adjusted p-values give a measure of the strength of an observed statistic with respect to making one or more type I errors. In the contest of $pFDR$, as a natural extension to it, the q-value can be defined as follows (Storey,

2002):

$$q(t) = \inf_{\{\Gamma: t \epsilon \Gamma\}} \{pFDR(\Gamma)\}. \tag{2.31}$$

So the q-value is the minimum $pFDR$ that can occur when rejecting a statistic T with value t for the set of rejection regions Γ. In terms of independent p-values, the q-value of an observed p-value p is

$$q(p) = \inf_{\gamma \geq p}\{pFDR(\gamma)\} = \inf_{\gamma \geq p}\Big\{\frac{\pi_0 \gamma}{Pr(P \leq \gamma)}\Big\}. \tag{2.32}$$

The q-values can be used in practice in the following way: $\hat{q}(p_{(i)})$ gives us the minimum $pFDR$ that we can achieve for rejection regions containing $[0, p_{(i)}]$ for $i = 1, \ldots, m$. In other words, for each p-value there is a rejection region with $pFDR = q(p_{(i)})$ so that at least $p_{(1)}, \ldots, p_{(i)}$ are rejected, where $\hat{q}(p_{(i)})$ is an estimate of $q(p_{(i)})$.

2.3.3 Looking for the optimal value of the tuning parameter

The method, proposed by Storey (2002), to calculate the optimal value of the tuning parameter for $\widehat{pFDR}$ and $\widehat{FDR}$, consists in minimize the mean square error (MSE) of $\widehat{pFDR}$ and $\widehat{FDR}$. The author uses the Bootstrap method to estimate this parameter, λ_1,

$$\lambda_1 = \text{argmin}_{\lambda \epsilon [0,1]}\{E[(\widehat{pFDR}_\lambda(\gamma) - pFDR(\gamma))^2]\} \tag{2.33}$$

where $E\{[\widehat{pFDR}_\lambda(\gamma) - pFDR(\gamma)]^2\} = MSE(\lambda)$. In order to estimate MSE over a range of λ from 0 to 0.95 with step 0.05, $\widehat{pFDR}^{*b}_\lambda(\gamma)$ (for $b = 1, \ldots, B$), the version of the estimate $\widehat{pFDR}_\lambda(\gamma)$ is used for any fixed λ. Then the Bootstrap estimate of the $MSE(\lambda)$ would be

$$\frac{1}{B}\sum_{b=1}^{B}[\widehat{pFDR}^{*b}_\lambda(\gamma) - pFDR(\gamma)]^2, \tag{2.34}$$

but because we don't know $pFDR(\gamma)$, we must form a plug-in estimate of it, that is, according to Shao and Tu (1996), $\{\widehat{pFDR}_\lambda(\gamma)\}$, so the estimate

used for $MSE_{pFDR}(\lambda)$ is

$$\widehat{MSE_{pFDR}}(\lambda) = \frac{1}{B}\sum_{b=1}^{B}[\widehat{pFDR}_{\lambda}^{*b}(\gamma) - \{\widehat{pFDR}_{\lambda}(\gamma)\}]^2. \tag{2.35}$$

We note that another method to calculate the optimal value of λ consisting in calculate the shortest length confidence interval of $\widehat{pFDR}$ and $\widehat{FDR}$, using $\widehat{pFDR}_{\lambda}^{*b}(\gamma)$ and $\widehat{FDR}_{\lambda}^{*b}(\gamma)$ for any fixed λ. We use the Bootstrap Percentile method identifying that value of λ corresponding to the minimum size confidence interval at level 95%. We found that the two methods lead approximately to the same value for λ_1.

2.4 The False Nondiscovery Rate (FNR)

The $pFDR$ (and the FDR) by itself don't tell us what proportion of the true hypotheses were detected (Genovese *et al.*, 2002). For this purpose we also need to quantify the size of a type II error (the acceptance of H_0 when is false) and to verify that the two kinds of error are reasonably small.
Let F be the proportion of alternative true declared false, that is, according to Table 2.2, the number of false acceptance T among the $W = (m - R)$ total acceptances, the false non-discovery rate (FNR) is defined by

$$FNR = E(F), \tag{2.36}$$

where

$$F = \frac{T}{W} = \frac{m_1 - S}{m - R} = \frac{m - m_0 - (R - V)}{m - R}. \tag{2.37}$$

As we have seen in (2.12), the (2.36) can be also written:

$$FNR = E\Big(\frac{T}{W}|W > 0\Big)P(W > 0) \tag{2.38}$$

and we can define the $pFNR$, in analogy with (2.12), as

$$pFNR = E\Big(\frac{T}{W}|W > 0\Big) \tag{2.39}$$

(Storey, 2003). Furthermore we obtain the following characterization of $pFNR$.

Theorem 3.

Under the same hypotheses of the Theorem 1

$$pFNR(\gamma) = \frac{\beta\pi_1}{Pr(P > \gamma)}. \tag{2.40}$$

Proof.

Applying Theorem 1 to T, W, H_1 and $P > \gamma$ in place of R, V, H_0 and Γ respectively we obtain that

$$pFNR(\gamma) = E(\frac{T}{W}|W > 0) = Pr(H_1 \text{ is true}|P > \gamma).$$

Then

$$Pr(H_1 \text{ is true}|P > \gamma) = \frac{Pr(P>\gamma|H_1 \text{ is true})Pr(H_1 \text{ is true})}{Pr(P>\gamma)} = \frac{\beta\pi_1}{Pr(P>\gamma)}.$$

□

From (2.40) it follows that $pFNR \geq FNR$. Then we can notice that the definition of the $pFNR$ includes the value of β that is the probability of type II error. If we suppose to know the distribution of the null and alternative hypothesis, for example:
$H_0 : Z \sim N(0,1)$ and $H_1 : Z \sim N(2,1)$, where $N(\mu, \sigma^2)$ is the Normal distribution with mean μ and variance σ^2, then

$$\beta = Pr(Z \leq z_{1-\gamma}|H_1 \text{ is true}) = \Phi(z_{1-\gamma} - 2),$$

where $z_{1-\gamma}$ is the $1-\gamma$-quantile of the Standard Normal distribution and Φ is its cumulative distribution function.
In real problem, β is unknown and has to be estimated. For this reason $pFNR$ is more difficult to estimate that the $pFDR$. With the purpose to find a good estimate of $pFNR$ we can notice that, for unbiased test (where

$1-\beta \geq \gamma$ that is $\beta \leq 1-\gamma$) we can write the following inequality:

$$FNR \leq pFNR \leq \frac{(1-\gamma)\pi_1}{Pr(P>\gamma)},$$

then we can estimate π_1 and $Pr(P>\gamma)$ respectively:

$$\hat{\pi_1} = \frac{\#\{p_i \leq \lambda\}}{m} = \frac{R(\lambda)}{m}$$

$$\hat{Pr}(P>\gamma) = \frac{\#\{p_i > \gamma\}}{m} = 1 - \frac{R(\gamma)}{m}, \tag{2.41}$$

then a conservative estimate of FNR can be written as:

$$\tilde{FNR} = \frac{(1-\gamma)R(\lambda)}{m - R(\gamma)}. \tag{2.42}$$

Nevertheless we found that (2.42) is too conservative. The discussions so far have focused on type II error rate, however in addition to it, of importance is also the power Pw. The power, in the context of multiple hypothesis testing, can be defined as:

$$Pw = \frac{S}{m_1} = \frac{R-V}{m-m_0} = \frac{(1-\frac{V}{R})R}{(1-\frac{m_0}{m})m} = \frac{(1-Q)R}{(1-\pi_0)m} \tag{2.43}$$

and

$$Pwa = E(Pw) = E\left(\frac{S}{m_1}\right)$$

is the average power which is equal to $Pr\left(P \leq \gamma | H_1 \text{ is true}\right) = 1-\beta$ for simple hypothesis. Then we can propose the following estimate for P_{wa} from (2.43):

$$\hat{Pwa} = \frac{(1-\widehat{pFDR})R}{(1-\hat{\pi_0})m}. \tag{2.44}$$

Let consider the complement to one of P_{wa}, the NDR:

$$NDR = 1 - Pwa,$$

$$\widehat{NDR} = 1 - \hat{Pwa}, \tag{2.45}$$

according to Craiu and Sun (2008).
Furthermore we can see that the quantity in (2.45) can be a candidate as estimator of β in (2.40). From (2.40), using (2.41) and (2.45) we obtain

$$\widehat{pFNR} = \frac{(1-\hat{\pi}_0)m\widehat{NDR}}{m-R} = \frac{(1-\hat{\pi}_0)m-(1-\widehat{pFDR})R}{m-R} \tag{2.46}$$

which represents our proposal for estimating the $pFNR$.
In analogy with the method proposed by Storey (2002), we can calculate the optimal value of the tuning parameter, minimizing the mean square error (MSE) of $\widehat{pFNR}$ as in (2.33):

$$\lambda_2 = \text{argmin}_{\lambda\epsilon[0,1]}\{E[(\widehat{pFNR}_\lambda(\gamma) - pFNR(\gamma))^2]\}. \tag{2.47}$$

In this case the estimate used for $MSE(\lambda)$ is

$$\widehat{MSE_{pFNR}}(\lambda) = \frac{1}{B}\sum_{b=1}^{B}[\widehat{pFNR}_\lambda^{*b}(\gamma) - \{\widehat{pFNR}_\lambda(\gamma)\}]^2, \tag{2.48}$$

where $b = 1, \ldots B$. Similarly, we define the optimal tuning parameter for $\widehat{Pwa}$ as:

$$\lambda_3 = \text{argmin}_{\lambda\epsilon[0,1]}\{E[\widehat{Pwa}_\lambda(\gamma) - Pwa(\gamma))^2]\} \tag{2.49}$$

and

$$\widehat{MSE_{Pwa}}(\lambda) = \frac{1}{B}\sum_{b=1}^{B}[\widehat{Pwa}_\lambda^{*b}(\gamma) - \{\widehat{Pwa}_\lambda(\gamma)\}]^2. \tag{2.50}$$

Besides, we propose an interval estimation of $pFNR$ and Pwa, using the Bootstrap Percentile method and finding a $1-\alpha$ upper confidence interval by taking the $1-\alpha$ quantile of the $\widehat{pFNR}_{\lambda_2}^{*1}(\gamma), \ldots, \widehat{pFNR}_{\lambda_2}^{*B}(\gamma)$ as the upper confidence bound and finding the α lower confidence interval by taking the α quantile of $\widehat{Pwa}_{\lambda_3}^{*1}(\gamma), \ldots, \widehat{Pwa}_{\lambda_3}^{*B}(\gamma)$ as the lower confidence bound.

Large Sample Results

Now we study the large sample properties of the proposal estimates for Pwa and $pFNR$ for a fixed rejection region. Recall the power as a function of

type I error in (2.22) and the quantity in (2.24).

Theorem 4.

With probability 1:

$$\lim_{m\to\infty}\{\widehat{Pwa}_\lambda(\gamma)\} = \frac{\gamma\pi_0/\pi_1[1-f(\lambda)]+1-\beta}{\pi_0/\pi_1[1-f(\lambda)]+1} \tag{2.51}$$

where $\frac{\gamma\pi_0/\pi_1[1-f(\lambda)]+1-\beta}{\pi_0/\pi_1[1-f(\lambda)]+1} \cong 1-\beta$ when $g(\lambda)\cong 1$.

Proof. By the asymptotic behavior of $\widehat{\pi_0}(\lambda)$ seen in (2.25), of $\frac{R(\gamma)}{m}$ in (2.26) and of $\widehat{pFDR}_\lambda(\gamma)$ in (2.27), we obtain that

$$\begin{aligned}\widehat{Pwa}_\lambda(\gamma) = \frac{[1-\widehat{pFDR}_\lambda(\gamma)]\frac{R(\gamma)}{m}}{1-\widehat{\pi_0}(\lambda)} \to \frac{\gamma\pi_0+(1-\beta)\pi_1-\gamma\pi_0 f(\lambda)}{1-\pi_0 f(\lambda)} = \\ = \frac{\gamma\pi_0[1-f(\lambda)]+(1-\beta)\pi_1}{\pi_0[1-f(\lambda)]+\pi_1} = \frac{\gamma\pi_0/\pi_1[1-f(\lambda)]+1-\beta}{\pi_0/\pi_1[1-f(\lambda)]+1}\end{aligned} \tag{2.52}$$

almost surely.

□

Furthermore, from the previous result, it follows that, with probability 1:

$$\lim_{m\to\infty}\{\widehat{NDR}_\lambda(\gamma)\} = \lim_{m\to\infty}\{1-\widehat{Pwa}_\lambda(\gamma)\} = \frac{(1-\gamma)[1-f(\lambda)]\pi_0/\pi_1+\beta}{\pi_0/\pi_1[1-f(\lambda)]+1} \tag{2.53}$$

where $\frac{(1-\gamma)[1-f(\lambda)]\pi_0/\pi_1+\beta}{\pi_0/\pi_1[1-f(\lambda)]+1} \cong \beta$ when $g(\lambda)\cong 1$.

Theorem 5.

With probability 1:

$$\lim_{m\to\infty}\{\widehat{pFNR}_\lambda(\gamma)\} = \frac{(1-\gamma)\pi_0[1-f(\lambda)]}{Pr(p>\gamma)} + pFNR(\gamma) \tag{2.54}$$

where the limit is $\cong pFNR$ when $g(\lambda)\cong 1$.

Proof. By the asymptotic behavior of $\widehat{\pi_0}(\lambda)$ in (2.25), of $\frac{R(\gamma)}{m}$ in (2.26) and

of $\widehat{pFDR}_\lambda(\gamma)$ in (2.27), we obtain that:

$$\begin{aligned}\widehat{pFNR}_\lambda(\gamma)) &= \frac{[1-\widehat{\pi_0(\lambda)}]m-[1-\widehat{pFDR}_\lambda(\gamma)]R(\gamma)}{m-R(\gamma)} = \\ &= \frac{1-\widehat{\pi_0}(\lambda)-R(\gamma)/m+\widehat{pFDR}_\lambda(\gamma)R(\gamma)/m}{1-R(\gamma)/m} \rightarrow \\ &\rightarrow \frac{1-\pi_0 f(\lambda)-\gamma\pi_0-\pi_1+\beta\pi_1+\gamma\pi_0 f(\lambda)}{Pr(p>\gamma)} = \\ &= \frac{\pi_0(1-\gamma)-\pi_0 f(\lambda)(1-\gamma)+\beta\pi_1}{Pr(p>\gamma)} = \frac{(1-\gamma)\pi_0[1-f(\lambda)]}{Pr(p>\gamma)} + \frac{\beta\pi_1}{Pr(p>\lambda)}\end{aligned} \tag{2.55}$$

almost surely.

□

2.5 The Trade-off between Type I Error Rate and Type II Error Rate in multiple tests

The dependence between α and β for single hypothesis testing is well documented in the literature (Lehmann and Romano, 2005). In the context of simple hypothesis testing, α is typically pre-specified at a small standard level (e.g., $\alpha = 0.05$, 0.01 or 0.001). Such a procedure controls the probability of false rejection at the desired level α but leaves the power of the test and hence the probability of a type II error to the mercy of the experiment, so β is mainly discussed when design and sample size are of concern. In Figure 2.1 the problem of how to balance α against β is shown. We can notice that when α decreases the probability of type II error increases. It seems then reasonable that the choice of test should involve β in addition to α. In this contest, there are several proposals of solution to balance the errors such that

$$\alpha \leq \beta \tag{2.56}$$

(Landenna *et al.*, 1997). This inequality is commonly justified on the grounds that the error of the first kind is of higher order of importance, and should be controlled at the prescribed level.

Similar to the trade-off between α and β in the context of single hypothesis, there is one between $pFDR$ and $pFNR$ for multiple comparisons. Untill now,

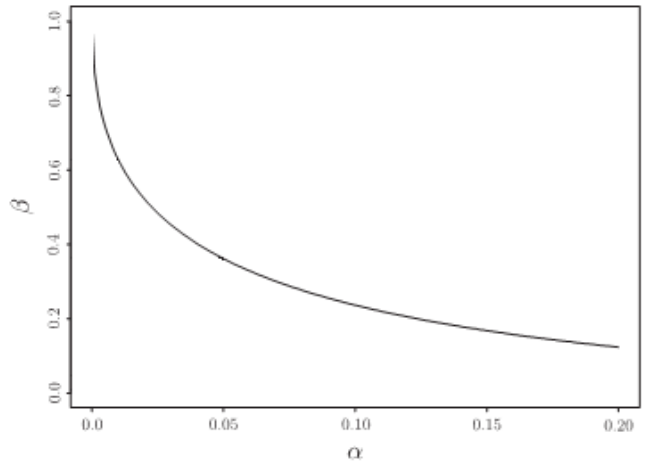

Figure 2.1: Illustration of the trade-off between type I and type II error rates

however, such standard statistical practice is still challenging for multiple hypothesis testing utilizing $pFDR$ and $pFNR$, and the choice of $pFDR$ level seems to be somewhat arbitrary from study to study.
In the following we propose a method to balance the two types of error a posteriori. Figure 2.2 shows the trade-off between $pFDR$ and $pFNR$. We can notice that the control of the $pFDR$ alone is not sufficient, in analogy with the case of the single hypothesis testing described before, if we reduce $pFDR$ there is an increase of $pFNR$. The rule to balance the two kinds of error is to look at graph itself, considering the set of pairs $(pFDR(\gamma), pFNR(\gamma))$ and choosing a suitable pair such that:

$$pFDR \leq pFNR \tag{2.57}$$

in analogy with the (2.56). We can then increase the $pFDR$ until when the consequent decrease of $pFNR$ is considered adequate. In this way we can derive the corresponding value γ for the rejection region.
In the context of network analysis, the inequality (2.57) is due to the fact that the error of putting a wrong link in the network is more serious than the error of omitting an actual link. We can notice that there is not a unique pair of $(pFDR(\gamma), pFNR(\gamma))$ satisfying the (2.57). For instance, a possible

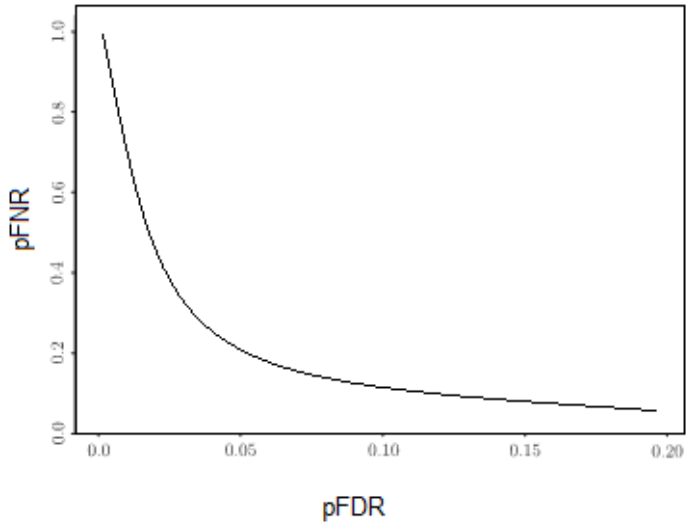

Figure 2.2: Illustration of the dependence between $pFDR$ and $pFNR$ for multiple hypothesis testing

way to choose one pair of values, according to Lehmann (1958), is to assure a pre-fixed relation between the errors $pFNR = r(pFDR)$ where $r = \frac{pFNR}{pFDR} \geq 1$ for the (2.57).

2.6 Interval estimation of pFDR, pFNR and Pwa

In the following section we show some results with the proposed estimators for $pFDR$, $pFNR$ and Pwa seen previously. In particular we derive the confidence intervals for them using the Percentile Bootstrap method, and we evaluate their coverage probability.

2.6.1 Simulation Studies

For the sake of simplicity, we simulate multiple tests on correlation coefficient,introduced in section 2.1, using its normal asymptotic distribution.

We consider the $\{n \times n\}$ correlation matrix for each network, where n is the number of nodes, and it has been fixed to 90. We performed $m = \binom{n}{2} = 4005$ hypothesis tests of $\mu = 0$ versus $\mu = 2$ for independent random variables $Z_i \sim N(\mu, 1)$, $i = 1, \ldots, 4005$, over B iterations ($b = 1, \ldots, 500$). The null hypothesis for each test is that $\mu = 0$, so the proportion of $Z_i \sim N(0,1)$ is set to π_0; hence the proportion of the alternative distribution $N(2,1)$ is π_1 (Storey, 2002). For each test the p-value is defined as $p_{ib} = Pr\{N(0,1) \geq z_i\}$ for the b-th iteration, where z_i is the observed value of Z_i.
The algorithm to estimate the FDR and FNR can be summarized through the following steps:

1. for the $m = m_0 + m_1$ hypothesis tests (m_0 true null hypotheses and m_1 false null hypotheses), calculate their respective p-values $p_1, \ldots, p_m$;

2. for each iteration b calculate $Q_b = \frac{V_b}{R_b}$, $F_b = \frac{m - m_0 - (R_b - V_b)}{m - R_b}$ and $Pw = \frac{(1-Q_b)R_b}{(1-\pi_0)m}$ where $V_b = \#\{p_{ib} > \gamma\}$ and $R_b = \#\{p_{ib} \leq \gamma\}$ with rejection region of interest $[0, \gamma]$ and $\pi_0 = m_0/m$;

3. calculate $pFDR = E[Q]$ as $\frac{1}{B}\sum_{b=1}^{B} Q_b$, $pFNR = E[F]$ as $\frac{1}{B}\sum_{b=1}^{B} F_b$ and Pwa as $\frac{1}{B}\sum_{b=1}^{B} Pw_b$ where Q, F and Pw are calculated for each iteration $b = 1, \ldots, B$;

4. estimate π_0 and $Pr(P \leq \gamma)$ by

$$\widehat{\pi_0}(\lambda) = \frac{W(\lambda)}{(1-\lambda)m}$$

and

$$\widehat{Pr}(P \leq \gamma) = \frac{R(\gamma)}{m},$$

where $R(\gamma) = \#\{p_{ib} \leq \gamma\}$ and $W(\lambda) = \#\{p_{ib} > \lambda\}$;

5. for any rejection region of interest, estimate $pFDR(\gamma)$ by

$$\widehat{pFDR}_\lambda(\gamma) = \frac{\hat{\pi_0}(\lambda)(\gamma)}{\widehat{Pr}(P \leq \gamma)[1 - (1-\gamma)^m]},$$

$pFNR(\gamma)$ by

$$\widehat{pFNR}_\lambda(\gamma) = \frac{(1-\hat{\pi}_0)m - (1-\widehat{pFDR})R}{m-R}$$

and $Pwa(\gamma)$ by

$$\widehat{Pwa}_\lambda(\gamma) = \frac{(1-\hat{pFDR})R}{(1-\hat{\pi}_0)m}$$

for each λ in $L = \{0, 0.05, \ldots, 0.95\}$;

6. for each $\lambda \in L$ estimate the mean-square errors as

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

$$\widehat{MSE_{pFNR}}(\lambda) = \frac{1}{B}\sum_{b=1}^{B}\left[\widehat{pFNR}_\lambda^{*b}(\gamma) - pFNR(\gamma)\right]^2,$$

and

$$\widehat{MSE_{Pwa}}(\lambda) = \frac{1}{B}\sum_{b=1}^{B}\left[\widehat{Pwa}_\lambda^{*b}(\gamma) - Pwa(\gamma)\right]^2.$$

Note that in this case we have the true value of $pFDR$, $pFNR$ and Pwa, instead, in real situations, we need to use a plug-in for them as described in section 2.3.3;

7. set $\hat{\lambda_1} = \text{argmin}_{\lambda\in L}\{\widehat{MSE_{pFDR}}(\lambda)\}$, $\hat{\lambda_2} = \text{argmin}_{\lambda\in L}\{\widehat{MSE_{pFNR}}(\lambda)\}$ and $\hat{\lambda_3} = \text{argmin}_{\lambda\in L}\{\widehat{MSE_{Pwa}}(\lambda)\}$. Our overall estimate of $pFDR(\gamma)$, $pFNR(\gamma)$ and $Pwa(\gamma)$ are respectively: $\widehat{pFDR}(\gamma) = \widehat{pFDR}_{\hat{\lambda_1}}(\gamma)$, $\widehat{pFNR}(\gamma) = \widehat{pFNR}_{\hat{\lambda_2}}(\gamma)$ and $\widehat{Pwa}(\gamma) = \widehat{Pwa}_{\hat{\lambda_3}}(\gamma)$;

8. form a $1-\alpha$ upper confidence interval for $pFDR(\gamma)$ by taking the $1-\alpha$ quantile of the $\widehat{pFDR}_{\hat{\lambda_1}}^{*1}(\gamma), \ldots, \widehat{pFDR}_{\hat{\lambda_1}}^{*B}(\gamma)$ as the upper confidence bound and the same procedure has to be performed for $pFNR(\gamma)$; $Pwa(\gamma)$ needs an α-lower confidence interval;

9. repeat the steps 1-8 1000 times to evaluate the coverage probability of the confidence interval obtained for $pFDR$, $pFNR$ and Pwa for a fixed rejection region γ.

$\pi_0 = 0.62$	True Value	Estimate	MSE	CI 95%
$\tau = 0.3$ $\gamma = 7.948544e-06$				
$pFDR$	0.001523	0.061898	0.002097	[0,0.068212]
$pFNR$	0.373364	0.386439	0.000355	[0,0.388173]
Pwa	0.010266	0.006320	0.000007	[0.005936,1]
$\tau = 0.2$ $\gamma = 0.0022$				
$pFDR$	0.018778	0.018969	0.000001	[0,0.020516]
$pFNR$	0.325189	0.324503	0.000413	[0,0.340279]
Pwa	0.201324	0.199559	0.000211	[0.186782,1]
$\tau = 0.1$ $\gamma = 0.0794$				
$pFDR$	0.154537	0.164708	0.000033	[0,0.165681]
$pFNR$	0.153168	0.169471	0.000647	[0,0.173149]
Pwa	0.723407	0.729512	0.001270	[0.695118,1]

Table 2.3: Results of the simulation study from step 1 to step 8, with $m_0 = 2500$ and $m_1 = 1505$: true values of $pFDR$, $pFNR$ and Pwa, their point estimates, the MSE and the 95% confidence interval

In the Table 2.3 we show the results of the simulation study from step 1 to step 8. We have reported the true values of $pFDR$, $pFNR$ and the average power Pwa, their point estimates, the MSE and the 95% confidence interval. First we have simulated a typical brain network with $m = 4005$, $m_0 = 2500$, $m_1 = 1505$ and $\pi_0 = 0.62$. We have furthermore calculated these values for different γ and associated the correspondent threshold τ on correlation coefficient. Besides we have evaluated the coverage probability of the confidence intervals, as we have seen in step 9 of the simulation, and we found that this is always equal to 100%. As we can see from the last column of Table 2.3, the confidence intervals at 95% level contain the true values of interest. We can notice that this result of over coverage is due to the conservative nature of the procedure considered. We found similar results with other values of m_0 and m_1. One example is shown in Table 2.4 with $m_0 = 1505$ and $m_1 = 2500$.

$\pi_0 = 0.37$	True Value	Estimate	MSE	CI 95%
$\tau = 0.3$ $\gamma = 7.948544e-06$				
$pFDR$	0.001042	0.014134	0.000224	[0,0.023029]
$pFNR$	0.621811	0.623932	0.000245	[0,0.635131]
Pwa	0.010212	0.010794	0.000004	[0.006721,1]
$\tau = 0.2$ $\gamma = 0.0022$				
$pFDR$	0.007023	0.007014	0.000000	[0,0.007591]
$pFNR$	0.571110	0.576322	0.000333	[0,0.585228]
Pwa	0.200224	0.182890	0.000115	[0.187253,1]
$\tau = 0.1$ $\gamma = 0.0794$				
$pFDR$	0.062954	0.062217	0.000006	[0,0.069030]
$pFNR$	0.333373	0.342642	0.000817	[0,0.346254]
Pwa	0.723172	0.713896	0.000343	[0.710298,1]

Table 2.4: Results of the simulation study from step 1 to step 8, with m_0 = 1505 and m_1 = 2500:true values of $pFDR$, $pFNR$ and Pwa, their point estimates, the MSE and the 95% confidence interval

2.7 Applications to brain networks

The aim of this section is to apply the proposed estimates for $pFDR$, $pFNR$ and Pwa to real data and to show how the methods seen above can be useful in the application introduced in the previous chapter. First, we will calculate the estimates in the context of a real network obtained from fMRI data and we will see how to choose the threshold τ on correlation coefficient that provides a balance between $pFDR$ and $pFNR$. Second, we will go back to the application of the previous chapter and we will introduce another interesting application of $pFDR$ in the context of local network analysis. In particular, we will use q-values to create a ranking of the most central nodes in the networks. Third we will investigate the differences on the inter-regional connectivity between the control group and the bvFTD group. Here we will use the multiple hypothesis testing approach to create a group differences network. Finally we will propose a method to construct a group-based representative network that allows determining the type of dynamic information transfer of a brain network set.

	Estimate	MSE	CI 95%
$\tau = 0.3$			
$pFDR$	0.000135	0.000001	[0,0.000143]
$pFNR$	0.497559	0.000117	[0,0.516438]
Pwa	0.613299	0.000061	[0.600069,1]
$\tau = 0.2$			
$pFDR$	0.000882	0.000000	[0,0.000940]
$pFNR$	0.275375	0.000185	[0,0.296598]
Pwa	0.857937	0.000047	[0.846642,1]
$\tau = 0.1$			
$pFDR$	0.026522	0.000001	[0,0.028061]
$pFNR$	0.025096	0.000041	[0,0.037872]
Pwa	0.991938	0.000004	[0.988121,1]

Table 2.5: Results of the case study:the estimates of $pFDR$, $pFNR$ and Pwa, the MSE and the 95% confidence interval, calculated on a healthy subject's network

2.7.1 Point and interval estimation of pFDR, pFNR and Pwa

In Table 2.5 the estimates of $pFDR$, $pFNR$ and Pwa, the MSE and the 95% confidence interval, calculated on a healthy subject's network, are reported for some values of the threshold τ. In Figure 2.3, Figure 2.4 and Figure 2.5 plots of $\widehat{pFDR}$, $\widehat{pFNR}$ and $\widehat{Pwa}$ are respectively shown over a range of threshold values, τ, which is the threshold applied on the correlation coefficient in order to binarize the correlation matrix and it is related to the rejection region of the test, γ, as we have seen in (2.2) and (2.1) . In Figure 2.3 and in Figure 2.5 we can see that the $pFDR$ and the average power estimated decrease with the threshold and consequently they increase with γ, because τ and γ are inversely related. In Figure 2.4 we can note the opposite behavior for the $pFNR$ estimated. This allows us to draw the trade-off between the two types of error for real networks relative to healthy subjects or patients, as illustrated in Figure 2.6. So, we can balance the two kinds of error using the rule proposed in section 2.5, considering the set of pairs $(p\widehat{FDR}(\gamma), p\widehat{FNR}(\gamma))$ and choosing a suitable pair satisfying the

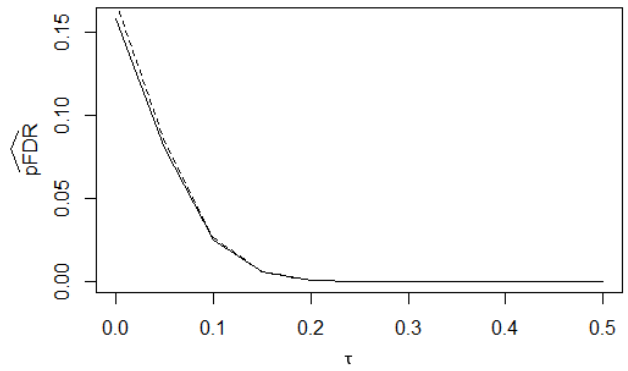

Figure 2.3: Plot of $\widehat{pFDR}$ over a range of τ values, with the respective 95% upper confidence bound (dashed line)

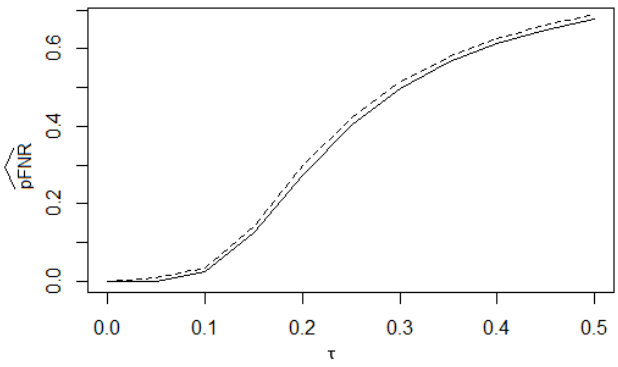

Figure 2.4: Plot of $\widehat{pFNR}$ over a range of τ values, with the respective 95% upper confidence bound (dashed line)

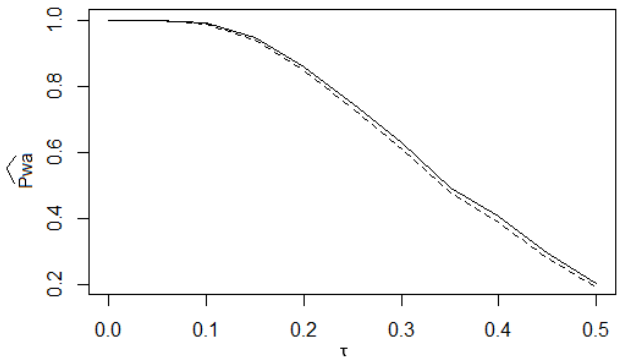

Figure 2.5: Plot of $\widehat{Pwa}$ over a range of τ values, with the respective 95% lower confidence bound (dashed line)

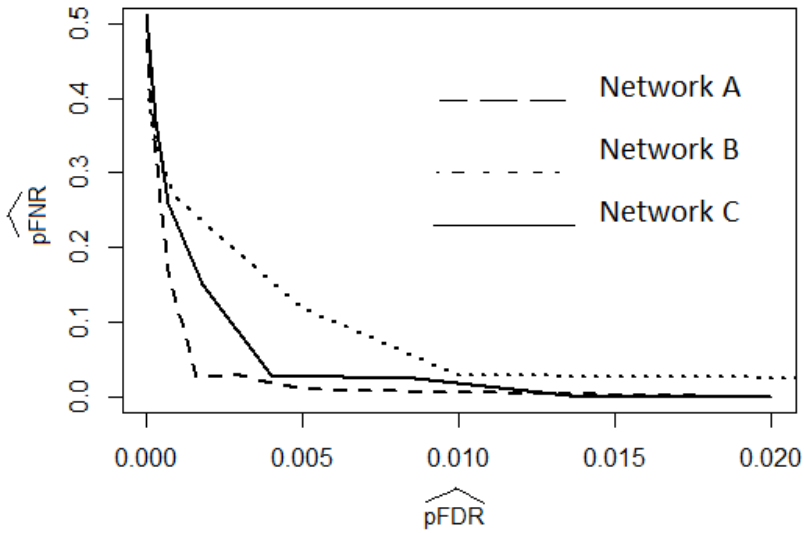

Figure 2.6: Trade-off between $\widehat{pFDR}$ and $\widehat{pFNR}$ for three networks of healthy subjects

	Estimate	MSE	CI 95%
$\tau = 0.14$			
$pFDR$	0.008	0.000001	[0,0.008426]
$pFNR$	0.08	0.00006	[0,0.10]
Pwa	0.96	0.000013	[0.955852,1]

Table 2.6: Example of balancing threshold (τ = 0.14): the estimates of $pFDR$, $pFNR$ and Pwa, the MSE and the 95% confidence interval, calculated on a healthy subject's network

(2.57). In this case we can consider τ = 0.14 as a balancing threshold with respect to $pFDR$ and $pFNR$ (Table 2.6). In the previous chapter we have seen that a crucial issue is the choice of threshold used to generate the adjacency matrix from the correlation matrix. We have seen that, in previous study, there was no unique way to choose τ, and, as a preliminary analysis, one strategy adopted to solve this problem was to examine several possible network configurations for a range of values of τ, exploring the consistency of the results over this range. In this chapter, we have dealt with this problem with the multiple hypothesis testing approach This allows us not only to associate a measure of the errors in terms of the $pFDR$ and $pFNR$ over this range of values for τ but also to choose an optimal value of τ itself using the balance rule mentioned above.

2.7.2 Local network analysis

Ranking of the nodes

Regional integrated network parameters X_{nod} (see section 1.4.3) of $Deg(i)$, $Bet(i)$ and $Eff(i)$ from all 90 examined cortical regions (i.e., including both hub and non-hub brain regions) is compared between groups by using the Wilcoxon-Mann-Whitney non-parametric test. Between-groups comparison of regional network metrics on several brain regions involves the issue of multiple comparisons. We solve this problem controlling for the $pFDR$. This approach allows us not only to control for multiple comparisons, but also to rank brain regions according to their importance in explaining differences

Brain Region	Healthy controls X_{nod}	Patients X_{nod}	q-value
Superior temporal pole R	15.76	12.98	0.0087
Anterior cingulate cortex L	16.37	13.36	0.0087
Superior parietal lobe R	16.44	13.87	0.0087
Middle occipital gyrus R	16.81	14.43	0.0087
Superior temporal pole L	14.89	12.34	0.0087
Superior parietal lobe L	16.56	14.21	0.0087
Anterior cingulate cortex L	16.16	13.71	0.0087
Inferior temporal gyrus L	16.66	14.71	0.0105
Supplementary motor area R	16.32	14.15	0.0105
Insula L	15.81	14.07	0.0105
Caudate nucleus L	14.7	11.86	0.0105
Superior occipital gyrus L	16.5	14.45	0.0105
Superior orbital frontal gyrus L	15.74	12.73	0.0105

Table 2.7: Integrated nodal parameter (X_{nod}) for degree ranked by q-values (R: right hemisphere, L: left hemisphere)

between groups. Ranking of significance is measured by the q-value seen in section 2.3.2.

For instance the results of the comparison of the nodal degree between healthy controls and patients with the behavioural variant of frontotemporal dementia (bvFTD) are shown in Table 2.7. Correction for multiple comparisons is conducted by controlling for the positive false discovery rate ($pFDR$). Brain regions were ranked according to their importance in explaining group differences by means of the q-value. So the brain regions of Table 2.7 can be considered different between controls and patients at a $pFDR$ level given by $\hat{q} = 0.0105$. The same analysis can be done with the X_{nod} of betweennes centrality and local efficiency.

2.7.3 Inter-regional correlation analysis and group differences network construction

In order to investigate differences in the brain functional organization between controls and patients, we study the inter-regional connectivity, similarly to Supekar *et al.* (2008). For this purpose, as explained in details in

Liang *et al.* (2006), we examined regional connectivity, looking at the correlation matrices of the two groups. The Pearson's correlation coefficient represent the strength of the connectivity between pairs of brain regions. The correlation coefficient values of 4005 pairs of anatomical regions are first z-normalized and then compared between the two subject groups using a one-tailed T-test. The correlation coefficients are transformed to z values using Fisher ρ-to-z transformation [2] to improve normality (Fisher, 1915). One-tailed T-test is used to test if there is a decrease of regional connectivity in bvFTD group, with $H_0 : \mu_{ij}^{controls} \leq \mu_{ij}^{patients}$ and $H_1 : \mu_{ij}^{controls} > \mu_{ij}^{patients}$, where $\mu_{ij}^{controls}$ is the mean value of the z-normalized ρ_{ij} for controls and $\mu_{ij}^{patients}$ is the mean value of the z-normalized ρ_{ij} for patients. This is another example of a multiple comparisons problem similar to those we have seen above, so it is useful to control for the $pFDR$.
The correlation values (z-normalized) of 90 pairs of anatomical regions are significantly lower in the bfFTD group as compared to the control group with $pFDR < 0.01$. In Table 2.8 we show the first connections between node i and node j among the 90 pairs, ranked by q-values.

In Figure 2.7 a group differences network is shown illustrating the results of the previous analysis, where the balls represent the nodes of the ranking in Table 2.8. Whereas the links represent the inter-regional connectivity lost in bvFTD patients. We propose this new kind of network showing the significant differences in the inter-regional connectivity between healthy controls and bvFTD patients.

2.7.4 Group representative network construction

Despite the utility of network science in providing insight into the infrastructural properties of a given subject's brain, understanding these properties in a group of subjects still presents challenges and necessitates the development of comparison tools to focus on changes in complex brain function across different cognitive and disease states (Simpson *et al.*, 2012). A group-based

[2] The Fisher variance-stabilizing transformation is defined by $z = \frac{1}{2} \ln \frac{1+\rho}{1-\rho} = \text{artanh}(\rho)$ where 'ln' is the natural logarithm function and 'artanh' is the inverse hyperbolic tangent.

Node i	Node j	q-value
Insula R	Globus pallidus R	0.0008
Anterior cingulate cortex R	Middle occipital gyrus R	0.0019
Anterior cingulate cortex R	Middle temporal pole L	0.0019
Insula L	Superior temporal pole R	0.0019
Superior occipital gyrus L	Paracentral lobule L	0.0019
Superior occipital gyrus R	Paracentral lobule L	0.0019
Medial superior frontal gyrus L	Middle temporal pole L	0.002
Insula R	Superior temporal pole R	0.002
Supplementary motor area R	Superior temporal gyrus L	0.002
Postcentral R	Temporal Sup L	0.0029
Anterior cingulate cortex L	Middle occipital gyrus R	0.0031
Medial superior frontal gyrus R	Middle temporal pole L	0.0032
Heschl L	Rectus L	0.0034
Rectus L	Rolandic Oper L	0.004

Table 2.8: First pairs (node i and node j) of anatomical regions significantly lower in the bfFTD group as compared to the control group, ranked by q-values (R: right hemisphere, L: left hemisphere)

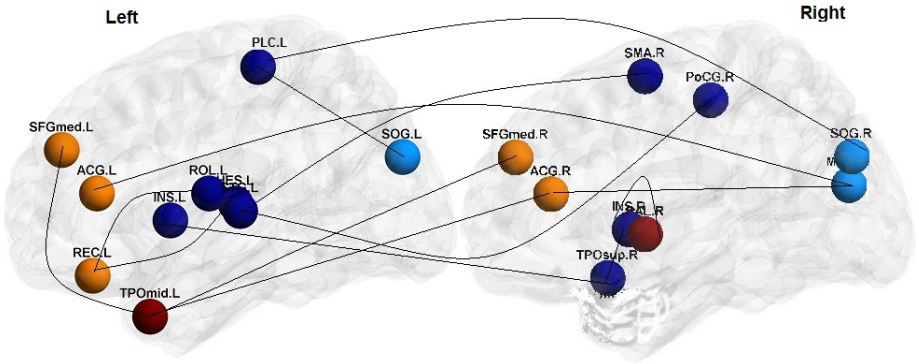

Figure 2.7: Group differences network: links represent the inter-regional connectivity lost in bvFTD patients

representative brain connectivity network can provide a graph that typifies the complex structure of a set of brain networks and it can be useful to capture the network characteristics from a group of subjects' brain networks, serving as null networks against which other networks or network models could be compared. Creating these group-based 'representative' networks is a daunting challenge given the difficulties associated with accounting for the inter-subject variability (Van Wijk *et al.*, 2010), (Simpson *et al.*, 2012). The goal of these representative network techniques is to take a set of already constructed networks and produce a group network that typifies the topological structure of the individual graphs. Hence, these methods are independent of how the initial subject-level networks are constructed.

The first and most common method to generating a group-based representative functional network has been to take the mean of the functional connectivity matrices of the subjects in a group and threshold this group mean matrix to get a mean network (Meunier *et al.*, 2009), (Zuo *et al.*, 2011). Although this approach is intuitive and computationally straightforward, the resulting network may be influenced by one or more outlying functional connectivity values (Simpson *et al.*, 2012). Another similar approach taken by researchers is to take the median of the functional connectivity matrices of the subjects and threshold this group median matrix to get a median network (Song *et al.*, 2009). Furthermore, we propose an approach to creating group-based representative networks utilizing the Fisher z-to-ρ transformation. In this case the averaging is done across the individual entries of the matrices containing the correlation coefficients z-normalized as seen in the previous section. That is, first we calculate the mean value of the z-normalized correlation matrices and then we apply the inverse of the Fisher transformation[3]. This last approach provides more robustness to outlying connectivity values. We introduce an example to see how this method works. For the sight of simplicity we consider only the group of controls. First we consider the entries of the correlation matrices $\hat{\rho}_{ij}$ for each subject and we apply the Fisher transformation, deriving a matrix with entries the resulting z_{ij}. Then we

[3]The inverse of the Fisher transformation is defined by $\rho = \frac{exp(2z)-1}{exp(2z)+1}$=tanh($z$), where 'tanh' is the hyperbolic tangent.

	Estimate	MSE	CI 95%
$\tau = 0.14$			
$pFDR$	0.026064	0.000001	[0,0.027238]
$pFNR$	0.038643	0.000025	[0,0.045257]
Pwa	0.955707	0.000032	[0.947384,1]

Table 2.9: Example of balancing threshold ($\tau = 0.14$): the estimates of $pFDR$, $pFNR$ and Pwa, the MSE and the 95% confidence interval, calculated on the control group representative network

calculate the mean of z_{ij} over the group, $\bar{z}_{ij}$, and we apply the inverse transformation to obtain $\tilde{\rho}_{ij}$. The resulting correlation matrix with entries $\tilde{\rho}_{ij}$ can be used to construct a graph applying $\tau = 0.14$ as a balancing threshold with respect to $pFDR$ and $pFNR$ (Table 2.9). This threshold corresponds to that calculated on the single subject network in Table 2.6.
As we have found the representative network of the controls, the same procedure can be also applied to different groups of patients.

2.8 Closing Remarks

In previous works (Van Wijk *et al.*, 2010), the correlation matrices were converted to an adjacency matrix by considering an arbitrary threshold. In this chapter, a new method for the derivation of the adjacency matrix with a suitable threshold is proposed using a multiple test on correlation coefficient and offering a statistical evaluation of the error committed, considering the problem of multiple comparisons. With at least one test performed for every entry of the correlation matrix, some correction is needed to control the error rates, but standard procedures for multiple hypothesis testing (e.g., Bonferroni correction) are known to be too conservative in this context. So we have dealt with the problem of multiple comparisons using the method based on the estimation of positive false discovery rate ($pFDR$) proposed by Storey (2002). Furthermore we considered a similar measure involving false negatives (type II errors), called the positive false negatives rate ($pFNR$), we proposed new point and interval estimators for $pFNR$ and a method to

balance the two types of error. We demonstrated this approach using both simulations and functional magnetic resonance imaging data providing finite sample and large sample results for $\widehat{pFDR}$ and $\widehat{pFNR}$.

Everything that we have discussed in this section lies under the assumption that we are working with independent p-values. Even in case of dependence, this approach can be fully applied due to the fact that the effect of dependence is negligible if the number of hypothesis m is large for a large class of dependence models (Storey, 2003). Also, in more general cases, such as with dependence or in nonparametric scenarios, it is possible to implement very similar approaches (Storey and Tibshirani, 2001).

We continued the applicative study introduced in the previous chapter utilizing such balanced networks to investigate the brain network characteristics, we proposed a ranking of the most central nodes in the networks using q-values (which are the $pFDR$ analog of the p-values), we examined the differences on the inter-regional connectivity between the control group and the patient (bvFTD) group. Moreover, we considered the multiple hypothesis testing approach to create a group differences network. Finally, we proposed a method to construct a group-based representative network that allows to determine the type of connectivity of a brain network set.

Chapter 3

Network Models

In previous chapters we reported that whole-brain connectivity analysis is useful to understand how each regions of the brain interacts with each another one. The complex interactions of these regions necessitates to study the brain as a whole rather than just its individual parts. The application of network and graph theory to the brain allows these whole-brain analyses and helps uncover new insights into the function of the nervous system (Simpson *et al.*, 2011), for example after having calculated the global and local network measures it has been shown that the brain exhibits the small-world properties (see sessions 1.3.4 and 1.4.3).
Another finding of the previous work is that network analysis might contribute to detect abnormalities of network connectivity in different brain disorders (see session 1.4).
In order to gain deeper insights into the complex neurobiological interactions and changes that occur in many neurological diseases, methods that enable assessing several properties simultaneously are needed given the statistical dependencies among network measures described in Chapter 1. In this chapter we will introduce the general theory of network model, in particular the exponential random graph models (ERGMs) provide a statistical approach to the assessment of how a set of interacting brain network features gives rise to the global structure. We will illustrate the utility of ERGMs for modeling and comparing complex whole-brain networks.

3.1 Models for Networks Graphs

A model for a network graph is a collection

$$\{P_\theta(G) : G \in \mathbf{G}; \theta \in \Theta\} \tag{3.1}$$

where $\mathbf{G}$ is a collection of possible graphs, P_θ is a probability distribution on $\mathbf{G}$, and θ is a vector of parameters in the space Θ. There are different graph models depending on the choice of P. Some approaches, for example, let $P(\cdot)$ be uniform on $\mathbf{G}$, but restrict $\mathbf{G}$ to contain only those graphs G satisfying certain properties of interest. Other approaches induce $P(\cdot)$ implicitly through the recurrent application of simple generative mechanism (Kolaczyk, 2009).

To understand the logic of these models and their utility in the analysis of the brain, it is helpful to introduce the concept of the observed network that is the network data the researcher has collected and is interested in modeling (Robins *et al.*, 2007a). The observed network is one realization from a set of possible network with similar characteristics (for example the same number of nodes): an outcome of some stochastic process. In general the stochastic process that generates the observed network is unknown and the aim of the model is to suggest a plausible hypothesis for this process. For instance, one research question may be whether in the observed network there are significantly more, or less, structural characteristics of interest than expected by chance. For example, it can be interesting to understand whether the observed network shows a strong tendency to form clusters, over the appearance of a number of clusters occurred at random. In this context this characteristic (presence of triangles) is the outcome of a connectivity process. In general, the structural characteristics in question helps shape the form of the model. A statistical model for a network on a given set of nodes assigns a probability to all possible networks on those nodes. As before a network can be represented as a graph constituted by nodes and edges. For a given model, the node set has to be fixed. The range of possible networks and their probability of occurrence under the model are represented by a probability

distribution on the set of all possible graphs with this number of nodes (see (3.1)). The observed network is a particular graph in this distribution and so it has a particular probability. The parameter values to use are not known in assigning probabilities to graphs in the distribution. The goal is to find the best values (by estimating model parameters) using the observed network. The maximum likelihood criterion is to choose parameter values in such a way that the most probable degree of clusterization is the one which occurs in the observed network. If we estimate a clustering parameter for the observed network, and if we can be confident that this parameter is positive, we may infer that there is more clusterization in the observed network than expected by chance (Kolaczyk, 2009).
In the following we will discuss the most important categories of network graph models used in the context of neuroimaging: the random networks, the small-world models and the class of exponential random graph models.

3.1.1 Random Networks

The term random networks typically refers to a model specifying a collection $\mathbf{G}$ and a uniform probability $P(\cdot)$ over a finite $\mathbf{G}$. Random graph models have been widely used for the precise analytical characterization of many of the descriptive measures introduced in Chapter 1. As previously reported, these models are used as term of comparisons to test some structural characteristics in observed network graphs.
Consider the task of estimating a given characteristic $\eta(G)$ of a network graph G, based on a sampled version of that graph, G^*. G is assumed to have been generated uniformly at random from a collection $\mathbf{G}$, prior to obtaining G^*. Inference on η should incorporate both randomness due to selection of G from $\mathbf{G}$ and randomness due to sampling G^* from G. If we are interested in $\eta(G^{obs})$, where η is a structural characteristic of the observed graph (G^{obs}) and we want to know if this property is unexpected in comparison to an appropriate frame of reference, we can use random graph models to set up such a comparison.
A collection of random graphs $\mathbf{G}$ is defined, and the value $\eta(G^{obs})$ is com-

pared to the collection of values $\{\eta(G) : G \in \mathbf{G}\}$. If $\eta(G^{obs})$ is judged to be extreme with respect to this collection, then there would be an evidence that $\eta(G^{obs})$ is unusual. Formally, a random network is used to create a reference distribution which, under the assumption of uniform likelihood of elements in $\mathbf{G}$, takes the form

$$P_{\eta,\mathbf{G}}(t) = \frac{\#\{G \in \mathbf{G} : \eta(G) \leq t\}}{|\mathbf{G}|}. \tag{3.2}$$

If $\eta(G^{obs})$ is found to be sufficiently unlikely under the distribution (3.2), this is taken as evidence against the hypothesis that G^{obs} is a uniform draw from $\mathbf{G}$.

The problem is that we are not always able to explicitly enumerate all of the elements of $\mathbf{G}$, and therefore, we cannot calculate the probabilities in (3.2) exactly. Rather, approximations to these probabilities are used. There are analytical approximations for certain comparatively simple cases or Monte Carlo simulation methods to obtain numerical approximations are employed. The classical theory of random graph models, described in a series of papers by Erdos and Renyi (1959, 1960, 1961), rests upon a simple model that places equal probability on all graphs of a given order and size. Their model specifies a collection G_{N_v,N_e} of all graphs $G = (V, E)$ with $|V| = N_v$ and $|E| = N_e$, and assigns probability $P(G) = \binom{N}{N_e}^{-1}$ to each $G \in \mathbf{G}_{N_v,Ne}$, where $N = \binom{N_v}{2}$ is the total number of distinct vertex pairs. Erdos and Renyi (1959) developed formal probabilistic results concerning the characteristics of graphs G drawn randomly from $\mathbf{G}_{N_v,Ne}$ with respect to this $P(\cdot)$. A variant of $\mathbf{G}_{N_v,Ne}$ is suggested by Gilbert (1959), where a collection $\mathbf{G}_{N_v,p}$ consists of all graphs G of order N_v that may be obtained by assigning an edge independently to each pair of distinct vertices with probability $p \in (0, 1)$. When p is a function of N_v, and $N_e \sim pN_v$, these two classes of models are essentially equivalent for large N_v (Bollobas, 2001).

3.1.2 Small-World Models

Small-World Models are designed to mimic an observed property of the real world. This type of model has been introduced by Watts and Strogatz (1998). The interest in these models is largely due to their relevance to the topic of communication: if a graph has this topology, it means that it would be able to transmit information quickly across the network. Small worlds are relevant, for example, to the spread of news, gossip, rumors; but they also have been found to be of importance in the biological field, for instance, for the spread of diseases, as in the case of human infectious diseases or for the information flow among the neurons in the brain.
In the case of the classical random graph model, we observe small distance between nodes and a very little clustering. Instead many networks in the real world display small distances between most nodes, but high levels of clustering. Watts and Strogatz (1998) create a network graph with both the properties, introduced above. They consider a generic graph G assumed to be unweighted, simple (multiple edges between the same couple of nodes are not allowed), connected (there exist at least one path connecting any couple of nodes) and sparse ($N_e << N_v(N_v - 1)/2$, i.e. only a few of the total possible number of edges $N_v(N_v - 1)/2$ exist. The degree of a graph with these properties must be small enough to be sparse but, on the other side, it must be large enough to assure that there exist at least one path connecting any couple of nodes. For a random graph these properties is satisfied if $N_e >> N_v ln N_v$ (Latora and Marchiori, 2001).
The mathematical characterization of the small-world behavior is based on the evaluation of the two quantities we have defined in Chapter 1: the characteristic path length $pl(G)$ and the clustering coefficient $cl(G)$. Small-world networks are in between regular and random networks: they are highly clustered like regular lattices [1], yet having small characteristics path lengths like random graphs. Watts and Strogatz (1998) propose a one-parameter model (the WS model) to construct a class of graphs $\mathbf{G}$ which interpolates between a regular lattice and a random graph. The WS model is a method

[1] In a regular lattice each node is linked to its four immediate neighbors

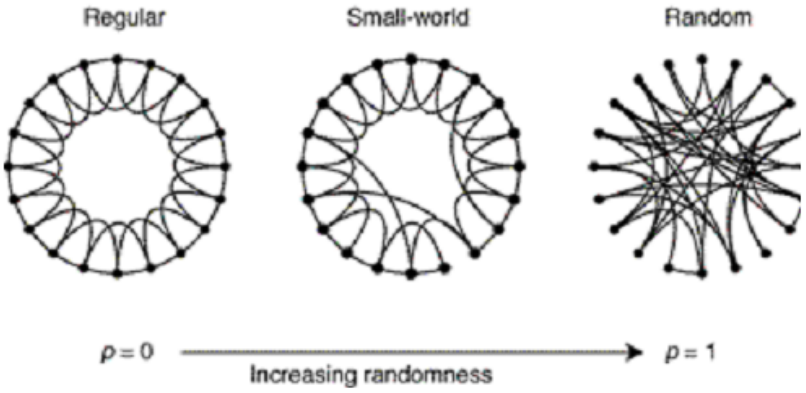

Figure 3.1: The rewiring procedure of the WS model interpolates between a regular lattice and a random graph without altering the number of nodes or edges. The regular lattice has $N_v = 20$ nodes, each connected to its 4 neighbors and a total number of edges $N_e = 40$. As the rewiring probability p increases, the network becomes increasingly disordered. For $p = 1$ a random graph is obtained

to produce a class of graphs with increasing randomness without altering the number of nodes or edges. As we can see in Figure 3.1, the WS model starts with a one-dimensional lattice with N_v vertices, N_e edges, and periodic boundary conditions. Every vertex in the lattice is connected to its k neighbors (4 in the Figure 3.1). The random rewiring procedure consists in going through each of the edges in turn and independently with some probability p rewire it. Rewiring means shifting one end of the edge to a new vertex chosen randomly with a uniform probability, avoiding multiple edges, self-connections and disconnected graphs (Latora and Marchiori, 2003). In this way it is possible to tune G in a continuous manner from a regular lattice ($p = 0$) into a random graph ($p = 1$), without altering the average number of neighbors. The behavior of $pl(G)$ and $cl(G)$ in the two limiting cases is: for the regular lattice we expect $pl(G) \sim N_v/2k$ and a relatively high clustering coefficient $cl(G) = 3/4(k-2)/(k-1)$; for the random graph, we expect $pl(G) \sim lnN_v/ln(k-1)$ and $cl(G) \sim k/N_v$. It is interesting to notice how regular and random graphs behave differently when we change the size of the system N_v. If N_v increases, keeping fixed the average number

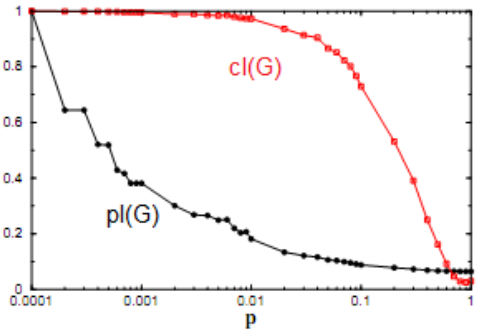

Figure 3.2: Characteristic path length $pl(G)$, and clustering coefficient $pl(G)$ for the class of graphs produced by the WS model with $N_v = 1000$ and $k = 10$. As function of p the rewiring procedure interpolates between a regular lattice ($p = 0$) and a random graph ($p = 1$), and produces the small-world behavior for p in the range 0.01-0.1

of edges per vertex, we see that for a regular graph $pl(G)$ increases with the size of the system, while for a random graph $pl(G)$ increases much slower, only logarithmically with N_v. On the other hand, the clustering coefficient $cl(G)$ does not depend on N_v for a regular lattice, while it goes to zero in large random graphs. From these two limiting cases one could argue that short $pl(G)$ is always associated with small $cl(G)$, and long $pl(G)$ with large $cl(G)$. Now we can come back to the WS model. To understand the co-existence of small characteristic path length and high clustering, typical of the small-world behavior, we can see the behavior of $pl(G)$ and $cl(G)$ as a function of the rewiring probability p for a graph with $N_v = 1000$ and $k = 10$ in Figure 3.2.

Although in the two limiting cases large $cl(G)$ is associated to large $pl(G)$ and viceversa, the numerical experiment reveals very interesting properties in the intermediate regime: only few rewired edges (small $p = 0$) are sufficient to produce a rapid drop of $pl(G)$, while $cl(G)$ is not affected and remains equal to the value for the regular lattice. In this intermediate regime the network is highly clustered like regular lattices and has small characteristic path lengths like random graphs. These networks are named small worlds in analogy with the small-world phenomenon empirically observed in social

systems more than 30 years ago by the social psychologist Milgram (1967) (Latora and Marchiori, 2003). The WS model is a way to construct networks with the characteristics of a small-world. In Chapter 1, we have applied this mathematical formalism to study the topological properties of real brain networks.

3.1.3 Exponential Random Graph Models

In network modeling, the goal is to predict the joint probability that a set of edges exists given a set of nodes in a network. In general, these edges are not independent, and that is the challenge for both model specification and estimation. ERGMs were developed to address the complex dependencies within relational data structures and provide a way to represent them (Kolaczyk, 2009). In Chapter 1, it has been shown that there are many techniques that measure properties of a network, of the nodes, or of subsets of nodes (e.g., small-worldness, centrality). These techniques are useful to describe and understand network features that might bear on particular research questions, otherwise, in the following, we explain why a well-fitted model of an observed network can be useful to improve descriptive results. A well-specified stochastic model allows us to understand the uncertainty associated with observed outcomes: we can learn about the distribution of possible outcomes for a given specification of a model, or we can estimate, for given observed data, the parameters of the hypothesized model from which the data may have been generated. Statistical models also allow inferences about whether certain network substructures are more commonly observed in the network than might be expected by chance. Besides, once we have defined a probability distribution on the set of all graphs with a fixed number of nodes, we can also draw graphs at random from the distribution according to their assigned probabilities, and we can compare the sampled graphs to the observed one on any other characteristic of interest. So, in the contest of the brain network analysis, ERGMs provide a statistically principled approach to the assessment of how a set of interacting local brain network features gives rise to the global structure and allow to explore several features simultaneously

and to understand how they interact to form the global network architecture (Simpson *et al.*, 2011). Simple examples of these features include degree distributions, triangles (number of triangles in the network) and other higher order cycles that lead to clustering. These observed network statistics are properly regarded as outcomes, and the goal of the model is to specify the process that leads to their joint distribution. The term used to indicated them is also network statistics.
ERGMs can be applied on a single subject network in order to quantify the relative significance of various network statistics, they allow asking specific questions about processes that may give rise to the network architecture. ERGMs can also be applied on a group representative network (see session 2.7.4) in order to create a network model for a particular group of subject, for example with a particular disease, and to find the network statistics having the most influential impact on overall functional brain network organization in this group. Moreover group-based network comparisons can potentially be performed by comparing the mean of the estimated model parameters among groups of subjects via hypothesis testing or classification techniques (Simpson *et al.*, 2011). A suitable test to compare two groups can be employed. It is important to note that if one were to just compare the mean of a estimated model parameter of a particular metrics among groups then the potential confounding from the other metrics included in the model would be inherently accounted for given that the estimates account for all other metrics in the model.
It is important to underline that the use of network analysis and models for comparing networks is not without challenges, the major difficulty arises from the fact that graph measures and the network statistics depend on the number of nodes and edges in the network. Van Wijk *et al.* (2010) propose a review of advantages and disadvantages of approach commonly used to compare networks, for instance fixing the threshold or the average degree.

The ERGM, also called p^* class of models, was first discussed by Frank and Strauss (1986) who viewed it as a generalization of model $p1$ introduced by Holland and Leinhardt (1981). Frank and Strauss (1986) presented the

general specification of p^* and, in particular, discussed the Markov model, a special case of p^* model. Further developments of this family of models has been given by Wasserman and Pattison (1996) and Pattison and Wasserman (1999) who showed how a Markov parametric assumption provides just one of many possible sets of parameters. Most of the initial investigations focused on nondirected, single, binary links. In literature there are a lot of works which present extensions to valued and multiple links (Pattison and Wasserman, 1999), (Koehly and Pattison, 2005). There are two different approaches to derive the model : it could be derived from the Hammersley-Clifford theorem (Besag, 1975) as a consequence of dependence graphs (Frank and Strauss, 1986) or it could be seen as a logistic regression model (Wasserman and Pattison, 1996).

Robins *et al.* (2007a) summarize the steps to follow in using the model. The first step implies a stochastic framework with a fixed node set. This means that the link is assumed to be a random variable. For i and j, distinct members of a set N_i of n nodes, there is a random variable Y_{ij} where $Y_{ij} = 1$ if there is a link from node i to node j, and where $Y_{ij} = 0$ if there is no link. y_{ij} is specified as the observed value of the variable Y_{ij} and $\mathbf{Y}$ as the matrix of all variables with $\mathbf{y}$ the matrix of observed links. We will restrict the attention to binary edges. The second steps regards the proposal of a dependence hypothesis, defining contingencies among the network variables. This hypothesis embodies the local processes that are assumed to generate the network links. For instance, edges may be assumed to be independent of each other, that is, brain regions form connections independently of their other links. This is not usually a very realistic assumption. The third step refers to the form for the model. It is important to notice that well-specified dependence assumptions imply a particular class of models (by the Hammersley and Clifford theorem). Each parameter corresponds to a configuration in the network, that is a subgraph structure for which there is a parameter in the model (Robins *et al.*, 2009). The model represents a distribution of random graphs which are assumed to be built up from the patterns represented by the configurations. Step four is the reduction of the number of parameters through homogeneity or other constraints. Finally the step five

is the estimation and the interpretation of the model parameters, the statistical basis of the model permits inferences about which configurations are important. In the following sections we discuss the steps introduced above.

Model specification

A discrete random vector $\mathbf{Z}$ is said to belong to an exponential family if its probability mass function may be expressed in the form

$$P_\theta(\mathbf{Z} = \mathbf{z}) = exp\{\theta^T \mathbf{g}(\mathbf{z}) - \psi(\theta)\} \tag{3.3}$$

where $\theta \in \mathbb{R}^p$ is a $p \times 1$ vector of parameters, $g(\cdot)$ is a p-dimensional function of $\mathbf{z}$ and $\psi(\theta)$ is a normalization term, ensuring that $P_\theta(\cdot)$ sums to one over its range. The class of discrete exponential families includes many familiar distributions, such as the Binomial, Geometric, and Poisson. In the case of continuous exponential families, where an analogous form of (3.3) holds for probability density functions, for example the Gaussian and Chi-square distributions. Exponential families all have a variety of useful algebraic and geometric properties, making this class of distributions mathematically convenient for purposes of inference and simulation (Kolaczyk, 2009). Consider $G = (V, E)$ as a random graph. Let $Y_{ij} = Y_{ji}$ be a binary random variable indicating the presence or absence of an edge $e \in E$ between two vertices i and j in V. The matrix $\mathbf{Y} = [Y_{ij}]$ is thus the (random) adjacency matrix for G. Denote by $\mathbf{y}=[y_{ij}]$ a particular realization of $\mathbf{Y}$. More precisely, an ERGM takes the form

$$P_\theta(\mathbf{Y} = \mathbf{y}) = \frac{1}{\kappa} exp\{\sum_H \theta_H g_H(\mathbf{y})\} \tag{3.4}$$

where: H is a configuration, which is defined to be a set of possible edges among a subset of the vertices in G; $g_H(\mathbf{y}) = \prod_{y_{ij} \epsilon H}(y_{ij})$ and is therefore either one if the configuration H occurs in $\mathbf{y}$,or zero, otherwise; a non-zero value for θ_H means that the Y_{ij} are dependent for all pairs of vertices i, j in H, conditional upon the rest of the graph; $\kappa = \kappa_{(\theta)}$ is a normalization

constant,

$$\kappa_\theta = \sum_{\mathbf{y}} exp\{\sum_{H} \theta_H g_H(\mathbf{y})\}. \tag{3.5}$$

The summation in (3.4) and (3.5) is over all possible configurations H.

Parameter Constraints

There are several ways in which constraints on the parameters of ERGMs may be applied. In the general form given by (3.4) the ERGM implies many parameters. In order to define a model clearly, we need to reduce the number of parameters. Some parameters need to be set to zero, equated or otherwise constrained by introducing hypothesized constraints on the values of parameters associated with larger configurations (Snijders *et al.*, 2006). Following Frank and Strauss (1986), often a homogeneity assumption is imposed by equating parameters when they refer to the same type of configuration. One of the examples of homogeneity constraint is to propose a single tendency for reciprocity across the entire network, by assuming that the reciprocity parameters for each possible reciprocated tie are all equal. Another method of applying constraints may be obtained by equating parameters for isomorphic configurations involving similar types of vertices.

Dependence Assumptions

As suggested by Frank and Strauss (1986), such a model implies a structure of dependence among the elements in $\mathbf{Y}$, so it is fundamental to define dependence structures for the random variables Y_{ij} because the entries of adjacency matrix cannot be assumed to be independent. For example, the simplest dependence assumption is the case of the Bernoulli Random Graph, for any given pair of vertices, the presence or absence of an edge between that pair is independent of the status of possible edges between any other pairs of vertices. That is, for each pair (i, j), we assume that Y_{ij} is independent of $Y_{i'j'}$, for any $(i'j') \neq (i, j)$. This assumption implies that $\theta_H = 0$ for all configurations H involving three or more vertices. In this case, the only relevant functions g_H are those of the form $g_H(\mathbf{y}) = g_{ij}(\mathbf{y}) = y_{ij}$, and the

ERGM in (3.4) reduces to

$$P_\theta(\mathbf{Y} = \mathbf{y}) = \frac{1}{\kappa} exp\{\sum_{i,j} \theta_{ij} y_{ij}\}. \tag{3.6}$$

This is another way to write that each edge i, j is present in the graph independently with probability

$$p_{ij} = \frac{exp(\theta_{ij})}{1 + exp(\theta_{ij})}. \tag{3.7}$$

This implies a model with N_v^2 parameters, which is likely far too overparameterized. In order to reduce the total number of parameters, it is common to impose an assumption of homogeneity across certain vertex pairs. For example, assuming homogeneity across all of G, the condition: $\theta_{i,j} \equiv \theta$, for all (i, j), yields

$$P_\theta(\mathbf{Y} = \mathbf{y}) = \frac{1}{\kappa} exp\{\theta N(\mathbf{y})\}, \tag{3.8}$$

where $N(\mathbf{y}) = \sum_{i,j} y_{i,j} = N_e$ is the number of edges in the graph and the parameter θ is related to the probability of a edge being observed. The parameter θ is called the edge or density parameter. The (3.8) is, furthermore, equivalent to the Erdos and Renyi model with probability $p = \frac{exp(\theta)}{1+exp(\theta)}$. There are other possibilities for imposing homogeneity. Suppose that vertices are known a priori to fall within either of two sets, say S_1 and S_2. If we impose homogeneity within and between sets, we arrive at a model of the form

$$P_\theta(\mathbf{Y} = \mathbf{y}) = \frac{1}{\kappa} exp\{\theta_{11} N_{11}(\mathbf{y}) + \theta_{12} N_{12}(\mathbf{y}) + \theta_{22} N_{22}(\mathbf{y})\}, \tag{3.9}$$

where $N_{11}(\mathbf{y})$ and $N_{22}(\mathbf{y})$ are the number of edges within sets S_1 and S_2, respectively, and $N_{12}(y)$ is the number of edges between S_1 and S_2. Unfortunately, assumptions of complete independence among possible edges are untenable in practice. Furthermore there are situations that indicate that Bernoulli-like random graphs lack the ability to reproduce many of the most basic structural characteristics observed in most real-world networks. But certain simple conditional independence assumptions may be used profitably

to create a much richer class of models.

Frank and Strauss (1986) introduced the notion of Markov dependence for network graph models, which specifies that two possible edges are dependent whenever they share a vertex, conditional on all other possible edges. That is, the presence or absence of i, j in the graph will depend upon that of i, k, for a given $k = j$, even given information on the status of all other possible edges in the network. A random graph G arising under Markov dependence conditions is called a Markov graph. Under a suitable assumption of homogeneity, using the Hammersley-Clifford theorem, Frank and Strauss (1986) shows that G is a Markov graph if and only if $P_\theta(\cdot)$ may be expressed as

$$P_\theta(\mathbf{Y} = \mathbf{y}) = \frac{1}{\kappa} exp\{\sum_{k=1}^{N_v - 1} \theta_k S_k(\mathbf{y}) + \theta_\tau T(\mathbf{y})\}, \tag{3.10}$$

where $S_1(\mathbf{y}) = N_e$ is the number of edges, $S_k(\mathbf{y})$ is the number of k-stars (a k-star is a tree, with one vertex of degree k, and k vertices of degree 1), for $2 \leq k \leq N_v - 1$, and $T(\mathbf{y})$ is the number of triangles. Note that the statistics S_k in (3.10), and also T, can be expected to be correlated. For example, more edges in G clearly allows for the possibility of more k-stars of certain orders. Similarly, more k-stars, for a given k, means more k'-stars, for $k' < k$. In this sense, we may view the θ_k, for increasingly larger k, as successively higher-order effects in the model. The inclusion of lower-order effects $\theta_{k'}$ means that θ_k is the effect due to k-stars, adjusted for the quantity of lower-order stars. Additionally, more k-stars means the potential for more triangles, and the value θ_τ is therefore the effect of triangles, adjusted for the levels of all k-stars. We can see from the nature of the statistics S_k and T that the transition from the independence assumptions underlying Bernoulli models to Markov dependence results in a model explicitly parameterized to account for some effects of transitivity, something lacking in the models introduced previously. In using Markov graph models, it has traditionally been common practice to include star counts S_k no higher than $k = 2$, or at most $k = 3$, by setting $\theta_4 = \ldots = \theta_{N_v - 1} = 0$. But experience has shown this practice to frequently produce models that fit quite poorly to real data. Investigation of this phenomena has found it to be related to the issue of

model degeneracy that we will discuss in the following. Otherwise, if we include a sufficiently large number of higher order terms, there would be a problem of model fitting. These problems are greater for network with more nodes (Hunter, 2007).

A solution to this dilemma, proposed by Snijders *et al.* (2006) , is to impose a parametric constraint of the form $\theta_k \propto (-1)^k \lambda^{2-k}$ upon the star parameters, for all $k \geq 2$, for some $\lambda \geq 1$. This tactic has the effect of combining all of the k-star statistics $S_k(\mathbf{y})$ in (3.1), for $k \geq 2$, into a single alternating k-star statistic of the form

$$AKS_\lambda(\mathbf{y}) = \sum_{k=2}^{N_v - 1} (-1)^k \frac{S_k(\mathbf{y})}{\lambda^{k-2}}, \tag{3.11}$$

and weighting that statistic by a single parameter θ_{AKS} that takes into account the star effects of all orders simultaneously. It may be shown by Hunter and Handcock (2006) and by Hunter (2007) that the statistic $AKS_\lambda(\mathbf{y})$ is a linear function of the geometrically weighted degree count, defined as:

$$GWD_\gamma(\mathbf{y}) = \sum_{d=0}^{N_v - 1} (e)^{-\gamma d} N_d(\mathbf{y}), \tag{3.12}$$

where $N_d(\mathbf{y})$ is the number of vertices of degree d and $\gamma > 0$ is related to λ through the expression $\gamma = log[\frac{\lambda}{(\lambda-1)}]$. So this approach in a sense attempts to model the degree distribution, with choice of γ influencing the extent to which higher-degree vertices are likely to occur in the graph G.A positive estimate is evidence that the network contains a skewed degree distribution with some higher degree nodes, while a negative parameter suggests that high degree nodes are improbable, with a smaller variance between the degrees. In other words, this parameter provides information about the spread of the degree distribution. Moreover, a positive parameter suggests a preference for connections between a larger number of low degree nodes and a smaller number of higher degree nodes.

Besides Snijders *et al.* (2006) discusses a number of other similar statistics, including a generalization of triadic structures based on alternating sums of

k-triangles, which takes the form

$$AKT_\lambda(\mathbf{y}) = 3T_1 + \sum_{k=2}^{N_v-2}(-1)^{k+1}\frac{T_k(\mathbf{y})}{\lambda^{k-1}}. \tag{3.13}$$

Here T_k is the number of k-triangles, where a k-triangle is defined to be a set of k individual triangles sharing a common base. There is an equivalent formulation of this definition, in terms of geometrically weighted counts of the neighbors common to adjacent vertices, called the edgewise shared partner parameters (GWESP) (Robins *et al.*, 2009) (Hunter, 2007) .
A positive alternating k-triangle effect and a negative alternating k-star effect together suggest a segmented network of multiple (but small) dense regions connected by low density paths (Robins *et al.*, 2007a).
Another similar statistics the alternating k-two-path effect (or alternating 2-path effect) is intended to assist with modeling localized multiple connectivity in the network. By this we mean tendencies for redundant short paths between pairs of nodes. A k-two path is a precursor to the transitivity construct represented by a k-triangle: the configuration is identical to a k-triangle except that there is no requirement for an edge at the base of the k-triangle. It represents k independent two-paths between two nodes.
Let the number of k-two paths in the graph be U_k. Then the alternating 2-path statistic is:

$$AK2P_\lambda(\mathbf{y}) = U_1 - \frac{2U_2}{\lambda} + \sum_{k=3}^{N_v-2}(-1)^{k-1}\frac{U_k(\mathbf{y})}{\lambda^{k-1}}. \tag{3.14}$$

If the k-triangle effect is positive in the presence of the k-2-path parameter, then there is evidence that transitivity in this network tends to occur because of the completion of the bases of k-triangles, rather than completion of the sides. In other words, multiple connectivity in the form of independent 2-paths tends to lead to network closure. There is also an equivalent interpretation of the parameter, the dyadwise shared partner parameter (GWDSP) (Robins *et al.*, 2009) (Hunter, 2007). Models involving the alternating k-star, alternating k-triangle and alternating k-two path statistics may be viewed as curved exponential family models (Hunter, 2007).

Estimation method

The goal of applying a ERG model for a network consists in describing the global features by a low number of local structures. To do it we need to estimate the unknown parameters θ of a set of configurations $g(y)$ and then interpret them. In standard settings, with independent and identically distributed realizations, exponential family models are fit using the method of maximum likelihood, and the resulting parameter estimates $\hat{\theta}$ are accompanied by asymptotically confidence intervals and test statistics. In the context of the ERGMs the maximum likelihood estimators (MLEs) $\hat{\theta}_H$ of the parameters θ_H are well defined, Handcock (2003a) presents conditions for the existence and uniqueness of the MLE for network models and he indicates that many properties of MLE can be derived from the statistical exponential family theory (Barndorff-Nielsen, 1978). The problem is that the calculation of MLE is non-trivial and that an asymptotic theory for confidence intervals and testing, taking into account the highly dependent nature of observations in a network graph, has yet to be established (Kolaczyk, 2009).
Consider the general definition of an ERGM in (3.4). The MLE for the vector $\theta = \theta_H$ is defined as $\hat{\theta} = \text{argmax}_\theta l(\theta)$, where $l(\theta)$ is the log-likelihood, which has the following form common to exponential families,

$$l(\theta) = \theta^T \mathbf{g}(\mathbf{y}) - \psi(\theta). \tag{3.15}$$

Here $\mathbf{g}$ denotes the vector of functions g_H and $\psi(\theta) = \log k(\theta)$. Alternatively, taking derivatives on each side

$$\frac{\partial l(\theta)}{\partial \theta} = \mathbf{g}(\mathbf{y}) - \frac{\partial \psi(\theta)}{\partial \theta}$$

and using the fact that $E_{\hat{\theta}}[\mathbf{g}(\mathbf{Y})] = \frac{\partial \psi(\theta)}{\partial \theta}$, the MLE can also be expressed as the solution to the system of equations

$$\frac{\partial \psi(\theta)}{\partial \theta} = \mathbf{g}(\mathbf{y})$$

which is equivalent to

$$E_{\hat{\theta}}[\mathbf{g}(\mathbf{Y})] = \mathbf{g}(\mathbf{y}). \tag{3.16}$$

The problem is that the function $\psi(\theta)$ cannot be evaluated explicitly, as it involves the summation in (3.5) over $2^{\binom{N_v}{2}}$ possible choices of $\mathbf{y}$, for each candidate θ. Therefore, it is necessary to use numerical methods to compute approximate values for $\hat{\theta}$. Two Monte Carlo approaches are commonly used for this purpose, one based on the stochastic approximation of the optimization of the log-likelihood in (3.15), and the other using a method for the stochastic approximation to solutions of systems of equations, applied to those in (3.16) (Robins *et al.*, 2007a). The first method, Markov Chain Monte Carlo (MCMC) maximum likelihood estimation, derives from fundamental work of Geyer and Thompson (1992). Note that optimization of the log-likelihood in (3.15) is equivalent to optimization of the logarithm of the likelihood ratio

$$r(\theta, \theta^{(0)}) = l(\theta) - l(\theta^{(0)}) = (\theta - \theta^{(0)})^T \mathbf{g}(\mathbf{y}) - [\psi(\theta) - \psi(\theta^{(0)})], \tag{3.17}$$

for arbitrary, fixed $\theta^{(0)}$. Furthermore, we can note that:

$$\begin{aligned} &exp\{\psi(\theta) - \psi(\theta^{(0)})\} = \\ &= \textstyle\sum_{\mathbf{y}} exp\{(\theta - \theta^{(0)})^T \mathbf{g}(\mathbf{y})\} \left(\frac{exp\{(\theta^{(0)})^T\}\mathbf{g}(\mathbf{y})}{exp\{\psi(\theta^{(0)})\}} \right) = \\ &= E_{\theta^{(0)}} \Big[exp\{(\theta - \theta^{(0)})^T\} \mathbf{g}(\mathbf{Y}) \Big]. \end{aligned} \tag{3.18}$$

Therefore, an approximation to the term $\psi(\theta) - \psi(\theta^{(0)})$ in (3.17) can be produced by:

- generating a Markov chain Monte Carlo sample $\mathbf{Y_1}, \ldots, \mathbf{Y_n}$ from the ERGM (3.4), under $\theta^{(0)}$;
- approximating the expectation in (3.18) by the corresponding average based on this sample;
- taking a logarithm of that average.

The resulting approximation to the log-likelihood ratio will converge to its target as n tends to infinity, and hence the optimum of this approximate log-likelihood ratio will approximate the MLE $\hat{\theta}$ (Hunter and Handcock, 2006). The second method utilizes the algorithm of Robbins and Monro (1951), which may be viewed as a stochastic version of the Newton-Raphson algorithm. Given a random vector Z, with distribution parameterized by a vector θ, the Robbins-Monro algorithm allows for the solution in θ of the system of equations $E_\theta[\mathbf{Z}] = 0$, through a sequence of iterations of the form

$$\hat{\theta}^{(i+1)} = \hat{\theta}^i - a_i D_i^{-1} \mathbf{Z_i}. \tag{3.19}$$

Here the $\mathbf{Z_i}$ are a sequence of random vectors for which the distribution of $\mathbf{Z_i}$, conditional on $\mathbf{Z_1}, \ldots, \mathbf{Z_{I-1}}$, is that of $\mathbf{Z}$ at $\theta = \hat{\theta}^{(i)}$; the a_i are a sequence of positive numbers tending to zero; and D_i is a matrix playing the role of the Hessian in the traditional Newton-Raphson algorithm. In the case of $a_i = \frac{1}{i}$ and $\mathbf{Z}$ distributed according to an exponential family, as in (3.3), the optimal choice of D_i is $COV_\theta(\mathbf{Z})$, where this quantity is usually approximated. To estimate the MLE $\hat{\theta}$, that solves the system of equations in (3.16), we set $\mathbf{Z} = \mathbf{g(Y)} - \mathbf{g(y)}$, where $\mathbf{Y}$ is distributed according to (3.4) (Snijders, 2002). For both of these methods of approximating the MLE, it is necessary to simulate draws of $\mathbf{Y}$ from the ERGM in (3.4). One natural approach to this task is to use the Gibbs sampler, a general Markov chain Monte Carlo method for simulating from the joint distribution of a vector $\mathbf{Z}$ that utilizes only the univariate conditional distributions of each element given all of the others. For ERGMs, such conditional distributions have a particularly simple form. Writing $\mathbf{Y}_{(-ij)}$ to be all of the elements of $\mathbf{Y}$ except $\mathbf{Y}_{(ij)}$, the distribution of $Y_{(ij)}$ conditional on $\mathbf{Y}_{(-ij)}$ is Bernoulli and satisfies the expression

$$\log\left[\frac{P_\theta(Y_{ij} = 1 | \mathbf{Y}_{(-ij)} = \mathbf{y}_{(-ij)})}{P_\theta(Y_{ij} = 0 | \mathbf{Y}_{(-ij)} = \mathbf{y}_{(-ij)})}\right] = \theta^T \Delta_{ij}(\mathbf{y}), \tag{3.20}$$

where $\Delta_{ij}(\mathbf{y})$ is the change statistic, denoting the difference between $\mathbf{g(y)}$ when $y_{ij} = 1$ and when $y_{ij} = 0$, which may be calculated in an efficient manner. Various other MCMC methods are also possible (Snijders, 2002). A

disadvantage of both of the methods above is their computationally intensive nature. An alternative is to estimate θ by maximizing not the actual log-likelihood (3.4), but rather the log-pseudo-likelihood

$$\sum_{\{i,j\}} \log P_\theta(Y_{ij} = 1 | \mathbf{Y}_{(-ij)} = \mathbf{y}_{(-ij)}). \tag{3.21}$$

This approach, originally proposed by Besag (1975) in the context of spatial data analysis, and adapted for ERGMs by Strauss and Ikeda (1990), will work best when dependencies among the elements of $\mathbf{Y}$ are relatively weak. Unfortunately, in many network contexts this is not likely to be the case. Nevertheless, pseudo-likelihood estimation is a method for obtaining at least some rough sense of the value of $\hat{\theta}$ (Kolaczyk, 2009). Maximum pseudo-likelihood estimation (MPLE) here is equivalent to logistic regression of the elements of $\mathbf{y}$ on the design matrix formed by the vectors $\Delta_{ij}(\mathbf{y})$, and may thus be carried out using standard software (van Duijn *et al.*, 2009). The computational tractability of the pseudolikelihood function makes it an attractive alternative to the full likelihood function and it has made ERGMs computationally available. For this reason this approach has been the primary method of estimation, until the more recent development of MCMC methods (Robins *et al.*, 2007b) which try to overcome the limits of this methods. When the dependence among observations becomes stronger (the dyadic dependence models), the statistical properties for the maximum pseudolikelihood estimators (MPLE) are not well understood and in practice MPLE does not provide a good performance (van Duijn *et al.*, 2009). In this case the parameter estimates may be biased and it could be risky to interprete standard errors from logistic regression output as though they are reasonable estimates of the standard deviations of the pseudolikelihood estimators.

Model Degeneracy and Goodness of Fit

The concept of model goodness of fit is important and well developed in standard modeling contexts, such as linear modeling, but it is still not so well developed in the context of network graph modeling (Kolaczyk, 2009). For

ERGMs, the current practice in assessing goodness of fit is to first simulate numerous random graphs from the fitted model and then compare high-level characteristics of these graphs with those of the originally observed graph (Hunter *et al.*, 2008b), (Goodreau *et al.*, 2008). If our model can generate a distribution of networks that have consistent network measures with the observed one we may say it is a good model. Otherwise, if the characteristics of the observed network graph are too poor of a match to the typical values arising from realizations of the fitted random graph model, then this suggests systematic differences between the specified class of models and the data, and therefore a lack of goodness of fit (Handcock *et al.*, 2008).
The method proposed by Hunter *et al.* (2008b) and Goodreau *et al.* (2008) consists of simulating many networks (at least 100) from the final coefficient estimates and compare features of the observed network with the same features of a set of simulated networks. If the original network is inconsistent with the networks generated from the model, this suggests that the structure of the network differs from those predicted by the model, and the model is not fitted well. This method provides some limitations because it gives possibility to compare only a single outcome (i.e. number of edges) from the simulations, so Handcock *et al.* (2008) propose graphical tests of goodness of fit (the gof function). Moreover the authors compare their graphical methods for assessing goodness of fit with more traditional methods such as AIC or BIC. Goodness of fit has been found to be particularly important where ERGMs are concerned, due in large part to the issue of model degeneracy (Kolaczyk, 2009). In this context the term is used to refer to a probability distribution that places a disproportionately large amount of its mass on a correspondingly small set of outcomes (Handcock, 2003a), (Handcock, 2003b).
A number of simple but originally popular Markov graph models have been shown to be degenerate. A common case is where the ERGM places most of its mass on either the empty graph, the complete graph, or a mixture of the two, depending on the value of θ. None of these, of course, is likely to be especially appropriate for modeling real data of any interest. In addition to its defining lack of richness, model degeneracy also can lead to difficulties

in fitting ERGMs. For any estimation procedure, it is very important that the model is nondegenerate one, otherwise it is difficult to obtain satisfactory convergence of parameter estimates. If degeneracy model is used for simulation and MCMC likelihood inference, the approximations to the true model will be very poor. Thus, for most applications we should seek to limit our space of viable models.
In summary, model specification for ERGMs is clearly a non-trivial task and one that should be approached in a manner closely informed by the above issues, with goodness of fit diagnostics playing an important facilitating role (Kolaczyk, 2009). The problem of degeneracy and poor fitting can be resolved using models based on recently developed statistics (like the alternating k-stars and the alternating k-triangles mentioned above) which have been proven to be more robust and very effective in representing real network data. These statistics seem to capture high-order dependency structure in networks and to contain a lot of significant information about network (Snijders *et al.*, 2006), (Hunter, 2007).

3.2 Exponential Random Graph Modeling for Complex Brain Networks

In this section we illustrate the utility of ERGMs for modeling brain network. Chapter 1 reported descriptive analysis based on a specific feature of the network such as characteristic path length or clustering coefficient. To examine group differences based on one of these features, inferential studies have been employed based on testing techniques. ERGMs provide an approach to explore several features simultaneously, to study their interactions and to compare different groups of networks (Simpson *et al.*, 2011).
Data used in this application include whole-brain functional connectivity networks for 55 normal subjects (mean age=41 years; range=20-65 years, 20/35 males/females) and 121 patients with a diagnosis of multiple sclerosis (MS) (mean age=40 years; range=19-63 years, 44/77 males/females). Multiple sclerosis (MS) is an inflammatory disease of the central nervous system in

which the myelin sheaths around the axons of the brain and spinal cord are damaged, leading to demyelination and a broad spectrum of signs and symptoms. MS affects the ability of nerve cells in the brain and spinal cord to communicate with each other effectively. Almost any neurological symptom can appear with the disease, and often progresses to physical and cognitive disability.

Each network comprises 116 nodes corresponding to the 116 brain regions (116 ROIs-Regions of Interest) defined by the Automated Anatomical Labeling atlas. Time series were extracted from each ROI by averaging the signal of all voxels within that region. Bivariate correlations between each ROI pair were obtained by calculating the Pearson's correlation coefficient ρ_{ij} between ROIs time courses i and j. Correlation matrices of size $\{116 \times 116\}$, obtained from all study subjects, were thresholded deriving an adjacency matrix A with entries:

$$A_{ij} = \begin{cases} 1, & \text{if } \rho_{ij} \geq \tau \\ 0, & \text{otherwise} \end{cases}$$

A is computed putting the A_{ij} elements to zero if $\rho_{ij} < \tau$ and unity if $\rho_{ij} \geq \tau$, where $0 < \tau < 1$ is a suitable threshold. This threshold is defined so that the relationship between the number of nodes and the average node degree $Deg(G)$ is the same across different subjects ($Deg(G) = k$). From a neurological point of view, for this particular pathology, it is advisable to fix the average degree (Schoonheim *et al.*, 2012). Consequently, using this approach, the threshold varies from subject to subject.

3.2.1 ERGMs for a Brain Network

Exponential random graph models have the form (3.4).

In this contest $\mathbf{Y}$ is a $\{116 \times 116\}$ random symmetric adjacency matrix representing a brain network from a particular class of networks, with $Y_{ij} = 1$ if an edge exists between nodes i and j and $Y_{ij} = 0$ otherwise. Nodes represent locations in the brain (e.g., ROIs) and edges represent functional connections between them. We statistically model the probability mass function $P_\theta(\mathbf{Y} = \mathbf{y})$ of this class of networks as a function of the pre-specified network

features defined by the p-dimensional vector $g(\mathbf{y})$. This vector of explanatory metrics (features) consists of covariates that are functions of the network $\mathbf{y}$ and can contain any graph statistic (e.g., number of paths of length two) or node statistic, also called attributes (e.g., brain location of the node). The parameter vector θ , associated with g($\mathbf{y}$), quantifies the relative significance of the network features to explain the structure of the network taking into account the contribution of all other network features in the model. More specifically, θ indicates the change in the log odds of an edge existing for each unit increase in the corresponding explanatory metric. If the θ value corresponding to a given metric is large and positive, then the metric plays a considerable role to explain the network architecture and is more prevalent than in the null model (random network with the probability of an edge existing $p = 0.5$). Conversely, if the θ value is large and negative, then this metric still plays a considerable role to explain the network architecture but is less prevalent than in the null model. Consequently, inferences can be made about whether certain local features are observed in the network more than it would be expected by chance enabling hypothesis development regarding the biological processes that produce these structural properties. The normalizing constant κ ensures that the probabilities sum to one. This approach allows representing the global network structure by specified explanatory metrics. The goal in defining g($\mathbf{y}$) is to identify metrics that summarize the whole-brain network structure.

In Simpson *et al.* (2011) a subset of mathematically compatible explanatory network metrics is defined (Saul and Filkov, 2007), (Handcock *et al.*, 2008), (Hunter and Handcock, 2006) and the most appropriate set of explanatory metrics and a 'best assessment' ERGM for complex brain networks are provided. The potential explanatory metrics are chosen based on properties of brain networks that are regarded as important in the literature. These metrics are analogous to typical brain network metrics (see Chapter 1) but have been developed to be statistically compatible with ERGMs.

The geometrically weighted statistics discussed in Hunter *et al.* (2008b) help address degeneracy issues described by Handcock (2003b) and in Rinaldo *et al.* (2009).

Here we use the Simpson's overall ERGM for whole-brain networks with a connectedness metric (Edges) [2], a clustering metric (GWESP) [3], and a global efficiency metric (GWNSP) [4]. That is,

$$P_\theta(\mathbf{Y} = \mathbf{y}) = \frac{1}{\kappa} exp\{\theta_1 Edges + \theta_2 GWESP + \theta_3 GWNSP\} \tag{3.22}$$

to model a healthy control network and a patient network, we examine several possible network configurations for $Deg(G) = k$, and we explored the results over $k = 4.5$, $k = 20$, $k = 45$. We started from the degree fixed by Simpson *et al.* (2011) and we explore other values for k that are closer to the values used in the descriptive analysis by Supekar *et al.* (2008). To this purpose we use the statnet package (Handcock *et al.*, 2008), (Hunter *et al.*, 2008a) for the R statistical computing environment.

The ν parameters associated with geometrically weighted statistics are all assumed to be fixed and known and set to $\nu = 0.75$ based on preliminary analyses as this value generally led to better fitting models.

Fitting of the ERGM is normally done with either Markov chain Monte Carlo maximum likelihood estimation (MCMC MLE) or maximum pseudo-likelihood estimation (MPLE) Robins *et al.* (2007a). Model fits with MPLE are much simpler computationally than MCMC MLE fits and afford higher convergence rates with large networks, but properties of the MPLE estimators are not well understood, and the estimates tend to be less accurate than those of MCMC MLE (van Duijn *et al.*, 2009). Here MCMC MLE is employed to fit the model, whereas MPLE fits is used as an appropriate alternative in the case of computational limitations which can cause problems of convergence. For this case we used the MCMC MLE to fit the networks with fixed average degree $k = 4.5$ which is the same used by Simpson *et al.* (2011) and for $k = 20$, whereas for $k = 45$ we used the MPLE.

[2] Number of edges in the network.

[3] Geometrically weighted edge-wise shared partner is a weighted sum of the number of connected nodes having exactly i shared partners weighted by the geometric sequence $(1 - exp(-\nu))^i$, where ν is a decay parameter.

[4] Geometrically weighted non-edge-wise shared partner is a weighted sum of the number of non-connected nodes having exactly i shared partners weighted by the geometric sequence $(1 - exp(-\nu))^i$, where ν is a decay parameter.

	Estimate	SE	p-value
$k = 4.5$			
$Edges$	-4.79546	0.21427	$< 1e-04$
$GWESP$	1.52919	0.09177	$< 1e-04$
$GWNSP$	-0.14301	0.11369	0.208
$k = 20$			
$Edges$	-2.22538	0.70118	0.001512
$GWESP$	0.99817	0.50068	0.046232
$GWNSP$	-0.26140	0.02073	$< 1e-04$
$k = 45$			
$Edges$	-5.27823	0.47417	$< 1e-04$
$GWESP$	2.73056	0.21708	$< 1e-04$
$GWNSP$	-0.39355	0.01473	$< 1e-04$

Table 3.1: ERGM estimates for a healthy control subject with $Deg(G)$=k=4.5, 20, 45

The resulting model and its corresponding parameter estimates are displayed in Table 3.1 where the results for a healthy control (HC) are reported and in Table 3.2 where the results for a patient (RRMS) are reported.

These estimates quantify the relative significance of the given metric in explaining the overall network structure. In order to test the significance of the estimates a one sample T-test is used, as the distribution of the statistic formed as the ratio of the estimate to its standard error is not known exactly, but likely to approximate a T-Student distribution (Snijders, 2002).
According to Simpson *et al.* (2011), these three metrics have an impact on overall functional brain network organization, the p-value is shown in Table 3.1 and in Table 3.2. The number of functional connections (Edges) is instrumental in information transfer while also playing a role in brain network organization. The negative density (Edges) parameter indicates that edges occur relatively rarely, especially if they are not part of higher order structures such as triangles.

Clustering (GWESP) is another critical feature of brain network architecture that allows the efficient local processing of information. In this case GWESP is an important metric in describing the structure of the subjects

	Estimate	SE	p-value
$k = 4.5$			
$Edges$	-4.81633	0.22254	$< 1e-04$
$GWESP$	1.52429	0.10495	$< 1e-04$
$GWNSP$	-0.10789	0.10431	0. 03
$k = 20$			
$Edges$	-2.94445	0.42465	$< 1e-04$
$GWESP$	1.17885	0.13826	$< 1e-04$
$GWNSP$	-0.23452	0.16260	0.03
$k = 45$			
$Edges$	-3.53978	0.39874	$< 1e-04$
$GWESP$	1.79015	0.18368	$< 1e-04$
$GWNSP$	-0.34945	0.01664	$< 1e-04$

Table 3.2: ERGM estimates for a RRMS patient with $Deg(G)$=k=4.5, 20, 45

network given the larger absolute value of the parameter estimate. The positiveness of the estimate associated with GWESP indicates that an edge that closes a triangle is more likely to exist than it would by chance (i.e., the network has more clustering than a random network where the probability of an edge is $p = 0.5$). The consistently negative θ_3 values associated with GWNSP indicate that if two brain areas are not functionally connected, they are less likely to have shared connections with other regions than they would by chance. That is, two regions are less likely than by chance to have a 2-path as the shortest path between them. Speculatively, this may result from the brain having direct connections when necessary, but allowing for slightly longer global connections (3-paths, etc.) to maintain efficiency otherwise (Simpson *et al.*, 2011).

	Estimate	SE	p-value
$k = 4.5$			
$Edges$	-4.04274	0.10410	$< 1e-04$
$GWESP$	1.36812	0.07549	$< 1e-04$
$GWNSP$	-0.26482	0.05207	$< 1e-04$
$k = 20$			
$Edges$	-3.25313	0.41328	$< 1e-04$
$GWESP$	1.32551	0.16515	$< 1e-04$
$GWNSP$	-0.19023	0.0641	0.00303
$k = 45$			
$Edges$	-5.26580	0.41741	$< 1e-04$
$GWESP$	2.53627	0.19288	$< 1e-04$
$GWNSP$	-0.31301	0.01362	$< 1e-04$

Table 3.3: ERGM estimates for the healthy controls representative network with $Deg(G)$=k=4.5, 20, 45

3.2.2 ERGMs for a Group Representative Brain Network

Group-based brain connectivity networks have great appeal for researchers interested in gaining further insight into complex brain function and how it changes across different mental states and disease conditions. Viable approaches to this task must engender networks that capture the constitutive topological properties of the group of subjects' networks that it is aiming to represent (Simpson *et al.*, 2012).

In Chapter 1 we have proposed an approach to creating group-based representative networks utilizing the z-to-ρ transformation, described in section 2.7.4. In this section we will model the patient representative network and the healthy controls representative network using the model in (3.22). Its corresponding parameter estimates are displayed in Table 3.3 and in Table 3.4 .

This approach could be useful to explore which network metrics best characterize a group of brain networks. We can notice that the results are consistent with those reported Table 3.3 and Table 3.4 for a single subject network.

	Estimate	SE	p-value
$k = 4.5$			
$Edges$	-3.93314	0.19331	$< 1e-04$
$GWESP$	1.52429	0.10495	$< 1e-04$
$GWNSP$	-0.28507	0.05424	$< 1e-04$
$k = 20$			
$Edges$	-2.58034	0.39183	$< 1e-04$
$GWESP$	1.00088	0.12338	$< 1e-04$
$GWNSP$	-0.27936	0.14138	0.0482
$k = 45$			
$Edges$	-4.72848	0.39039	$< 1e-04$
$GWESP$	2.27267	0.18025	$< 1e-04$
$GWNSP$	-0.31619	0.01396	$< 1e-04$

Table 3.4: ERGM estimates for the RRMS representative network with $Deg(G)$=k=4.5, 20, 45

3.2.3 ERGMs to compare two Groups of Networks

Once the most appropriate statistics have been established, parameter profiles (θ) can be utilized to classify and compare whole-brain networks.
The following analysis includes ERGM fits to thresholded networks with degree fixed as described in the introduction of this section. The parameter profile comparisons require the use of the same set of explanatory metrics for all networks (due to metric interdependencies) and same number of nodes for all networks.
Group-based network comparisons can potentially be performed by comparing the mean of the estimated θ_1, θ_2, and θ_3 values among groups via hypothesis testing or classification techniques (Simpson *et al.*, 2011). One of the advantage of this analysis is that if we want to compare, for instance,the mean of the estimated θ_1 (Edges) values among groups, potential confounding from the GWESP and GWNSP would be inherently accounted.
In the hypothesis testing framework one can exploit the fact that the $\hat{\theta}$s are approximate MLEs and thus asymptotically have a Gaussian distribution. For this applicative study, first a F test to verify the equality of variances and then a T-test can be employed in order to elucidate biologically inter-

	HC		SM		
	Mean	SE	Mean	SE	p-value
k = 4.5					
θ_1 (Edges)	-4.88	0.49	-4.9	0.41	0.7
θ_2 (GWESP)	1.68	0.26	1.71	0.27	0.4
θ_3 (GWNSP)	-0.15	0.09	-0.16	0.08	0.3
k = 20					
θ_1 (Edges)	-3.44	0.74	-3.69	0.71	0.03
θ_2 (GWESP)	1.65	0.3	1.75	0.30	0.04
θ_3 (GWNSP)	-0.29	0.05	-0.28	0.05	0.7
k = 45					
θ_1 (Edges)	-4.36	1.42	-4.9	1.82	0.04
θ_2 (GWESP)	2.17	0.65	2.37	0.85	0.08
θ_3 (GWNSP)	-0.35	0.04	-0.37	0.06	0.06

Table 3.5: Results of ERGM parameter estimate comparisons between healthy controls (HC) and multiple sclerosis (MS) patients with different average degree $Deg(G)$=k=4.5, 20, 45

esting differences between the two groups.

Here we implement the Simpson's best assessment ERGM from equation (3.22) to illustrate its utility for comparing groups of networks. We implement the model for each of the 176 subjects (55 healthy controls and 121 patients with multiple sclerosis). The aim is to assess if there were any discernible differences between controls and patients. The results of this analysis are exhibited in Table 3.5 . In particular in Table 3.5 we observe that, at $Deg(G)$ = 20, the two groups differ significantly in θ_1, with the SM group having a more negative value. An increased number of edges in MS patients compared with controls represents a tendency towards an increased functional connectivity in Ms patients, in the attempt to compensate structural damage due to demyelination. Moreover the two groups significantly differ in θ_2: in MS patients θ_2 is positive and higher than in HC group denoting a higher tendency to clusterization.

3.3 Closing Remarks

The general theory of network model and, in particular, exponential random graph models (ERGMs) are introduced in the first part of this chapter.
In the second part, a particular assessment of ERGMs for whole-brain networks has been used to model a healthy control and a patient network. Applying different assessments of this analysis provides a way of exploring which set of features is the most important in explaining the global architecture of the brain network. In this sense an important perspective could be the development of novel explanatory network metrics in order to define more appropriate model settings.
Furthermore the ERGM has been applied to a group representative brain network: this approach is able to capture the network characteristics from a group of subjects. These representative networks can serve as null networks against which other networks can be compared, as well as a visualization tool and as a means to define properties of network metrics in a group (Simpson *et al.*, 2012).
Finally, group-based network comparisons have been performed by comparing the mean of the estimated parameters between healthy controls and patients with multiple sclerosis via hypothesis testing. This application has shown the utility of ERGMs to characterize change of functional network properties in the context of this disease.
Another important perspective is that ERGMs can be applied in medical prediction problems,in particular, to characterize features of the pathological brain. As shown above, healthy controls and multiple sclerosis data can be used to establish which variables are influential in predicting the given outcome (Y. They can then be measured for a new patient, through placing in ERGM model, to calculate the probability of given outcome. For instance, we are interested in understating if a new subject can be classified as a patient (MS) given an observed outcome, that is $Pr(MS|Y = y)$. Let $Pr(Y = y|MS) = ERGM_1(y)$, $Pr(Y = y|\overline{MS}) = ERGM_0(y)$ and $Pr(MS)$

be the probability of occurrence of the disease, then

$$\frac{Pr(MS|Y = y)}{Pr(\overline{MS}|Y = y)} = \frac{Pr(Y = y|MS)Pr(MS)}{Pr(Y = y|\overline{MS})Pr(\overline{MS})}. \tag{3.23}$$

In this way we are able to know when $Pr(MS|Y = y) > Pr(\overline{MS}|Y = y)$, because this relation is verified if and only if

$$\frac{Pr(Y = y|MS)Pr(MS)}{Pr(Y = y|\overline{MS})Pr(\overline{MS})} = \frac{ERGM_1(y)Pr(MS)}{ERGM_0(y)\left[1 - Pr(MS)\right]} > 1. \tag{3.24}$$

Conclusions

The brain can be described as a complex system and modeled as a network where the vertices are brain regions and edges represent the interactions between them. An accurate description of the architecture and a characterization of the network properties can be important to understand the functional and structural connectivity of the brain. In this context the network analysis describes complex systems by quantifying properties of network representation.

In this work, the basic terminology and notation for graphs were discussed and the network analysis was applied to a particular case study in order to investigate the structure and the characteristics of brain networks and to provide new measures to quantify differences between patient groups. In our study, we quantified this difference by using metrics of graph theory, in order to explore the difference of brain functional connectivity between healthy subjects and patients with a diagnosis of behavioural variant of frontotemporal dementia (bvFTD). bvFTD is a clinical syndrome caused by the degeneration of the frontal and temporal lobes of the brain and it is the second-most common dementia after Alzheimer's disease. To this purpose we measured brain network connectivity with functional magnetic resonance images (fMRI) at resting state, i.e., with subjects lying still in the scanner and thinking nothing in particular. Global and local network characteristics were explored by assessing the principal measures of graph theory, which are able to detect various aspects of functional connectivity, e.g. quantify importance of brain regions, characterize patterns of local anatomical circuitry, test resilience of networks to insult and so on. In order to characterize the global organization

of brain networks, the average degree, the global efficiency, the clustering coefficient and the characteristic path length were adopted. Global network architecture subjects was quantified in terms of *small-worldness*, computing the normalized clustering coefficient and the normalized characteristic path length. Most of graph theoretical metrics were found to be significantly altered in bvFTD patients with respect to healthy controls and, in both controls and bvFTD patients, *small-worldness* properties were verified. Moreover local network measures were calculated in order to identify which brain regions could be considered 'hubs', i.e., regions interacting with several other brain areas, thus facilitating functional integration. Hubs are identified on the basis of their centrality in the network. To this aim, we calculated two metrics of centrality: the degree and the betweenness centrality. This applicative study confirmed that complex network analysis can be used to explore connectivity relationships in individual subjects and to explore connectivity differences between subject groups. By analyzing fMRI fluctuations during resting state condition, we have demonstrated that bvFTD condition induced changes in global and local brain functional connectivity in line with the expected neuropathological alterations. Global functional network organization was altered in bvFTD patients, suggesting a loss of efficiency in information exchange between brain areas. Moreover, the local network analysis reveals that the altered brain regions are mainly located in the frontal lobe (including the anterior cingulate cortex), temporal lobe and basal ganglia that are closely associated with the expected neuropathology of bvFTD.

In the second part of this work, the adjacency matrix has been constructed with an original methodology involving the problem of multiple comparisons. To simplify analysis, networks are often reduced to a binary (undirected) form, through an adjacency matrix. The correlation matrices are typically converted to a graph by considering a threshold τ, to create an adjacency matrix. In previous works, there was no unique way to choose τ, so, the possible network configurations were examined by constructing graphs for a range of values of τ within which the consistency of the network characteristics was explored. In this context, we proposed an innovative methodology for the

derivation of the adjacency matrix using a suitable threshold, performing a multiple test on correlation coefficient. Then we offered a statistical evaluation of the error committed, considering the problem of multiple comparisons. A correction is needed to control the error rates, but standard procedures for multiple hypothesis testing (e.g., Bonferroni correction) are known to be too conservative. So we have dealt with the problem of multiple comparisons using the method based on the estimation of positive false discovery rate ($pFDR$) proposed for the first time by Storey (2002). Furthermore we considered a similar measure involving false negatives (type II errors), called the positive nondiscovery negatives rate ($pFNR$), we proposed new point and interval estimators for $pFNR$ and a method to balance the two types of error. We demonstrated this approach using both simulations and fMRI data providing finite sample and large sample results for $\widehat{pFDR}$ and $\widehat{pFNR}$. We reported some results with the proposed estimators for $pFDR$, $pFNR$ and the average power (Pwa). In particular we derived the confidence intervals for them using the Bootstrap method, and we evaluated their coverage probability. We showed the results of the simulation study, reporting the true values of $pFDR$, $pFNR$ and the average power Pwa, their point estimates, the MSE and the 95% confidence interval. First we have simulated a typical brain network and calculated these values for different rejection regions and associated the correspondent threshold τ on correlation coefficient. Then we evaluated the coverage probability of the confidence intervals and we found that this is always equal to 100%. We found similar results with other brain network simulations. Furthermore we applied the proposed estimates for $pFDR$, $pFNR$ and Pwa to real data. We calculated the estimates in the context of a real network obtained from fMRI data and we described how to choose the threshold on correlation coefficient that provides a balance between pFDR and pFNR. We continued the applicative study introduced in the previous chapter utilizing such balanced networks to investigate the brain network characteristics, we proposed a ranking of the most central nodes in the networks, using q-values, which are the $pFDR$ analog of the p-values, we examined the differences on the inter-regional connectivity between the control group and the patient (bvFTD) group. We reported the results of the

comparison of the nodal degree between healthy controls and patients with the behavioral variant of frontotemporal dementia. Brain regions were ranked according to their importance in explaining group differences by means of the q-value. Moreover we considered the multiple hypothesis testing approach to create a group differences network. We proposed this new kind of network showing the significant differences in the inter-regional connectivity between healthy controls and bvFTD patients. Finally, we proposed a new method to construct a group-based representative network that allows to determine the type of connectivity of a brain network set and we showed an example to see how this method works, considering the group of controls. As we have found the representative network of the controls, the same procedure can be also applied to different groups of patients. A group-based representative brain connectivity network can provide a graph that typifies the complex structure of a set of brain networks and it can be useful to capture the network characteristics from a group of subjects' brain networks, serving as null networks against which other networks and network models could be compared.

In the third part of this work, the general theory of network model and exponential random graph models (ERGMs) were introduced. A particular assessment of ERGMs for whole-brain networks has been used to model a healthy control and a patient network. Data used in this application included whole-brain functional connectivity networks for healthy subjects and patients with a diagnosis of relapsing-remitting multiple sclerosis (RRMS). We used the ERGM for whole-brain networks with a Connectedness metric (Edges), i.e., the number of edges in the network, a Local Efficiency metric, i.e., the geometrically weighted edge-wise shared partner (GWESP) and a Global Efficiency metric, i.e., the geometrically weighted non-edge-wise shared partner (GWNSP), to model a healthy control network and a patient network, and we examined several possible network configurations. We displayed the resulting model and its corresponding parameter estimates. Furthermore we applied the ERGM to a group representative brain network: this approach is able to capture the network characteristics from a group of subjects. These representative networks can serve as null networks

against which other networks can be compared, as well as a visualization tool and as a means to define properties of network metrics in a group. We demonstrated that the three metrics, considered above, have an impact on overall functional brain network organization. Finally, group-based network comparisons have been performed by comparing the mean of the estimated parameters among groups via hypothesis testing. This application has shown the utility of ERGMs in the clinical context, for understand neurological processes (represented by the explanatory network metrics) that can play a role in different diseases. We observed that, in a particular network configuration, the two groups differed significantly in the Edegs coefficient, with the HC group having a more negative value. Furthermore the two groups significantly differed in GWESP coefficient, which, in RRMS patients was positive and higher than in control group denoting a higher tendency to clusterization.

Bibliography

Barndorff-Nielsen, O. (1978). *Information and Exponential Families in Statistical Theory*. John Wiley, New York.

Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, **57**(1), 289–300.

Besag, J. (1975). Statistical analysis of non-lattice data. *The statistician*, **24(3)**, 179–195.

Bollobas, B. (2001). *Random graphs*, volume 73. Cambridge Univ Pr.

Bullmore, E. and Sporns, O. (2009). Complex brain networks: graph theoretical analysis of structural and functional systems. *Nat Rev Neurosci*, **10**(3), 186–198.

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

De Haan, W., Pijnenburg, Y., Strijers, R., Van Der Made, Y., van der Flier, W., Scheltens, P., and Stam, C. (2009). Functional neural network analysis in frontotemporal dementia and alzheimer's disease using eeg and graph theory. *BMC neuroscience*, **10**(1), 101.

Erdos, P. and Renyi, A. (1959). On random graphs, i. *Publicationes Mathematicae (Debrecen)*, **6**, 290–297.

Erdos, P. and Renyi, A. (1960). On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci*, **5**, 17–61.

Erdos, P. and Renyi, A. (1961). On the strength of connectedness of a random graph. *Acta Mathematica Hungarica*, **12**(1), 261–267.

Fisher, R. (1915). Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population. *Biometrika*, **10**(4), 507–521.

Frank, O. and Strauss, D. (1986). Markov graphs. *Journal of the American Statistical Association*, **81(395)**, 832–842.

Freeman, L. (1977). A set of measures of centrality based on betweenness. *Sociometry*, **40**(4), 35–41.

Genovese, C., Lazar, N., and Nichols, T. (2002). Thresholding of statistical maps in functional neuroimaging using the false discovery rate. *Neuroimage*, **15**(4), 870–878.

Geyer, C. and Thompson, E. (1992). Constrained monte carlo maximum likelihood for dependent data. *Journal of the Royal Statistical Society. Series B (Methodological)*, **54(3)**, 657–699.

Gilbert, E. (1959). Random graphs. *The Annals of Mathematical Statistics*, **30**(4), 1141–1144.

Goodreau, S., Handcock, M., Hunter, D., Butts, C., and Morris, M. (2008). A statnet tutorial. *Journal of statistical software*, **24**(9), 1.

Handcock, M. (2003a). Assessing degeneracy in statistical models of social networks. *Center for Statistics and the Social Sciences, University of Washington*, **39**.

Handcock, M. (2003b). Statistical models for social networks: Inference and degeneracy. *Dynamic social network modeling and analysis*, **126**, 229–252.

Handcock, M., Hunter, D., Butts, C., Goodreau, S., and Morris, M. (2008). statnet: Software tools for the representation, visualization, analysis and simulation of network data. *Journal of Statistical Software*, **24**(1), 1548.

Holland, P. and Leinhardt, S. (1981). An exponential family of probability distributions for directed graphs. *Journal of the american Statistical association*, **76**(373), 33–50.

Hunter, D. (2007). Curved exponential family models for social networks. *Social Networks*, **29**(2), 216–230.

Hunter, D. and Handcock, M. (2006). Inference in curved exponential family models for networks. *Journal of Computational and Graphical Statistics*, **15**(3), 565–583.

Hunter, D., Handcock, M., Butts, C., Goodreau, S., and Morris, M. (2008a). ergm: A package to fit, simulate and diagnose exponential-family models for networks. *Journal of Statistical Software*, **24**(3), nihpa54860.

Hunter, D., Goodreau, S., and Handcock, M. (2008b). Goodness of fit of social network models. *Journal of the American Statistical Association*, **103**(481), 248–258.

Jackson, M. (2008). *Social and economic networks*. Princeton Univ Pr.

Koehly, L. and Pattison, P. (2005). Random graph models for social networks: Multiple relations or multiple raters. *Models and methods in social network analysis*, pages 162–191.

Kolaczyk, E. (2009). *Statistical analysis of network data: methods and models*. Springer Verlag.

Landenna, G., Marasini, D., and Ferrari, P. (1997). *La verifica di ipotesi statistiche*. Il Mulino ed., Bologna.

Latora, V. and Marchiori, M. (2001). Efficient behavior of small-world networks. *Physical Review Letters*, **87**(19), 198701.

Latora, V. and Marchiori, M. (2003). Economic small-world behavior in weighted networks. *The European Physical Journal B-Condensed Matter and Complex Systems*, **32**(2), 249–263.

Lehmann, E. (1958). Significance level and power. *The Annals of Mathematical Statistics*, **29**(4), 1167–1176.

Lehmann, E. and Romano, J. (2005). *Testing statistical hypotheses*. Springer Verlag.

Liang, M., Zhou, Y., Jiang, T., Liu, Z., Tian, L., Liu, H., and Hao, Y. (2006). Widespread functional disconnectivity in schizophrenia with resting-state functional magnetic resonance imaging. *Neuroreport*, **17**(2), 209.

Maslov, S. and Sneppen, K. (2002). Specificity and stability in topology of protein networks. *Science's STKE*, **296**(5569), 910.

Meunier, D., Lambiotte, R., Fornito, A., Ersche, K., and Bullmore, E. (2009). Hierarchical modularity in human brain functional networks. *Frontiers in neuroinformatics*, **3**(37).

Milgram, S. (1967). The small world problem. *Psychology today*, **2**(1), 60–67.

Pattison, P. and Wasserman, S. (1999). Logit models and logistic regressions for social networks: Ii. multivariate relations. *British Journal of Mathematical and Statistical Psychology*, **52**(2), 169–193.

Rinaldo, A., Fienberg, S., and Zhou, Y. (2009). On the geometry of discrete exponential families with application to exponential random graph models. *Electronic Journal of Statistics*, **3**, 446–484.

Robbins, H. and Monro, S. (1951). A stochastic approximation method. *The Annals of Mathematical Statistics*, **22**, 400–407.

Robins, G., Pattison, P., Kalish, Y., and Lusher, D. (2007a). An introduction to exponential random graph (p*) models for social networks. *Social Networks*, **29**(2), 173–191.

Robins, G., Snijders, T., Wang, P., Handcock, M., and Pattison, P. (2007b). Recent developments in exponential random graph (¡ i¿ p¡/i¿*) models for social networks. *Social Networks*, **29**(2), 192–215.

Robins, G., Pattison, P., and Wang, P. (2009). Closure, connectivity and degree distributions: Exponential random graph (¡ i¿ p¡/i¿*) models for directed social networks. *Social Networks*, **31**(2), 105–117.

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

Salvador, R., Suckling, J., Schwarzbauer, C., and Bullmore, E. (2005). Undirected graphs of frequency-dependent functional connectivity in whole brain networks. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **360**(1457), 937–946.

Sanz-Arigita, E., Schoonheim, M., Damoiseaux, J., Rombouts, S., Maris, E., Barkhof, F., Scheltens, P., and Stam, C. (2010). Loss of small-worldnetworks in alzheimer's disease: Graph analysis of fmri resting-state functional connectivity. *Plos One*, **5**(11), e13788.

Saul, Z. and Filkov, V. (2007). Exploring biological network structure using exponential random graph models. *Bioinformatics*, **23**(19), 2604–2611.

Schoonheim, M., Hulst, H., Landi, D., Ciccarelli, O., Roosendaal, S., Sanz-Arigita, E., Vrenken, H., Polman, C., Stam, C., Barkhof, F., *et al.* (2012). Gender-related differences in functional connectivity in multiple sclerosis. *Multiple Sclerosis Journal*, **18**(2), 164–173.

Shao, J. and Tu, D. (1996). *The jackknife and bootstrap*. Springer.

Simpson, S., Hayasaka, S., and Laurienti, P. (2011). Exponential random graph modeling for complex brain networks. *PloS one*, **6**(5), e20039.

Simpson, S., Moussa, M., and Laurienti, P. (2012). An exponential random graph modeling approach to creating group-based representative whole-brain connectivity networks. *NeuroImage*, **60**, 1117–1126.

Smith, S., Miller, K., Salimi-Khorshidi, G., Webster, M., Beckmann, C., Nichols, T., Ramsey, J., and Woolrich, M. (2011). Network modelling methods for fmri. *Neuroimage*, **54**(2), 875–891.

Snijders, T. (2002). Markov chain monte carlo estimation of exponential random graph models. *Journal of Social Structure*, **3**(2), 1–40.

Snijders, T., Pattison, P., Robins, G., and Handcock, M. (2006). New specifications for exponential random graph models. *Sociological Methodology*, **36**(1), 99–153.

Song, M., Liu, Y., Zhou, Y., Wang, K., Yu, C., and Jiang, T. (2009). Default network and intelligence difference. *Autonomous Mental Development, IEEE Transactions on*, **1**(2), 101–109.

Stam, C., De Haan, W., Daffertshofer, A., Jones, B., Manshanden, I., van Walsum, A., Montez, T., Verbunt, J., De Munck, J., Van Dijk, B., *et al.* (2009). Graph theoretical analysis of magnetoencephalographic functional connectivity in alzheimer's disease. *Brain*, **132**(1), 213–224.

Storey, J. (2002). A direct approach to false discovery rates. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **64**(3), 479–498.

Storey, J. (2003). The positive false discovery rate: A bayesian interpretation and the q-value. *Annals of Statistics*, **31**(6), 2013–2035.

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

Strauss, D. and Ikeda, M. (1990). Pseudolikelihood estimation for social networks. *Journal of the American Statistical Association*, **85(409)**, 204–212.

Supekar, K., Menon, V., Rubin, D., Musen, M., and Greicius, M. (2008). Network analysis of intrinsic functional brain connectivity in alzheimer's disease. *PLoS computational biology*, **4**(6), e1000100.

Tian, L., Wang, J., Yan, C., and He, Y. (2011). Hemisphere-and gender-related differences in small-world brain networks: a resting-state functional mri study. *Neuroimage*, **54**(1), 191–202.

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., and Joliot, M. (2002). Automated anatomical labeling of activations in spm using a macroscopic anatomical parcellation of the mni mri single-subject brain. *Neuroimage*, **15**(1), 273–289.

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

Van Den Heuvel, M., Mandl, R., Kahn, R., and Hulshoff Pol, H. (2009). Functionally linked resting-state networks reflect the underlying structural connectivity architecture of the human brain. *Human brain mapping*, **30**(10), 3127–3141.

van Duijn, M., Gile, K., and Handcock, M. (2009). A framework for the comparison of maximum pseudo-likelihood and maximum likelihood estimation of exponential family random graph models. *Social Networks*, **31**(1), 52–62.

Van Wijk, B., Stam, C., and Daffertshofer, A. (2010). Comparing brain networks of different size and connectivity density using graph theory. *PLoS One*, **5**(10), e13701.

Wasserman, S. (1994). *Social network analysis: Methods and applications*. Cambridge university press.

Wasserman, S. and Pattison, P. (1996). Logit models and logistic regressions for social networks: I. an introduction to markov graphs andp. *Psychometrika*, **61**(3), 401–425.

Watts, D. and Strogatz, S. (1998). Collective dynamics of small-world networks. *Nature*, **393**(6684), 440–442.

Wig, G., Schlaggar, B., and Petersen, S. (2011). Concepts and principles in the analysis of brain networks. *Annals of the New York Academy of Sciences*, **1224**(1), 126–146.

Yao, Z., Zhang, Y., Lin, L., Zhou, Y., Xu, C., and Jiang, T. (2010). Abnormal cortical networks in mild cognitive impairment and alzheimer's disease. *PLoS Computational Biology*, **6**(11), e1001006.

Zuo, X., Ehmke, R., Mennes, M., Imperati, D., Castellanos, F., Sporns, O., and Milham, M. (2011). Network centrality in the human functional connectome. *Cerebral Cortex*.