UNIVERSITÀ DI VERONA

Università degli studi di Verona

Dipartimento di diagnostica e sanità pubblica

Dottorato di ricerca in nanoscienze e tecnologie avanzate
Curriculum in Nanomedicina Morfologica Clinica
XXIX Ciclo

The modular structure of brain functional connectivity networks: a graph theoretical approach

S.S.D. FIS/07

Coordinatore:

Franco Tagliaro

Tutor:

Angelo Bifone

Pasquina Marzola

Dottorando:

Carlo Nicolini

Keywords:
Brain networks, modularity, community detection, functional connectivity, Surprise.

A.A. 2014-2017

CONTENTS

ACKNOWLEDGMENTS

SYMBOLS AND NOTATIONS

V	A set of n vertices (or nodes)
E	A set of m edges (or links)
$G = (V, E)$	A graph with its vertices and edges
n	The number of nodes in G
m	The number of edges in G
p	number of vertex pairs $p = \binom{n}{2}$
$\mathbf{A}$	Adjacency matrix of a graph
$\mathbf{W}$	Weighted adjacency matrix of a graph
C_n	Cycle graph with n vertices
K_n	Clique or complete graph with n vertices
$H(\sigma)$	Hamiltonian (energy) of a partition
$\Pr(\cdot)$	Probability distribution
Q^N	Newman's Modularity function
S	Surprise function
$\hat{S}$	Surprise function $(-\log_{10}(S))$
S_a	Asymptotical Surprise function
$\delta(i, j)$	Kronecker delta
$\langle \cdot \rangle$	Expectation value (average)
P_{ij}	Null model probability of an edge
$\boldsymbol{\sigma}$	Community membership vector
σ_i	Community membership of vertex i
k, k^{out}, k^{in}	undirected, outgoing and incoming degree
k^{int}, k^{ext}	internal and external degree
s, s^{out}, s^{in}	undirected, outgoing and incoming strength
s^{int}, s^{ext}	internal and external strength
$\Gamma(u)$	Neighboring vertices of vertex u
$H(X)$	Shannon Entropy of X
$H(X|Y)$	Conditional Entropy
$NMI(X, Y)$	Normalized Mutual Information

ACRONYMS

BOLD	Blood Oxygen Level Dependent
CPM	Constant Potts Model
ER	Erdos-Renyi
FC	Functional Connectivity
LFR	Lancichinetti-Fortunato-Radicchi model
fMRI	Functional Magnetic Resonance Imaging
RB	Reichardt-Bornholdt
TMS	Transcranial magnetic stimulation

SOMMARIO

Fin dal loro inizio, le neuroscienze hanno investigato il cervello a diverse scale, partendo dalla neurobiologia molecolare alla neuroanatomia. A tal fine si è compreso che per ottenere una comprensione profonda del funzionamento del cervello, è necessario un approccio multiscala in cui gli elementi costitutivi vengano valutati nelle loro interazioni a diversi livelli.

Fra i metodi di indagine neuroscientifica più interessanti, la risonanza magnetica funzionale (fMRI) ha dimostrato che è possibile misurare in tempo reale i cambiamenti nel flusso ematico locale legato all'aumento del metabolismo neuronale. Questa disciplina ha preso il nome di neuroimaging funzionale in quanto permette di quantificare le attività cerebrali in relazione alle loro funzioni neurofisiologiche e cognitive. In particolare, attraverso misure della cosiddetta connettività funzionale è possibile quantificare la correlazione temporale di eventi neurofisiologici in aree neurali spazialmente remote, esprimendo la loro interdipendenza come deviazione statistica rispetto all'attività di regioni distribuite sull'intera corteccia.

Diversi approcci di analisi multivariata sono applicabili allo studio dei dati ottenuti tramite fMRI, come l'analisi delle componenti indipendenti (ICA) o metodi seed-based. Ad oggi però, la teoria delle reti complesse, un approccio multidisciplinare di cui la teoria dei grafi è il fondamento teorico, ha offerto gli strumenti più avanzati per indagare la complessità del cervello a diverse scale. Secondo questo approccio, il cervello viene considerato come un grafo i cui nodi rappresentano singole regioni cerebrali, le quali possono ricoprire un singolo voxel o arrivare a rappresentare una regione anatomicamente definita. Gli archi del grafo invece esprimono numericamente il livello di interazione fra regioni.

Partendo da tali premesse, la moderna teoria delle reti complesse ha dimostrato che dal punto di vista funzionale, la corteccia cerebrale umana sana è organizzata secondo un principio di architettura ad invarianza di scala, caratterizzata da una combinazione di dense interazioni locali ed interazioni a lungo raggio, critiche al fine di un corretto meccanismo di scambio dell'informazione. Funzionalmente, tale struttura si poggia su un insieme di nodi molto connessi che agiscono da "hubs" favorendo processi di integrazione dell'informazione elaborata a livello locale.

Contemporaneamente, le reti cerebrali mostrano, sia a livello strutturale che funzionale, una struttura modulare organizzata gerarchicamente che ne incrementa la robustezza a perturbazioni esterne, quali insulti fisici, patologie o alterazioni durante il neurosviluppo. L'architettura modulare di queste reti funzionali, spiega la capacità del cervello di elaborare una grande quantità di input paralleli provenienti in ogni momento dai vari organi di senso, in quanto ognuno di questi moduli è altamente efficiente nel suo compito e le connessioni fra moduli consentono un efficace scambio dell'informazione.

I moduli, o comunità, nelle reti di connettività funzionale, rappresentano sottogruppi di nodi che sono strettamente connessi fra di loro più di quanto non lo siano rispetto a nodi esterni. In altri termini i moduli si possono intendere come

aree per cui la dipendenza statistica della loro attività temporale è più elevata che rispetto ad altre regioni.

Diversi sistemi complessi in natura esibiscono un'architettura modulare di questo tipo e la teoria delle reti complesse fornisce potenti strumenti per la loro indagine, ispirati per la maggior parte alla fisica dei sistemi disordinati. A questo scopo, la specialità della teoria delle reti che si occupa di identificare i moduli costitutivi dei sistemi complessi, è nota in letteratura con il nome di "community detection".

Non tutti gli algoritmi di community detection proposti negli anni sono però capaci di risolvere la struttura modulare delle reti funzionali, mettendo in luce la presenza di moduli e sottomoduli a diverse scale. Uno dei principali problemi che limitano la capacità di individuazione delle cosiddette comunità, è infatti il cosiddetto "resolution limit". Questo problema piuttosto generale e passato inosservato nella letteratura neuroscientifica, condiziona la capacità dei metodi di community detection di individuare moduli le cui dimensioni siano minori di una scala determinata dalla dimensione complessiva della rete. Precisamente, i metodi di community detection più ampiamente utilizzati sono basati sulla funzione Modularity introdotta originalmente da Mark Newman. L'ottimizzazione di questa funzione, vincolata dal limite di risoluzione, fornisce una visione ristretta sulle comunità corticali, mostrando che quasi indipendentemente da ogni altro parametro, la corteccia umana sarebbe suddivisa funzionalmente in un numero variabile da quattro a sei moduli, a seconda degli studi. Questo effetto artefattuale ha non solo inficiato il potere risolutivo dei metodi di community detection ma ha anche diminuito la capacità di effettuare discriminazioni di piccole differenze che possono portare ad utili diagnosi riguardanti alterazioni della connettività.

In questo lavoro mi sono occupato esplicitamente di analizzare, tramite la teoria delle reti complesse, gli effetti del limite di risoluzione sullo studio della modularità delle reti di connettività funzionale e di fornire una soluzione che permetta una descrizione multiscala dei moduli di queste reti.

Nel primo capitolo il lettore viene guidato alla teoria delle reti complesse, attraverso un'introduzione ai metodi di neuroimaging ed all'applicazione della teoria dei grafi all'analisi dei dataset fMRI. Successivamente nel secondo capitolo ho trattato nel dettaglio le caratteristiche del problema del limite di risoluzione, sia in relazione al suo aspetto teorico che a quello pratico sulle reti di connettività funzionale.

Nel terzo capitolo ho investigato gli effetti del resolution limit negli studi di neuroimaging, trovando che una varietà di studi dove la teoria delle reti è stata applicata a dataset di connettività funzionale, mostra i chiari effetti di tale problema. Per questo motivo ho sfruttato un nuovo metodo basato sull'ottimizzazione numerica della funzione costo nota come Surprise, centrale per questo lavoro, studiandone le proprietà teoriche in dettaglio e trovando che essa non soffre del limite di risoluzione. I risultati ottenuti dalla massimizzazione di Surprise su dati reali, hanno mostrato che dal punto di vista funzionale le reti di connettività funzionale esibiscono una struttura modulare caratterizzata da moduli di dimensione eterogenea, differentemente da quanto osservato utilizzato approcci a risoluzione limitata. Alla luce della diversa e più fine partizione in comunità perciò ho rivalutato il ruolo dei singoli nodi all'interno della rete, facendo emergere l'indicazione che alcune aree, in particolare il precuneo, svolgono un ruolo

fondamentale nell'integrazione dell'informazione fra moduli. Quest'ipotesi risulta consistente con il ruolo del precuneo come area integrativa di un ampio spettro di funzioni, dalle abilità visuo-spaziali al recupero di memorie episodiche.

In un secondo studio, contenuto nel capitolo quarto, ho esteso a reti pesate la tecnica prima sviluppata. Ho prima testato la validità del nuovo approccio detto Asymptotical Surprise su reti sintetiche dove ho simulato gli effetti dei fattori di disturbo in fMRI, variando il rapporto segnale rumore ed il numero di partecipanti in un esperimento virtuale. Successivamente ho dimostrato che similarmente a Surprise, anche l'ottimizzazione di Asymptotical Surprise permette sia di individuare moduli di diversa scala su reti di connettività funzionale umana che di rivalutare il ruolo dei singoli nodi all'interno della struttura modulare in esame.

Come ultima applicazione, ho voluto dimostrare con uno studio preliminare, come la disponibilità di un metodo di community detection a maggiore risoluzione permetta di discriminare alterazioni delle struttura modulare della connettività funzionale di pazienti schizofrenici rispetto a soggetti di controllo. Nello specifico, ho notato come le corteccie primarie (percettiva e motoria) mostrino una disgregazione in sotto-moduli indipendenti. Quest'osservazione risulterebbe in accordo con una teoria della schizofrenia bottom-up che vede questa devastante malattia come un disordine cognitivo originato da un deficit negli stadi iniziali dell'elaborazione sensoriale.

In definitiva, il limite di risoluzione sembra avere mascherato in maniera artefattuale la meravigliosa complessità delle reti di connettività funzionale e lo studio qui presentato è uno dei primi che ha cercato di svelare le cause di questa apparente cecità.

PUBBLICAZIONI

- Nicolini C., Bifone A., "Modular structure of brain functional networks: breaking the resolution limit by Surprise", Scientific Reports, 6, 19250 (2016).

- Nicolini C., Bordier C., Bifone A. "Community detection in weighted brain connectivity networks beyond the resolution limit". Neuroimage, 146, 28-39 (2017).

- Bordier C., Nicolini C., Bifone A. "Graph analysis and modularity of brain functional connectivity networks: searching for the weakest link". Frontiers in Physics (2017), submitted.

- Bordier C., Nicolini C., Bifone A. "Disrupted modular organization of primary sensory areas in schizophrenics". (in preparation).

CONFERENZE

Il lavoro presentato in questa tesi è stato presentato in forma orale e di poster presso diverse conferenze nazionali ed internazionali.

- Carlo Nicolini, Cècile Bordier, Angelo Bifone, "Community detection in brain functional networks beyond the resolution limit", *Conference of the Complex Systems Society*, Amsterdam, The Netherlands, September 19-22, 2016. Presentazione orale.

- Carlo Nicolini, Cècile Bordier, Angelo Bifone, "Community detection in resting state functional connectivity networks beyond the resolution limit", AISMRM Annual meeting 2016, Bologna, Italy, February 04-05, 2016. Poster.

- Carlo Nicolini, Cècile Bordier, Angelo Bifone, "Community detection in weighted brain connectivity networks beyond the resolution limit", *Pattern Recognition in Neuroimaging*, Trento, Italy, June 20-24, 2016. Poster.

- Carlo Nicolini, Cècile Bordier, Angelo Bifone, "Modular organization of resting state functional connectivity networks: breaking the resolution limit by Surprise", *Resting state conference*, Vienna, Austria, September 21-23, 2016.

ABSTRACT

Complex networks theory offers a framework for the analysis of brain functional connectivity as measured by magnetic resonance imaging. Within this approach the brain is represented as a graph comprising nodes connected by links, with nodes corresponding to brain regions and the links to measures of inter-regional interaction. A number of graph theoretical methods have been proposed to analyze the modular structure of these networks. The most widely used metric is Newman's Modularity, which identifies modules within which links are more abundant than expected on the basis of a random network. However, Modularity is limited in its ability to detect relatively small communities, a problem known as "resolution limit". As a consequence, unambiguously identifiable modules, like complete sub-graphs, may be unduly merged into larger communities when they are too small compared to the size of the network. This limit, first demonstrated for Newman's Modularity, is quite general and affects, to a different extent, all methods that seek to identify the community structure of a network through the optimization of a global quality function. Hence, the resolution limit may represent a critical shortcoming for the study of brain networks, and is likely to have affected many of the studies reported in the literature. This work pioneers the use of Surprise and Asymptotical Surprise, two quality functions rooted in probability theory that aims at overcoming the resolution limit for both binary and weighted networks. Hereby, heuristics for their optimization are developed and tested, showing that the resulting optimal partitioning can highlight anatomically and functionally plausible modules from brain connectivity datasets, on binary and weighted networks. This novel approach is applied to the partitioning of two different human brain networks that have been extensively characterized in the literature, to address the resolution-limit issue in the study of the brain modular structure. Surprise maximization in human resting state networks revealed the presence of a rich structure of modules with heterogeneous size distribution undetectable by current methods. Moreover, Surprise led to different, more accurate classification of the network's connector hubs, the elements that integrate the brain modules into a cohesive structure. In synthetic networks, Asymptotical Surprise showed high sensitivity and specificity in the detection of ground-truth structures, particularly in the presence of noise and variability such as those observed in experimental functional MRI data. Finally, the methodological advances hereby introduced are shown to be a helpful tool to better discern differences between the modular organization of functional connectivity of healthy subjects and schizophrenic patients. Importantly, these differences may point to new clinical hypotheses on the etiology of schizophrenia, and they would have gone unnoticed with resolution-limited methods. This may call for a revisitation of some of the current models of the modular organization of the healthy and diseased brain.

1 | INTRODUCTION

1.1 THE RISE OF NETWORK NEUROSCIENCE

Starting from 19th century's neuron doctrine, the brain is described as an incredibly complex system of intercommunicating elements, the neurons, intertwined with other cells that support neuronal activity.

The dynamical processes running on the biological substrate of neurons are the basis of information processing that gives rise to cognitive and psychological processes, from the perception of the external environment to the representation of mental states and emotions. Neuroscience tackles the investigation of the brain, the most complex organ in nature, at different scales. Molecular and cellular neuroscience focus on the finest level of workings of the nervous system cells. At this scale, molecular biology and genetics are used to address the study of very small populations of neurons.

At larger scales, hundreds or thousands of neurons are involved and a more comprehensive approach is needed. The view must shift from the microscopic level to the so-called systems neuroscience, which deals with widespread networks of neurons, dubbed neuronal circuits.

System neuroscience studies how these large-scale circuits form and evolve during development, how they produce basic processes like reflexes, motor coordination, circadian rhythms and, ultimately, how they give rise to behavioural and different internal states like motivation, emotions, and cognition. Hence, an extension of system neuroscience from the study of small groups of neurons to the whole brain requires radically different tools, both computational and experimental. Very small scale neuronal circuits can be studied by mathematically accurate models. The typical size of the smallest identified functional unit, the cortical column, is around hundred thousand neurons and, already at that level, many hypotheses are required to approximately reproduce the electrical and chemical activity of neurons, at the point that a comprehensive differential equations model of cortical columns is still lacking.

Yet, a large-scale understanding of the incredibly complicated network of all the neurons in the human brain, the *human connectome*, remains a challenging target to be reached in the future [1]. This scientific, technological and cultural challenge is named after the term *connectomics*, a word coined at the beginning of this decade by Giulio Tononi and Olaf Sporns [2]. Its most ambitious aim is to map all the connections in the brain at different scales of resolution, deep down to the synaptic level. Whether this will help answer some of the most profound questions about the human brain, it is not entirely known and still a topic of hot debate in the scientific community.

The science of brain connectivity takes advantage of well-consolidated technologies for looking at the brain at work. These non-invasive methods of analysis of

brain functions include neuroimaging techniques like magnetic resonance imaging (MRI), electroencephalography (EEG) and transcranial magnetic stimulation (TMS). Among these modalities, MRI has emerged as a dominant non-invasive imaging method because it enables the investigation of both brain structure and function with a good compromise in terms of spatial and temporal resolution.

At the morphological level, MRI provides a comprehensive view of the brain, as it can identify the white matter fiber pathways connecting gray matter regions otherwise impossible to obtain in-vivo. The pattern of white matter fibers is often named after the term *structural connectivity* as it relates to the morphological features of the brain connections. Structural networks correspond to a pattern of anatomical connections, summarizing bundles of axons connecting different brain regions. Borrowing ideas from the electronics, structural connectivity may be interpreted as the electrical wiring of the different parts of a processor. Although useful, the wiring diagram of the brain is only a starting point for making hypotheses about the brain at work; indeed it cannot reveal how neurons behave in real-time, nor does it account for the yet unknown regulatory mechanisms that neurons exert on each other. Unfortunately, these dynamics are and remain partially unpredictable, despite many notable and useful attempts [3], as also the synaptic weights are changing dynamically due to synaptic plasticity.

The discovery of the BOLD effect in the 90's by Ogawa [4], led to functional MRI, a method to spatially map activity in the whole brain. Considered over the large scale, the BOLD response is descriptive of the temporal metabolic demands of large groups of cortical and subcortical structures, a quantity strictly related to the electrical activity of neurons [5]. Indeed an active exchange of information between neural populations is thought to be at the heart of the BOLD fMRI signal [5]. The local increase of metabolic demand for oxygenated blood to fuel the neuronal activity increments the vascular flow and favors the consumption of oxygen from neurons. This metabolic cascade results in the change of the magnetic properties of hemoglobin that transitions from oxygenated to deoxygenated state, yielding local magnetic susceptibility variations measured by MRI. Possible confounds of the fMRI signal, such as the influence of the cardiac and respiratory rhythms on the BOLD oscillations have been ruled out by a number of studies [6, 7, 5, 8].

The tools offered by fMRI helped scientists to unveil another pattern of connectivity that exists in the brain. *Functional connectivity* examines functional associations between temporal fluctuations of neuronal activities from different brain regions [9]. Activation maps have been estimated for almost every kind of cognitive task and collected in publicly available datasets, like the BrainMap database [10], the Neurovault repository [11] and the UCLA multimodal connectivity database [12]. In contrast to structural connectivity, functional connectivity results from estimates of statistical dependencies between neuronal or regional time courses. Functional networks may vary over time, unlike structural networks which are stable over long time scales. For instance, electrical patterns of cortical activity may rearrange over time-scales of milliseconds, in contrast to structural ones, involved in processes like nervous system development, learning, and aging. Furthermore, the temporal evolution of functional networks is variable as it exhibits spontaneous changes during rest and characteristic modulations for different task conditions [13, 14].

When cognitively at rest, the brain engages in a characteristic pattern of dynamic neural activity and most brain regions show spontaneous oscillations in the BOLD signal [9, 15, 16]. In the resting state, a subject is quietly awake and alert but does not engage in or attend to any specific cognitive or behavioral task. Interestingly, this background activity dubbed *resting-state connectivity* is not extinguished but only attenuated during tasks [17], making evident its importance in response to changes in the external demands. During a typical resting state fMRI (rs-fMRI) session, volunteers are left in the MRI scanner, instructed to relax and not to think anything in particular. On the other end, animals like mice or rats, are lightly sedated and the BOLD signal is measured for longer time intervals, yielding more stable estimates [18]. In humans, the fluctuations of resting-state connectivity typically exhibit the highest correlation at low temporal frequencies (<0.1 Hz) and can be observed during alertness [19], sleep [20], light sedation [21] and general anesthesia [22]. The structure of temporal correlations of the BOLD signals defines a network over spatially remote brain regions. Interestingly, these networks are altered in neuropsychiatric diseases like autism [23] and schizophrenia [24]. The extent of these alterations may be a marker for possible diagnoses of brain disease [25, 26].

Complex network science provides a framework to study the structure of functional connectivity networks and to assess alterations induced by pathological conditions [27, 25, 28, 26]. The methods of network theory date back to the problem of Königsberg bridges, in 1735, when the famous mathematician Leonhard Euler was faced with the problem of finding a path through the city that would cross all the seven bridges just once. The knowledge accumulated since then forms the basis of modern graph theory. In totally general settings, simple graphs are usually defined as a set of nodes, connected by a set of links, denoting the strength of the relation between pairs of nodes. Graphs can be used to represent any type of relation between interacting elements and the idea that graph-theory can be applied to brain networks has been pioneered starting from the middle nineties defining a new specialization of systems biology, dubbed *network neuroscience*.

The network neuroscience relies upon the treatment of the brain as a graph, whose nodes are indicative of particular brain regions connected by links representing the strength of their mutual interactions. The classical general pipeline for the definition of a brain graph is given in Figure 1.1. In this framework, the modelling of the brain in terms of graphs starts from the definition of its nodes [29]. These can be defined as anatomically or functionally defined regions [30]. A clear delineation of network nodes is one of the first steps in the whole pipeline of the graph-theoretical analyses and is highly important for the subsequent analyses [29]. The second step in the definition of a brain network is an estimate of the continuous association between nodes: the links of the graph indeed measure the interregional coordination or similarity between nodes [31]. To be more specific, in the case of fMRI studies of resting state functional connectivity, link weights are defined as interregional temporal correlations in the fluctuations of the BOLD signals, and the resulting graph can be represented by a square matrix (adjacency matrix) where the strongest correlations indicate the presence of a link between any two specific areas. In other applications, like Diffusion Tensor Imaging, a technique that is able to show the effective directions of nervous fibers traits in the

brain, one can define links whose strength is directly proportional to the number of white matter tracts connecting any two regions. In addition, brain networks have also been defined on the basis of inter-subject anatomical covariance [32, 33], co-activation of different brain regions across individuals subjected to experimental tasks [34], pharmacological challenges [35, 36] or spectral coherence of EEG electrical time-courses [31]. All of these networks are "weighted" by definition, i.e. their links are associated with real numbers representing a measure of the strength of pairwise interactions between nodes.
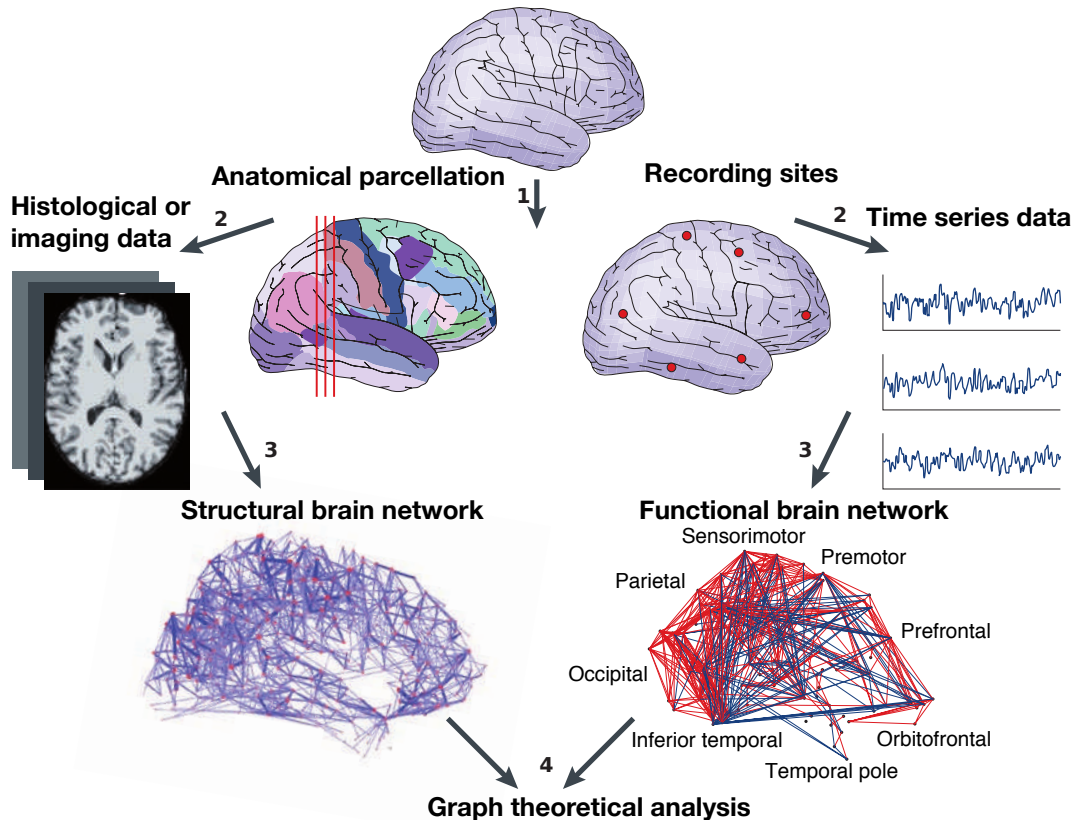
Figure 1.1: Structural and functional brain networks can be explored using graph theory through four steps. Adapted from [27].

The rise of network neuroscience in the last decade has been driven by the evidence that almost independently from how the network links are defined, brain networks share some common features at several spatial and temporal scales.

For example, they exhibit the *small-world* property [37, 38, 39], whereby most nodes in a graph can be reached from any other node in just a few hops. Small-worldness in brain networks [40] has been found to positively correlate with higher IQ in humans [41], suggesting that human intellectual performance may be related to how efficiently our brain integrates information between multiple brain regions.

Other measures have been correlated with phenotypical and structural characteristics. *Rich-clubs* are subsets of highly connected nodes who are also highly connected to each other, forming a communication backbone. It has been hypothesized, that rich clubs might be valuable in supporting high-level and more cognitively advanced forms of information processing [42]. Moreover, it is thought that the presence of a rich club increases the diversity of the functional repertoire

over and above the effects produced by scale-freeness alone [43], a property of a network of being self-similar at different scales. Forms of rich-club organization have been observed in cats, macaque, and humans [44, 45, 46, 42], comprising portions of prefrontal, parietal, temporal and insular cortex. Rich club nodes and edges participate in a large number of short paths across the network, contributing disproportionately to global communication.

From the functional point of view, central hubs of rich clubs are concentrated in heteromodal association cortices, areas that receive input from multiple unimodal sensory association areas and/or other heteromodal cortices [47, 48, 49]. Conversely, primary sensory cortices have low topological centrality and degree [44, 50, 45]. Damage to primary sensory areas only affects them locally, while perturbation to the rich-club backbone has diffuse and severe effects on the whole network architecture [51, 52]. Abnormal rich-club connectivity is hypothesized to be related to familial vulnerability for schizophrenia [53] and other diseases as well.

More generally, network neuroscience has matured the ability to map alterations at the level of structural or functional connectome and may be useful for predicting and generating hypotheses about underlying pathophysiological mechanisms of disease. To this end, the definition and validation of data-driven biomarkers based on graph-theoretical observations of the effects of brain diseases is a highly desirable objective.

1.2 BRAIN NETWORKS AT THE MESOSCALE

Many natural and artificial networks display a modular organization, i.e. nodes group together in some kind of *clusters*, resulting in a particular mesoscopic structure. At this level of mesoscale organization, the topological properties of nodes are more dependent on their neighbourhood than on the rest of the graph. When the structure tends to group nodes with similar properties, the mesoscale structure is called *assortative*. Graphs with an assortative structure consist of tightly connected groups of nodes that are themselves loosely connected with other groups. These groups are therefore dubbed as *communities*.

The organization into communities is one of the most studied properties of networks and a large amount of literature has grown on the topic, as reviewed in [54]. The problem of detecting the communities, *community detection* or *graph-partitioning* in the computer-science jargon, is of high theoretical and practical importance. From the theoretical point of view, communities are representative of the generative probabilistic model underlying the formation of links [55] and tell much about the statistical properties of networks. On the other side, practical applications of community detection are fundamental not only in neuroscience but also in sociology, computer science, biology and many other fields. Modules derive from a decomposition of the network into subcomponents that are internally strongly coupled but externally only weakly coupled. The property of a system of being modular or *nearly-decomposable*, as already noted by Nobel prize laureate Herbert Simon [56, 57], helps any system to resist external perturbations by separating different functions into topologically distinct part of the system.

Near decomposability is in many respects a hallmark of complexity in natural systems [56].

Strong experimental evidence shows that the nervous system in animals and humans, from the neuronal level up to macroscopic scale, displays an organization that favours the exchange of information over a communication backbone supported by well-connected areas, the connector hubs of the network [46], responsible in turn for sending information to local processors, identified as the communities or modules [58]. Functional brain networks of humans and other animals consist of connected subnetworks (or modules) associated with cognitive functions, of vital importance for the daily life and the interaction with the environment. The modular organization is exhibited hierarchically at different scales, with modules inside modules. Within this architecture that confers robustness and adaptability to the network [59], information flows from lower level sensory cortices to higher multimodal and associative cortices [60], thus being refined and filtered in its content at multiple stages of elaboration. The specialized stages of neural information processing are the regions of *functional segregation*, a common feature at multiple scales in the brain, from the neuronal networks to macroscopic anatomically defined regions. From the functional point of view, segregation is meant as the statistical dependence of processing elements within the same region [61]. *Functional integration* instead is defined as a significant deviation from independence of large subsystems, and is typically realized through synchronization processes that increase the statistical dependence between functional modules, a mechanism required to access widely distributed resources, needed for a coherent behaviour [62]. A delicate balance between functional integration and segregation is realized at the edge of metastability as cortical areas are neither fully synchronized nor full desynchronized [63].

When the subtle balance between functional integration and segregation is altered, like in pathological states or for external insults, the brain responds in an adaptive manner, in order to maintain homoeostasis, where possible. The mechanisms of homoeostasis adopted are compensation, degeneracy and redundancy [26]. Through compensation, extra neural resources are recruited to maintain normal function and thanks to neural degeneracy and redundancy [64], it is possible for distinct regions to carry out the same function, to some extent. When these compensatory mechanisms set in after homoeostatic perturbations, the result is a reorganization of the brain structure, involving changes in its network properties. Among the network level changes observed after specific alterations, those relative to the brain modular structure can be investigated within the framework of complex networks science [65, 66, 67].

Using the tools of complex network science, the evaluation of the modular structure of FC networks is cast to the problem of community detection in graphs, i.e. the problem of identifying independent groups of brain regions that exhibit strong interconnection between each other but that are weakly connected to other modules.

Several methods have been proposed to resolve the community structure of complex networks [54, 69, 70]. Many of these methods involve the definition of a fitness function that assigns positive or negative scores to edges connecting nodes within or outside the same community, and heuristics to find the optimal partition
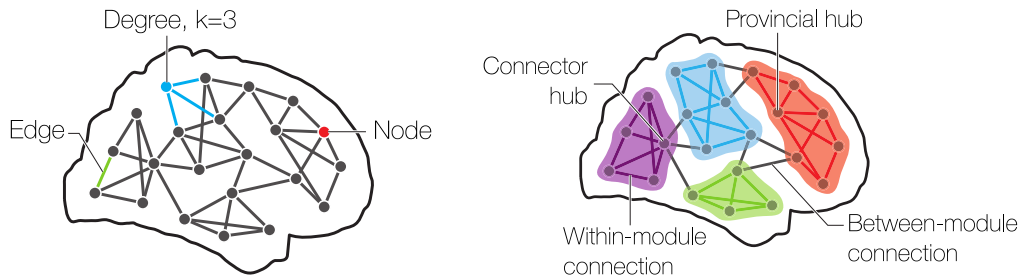
Figure 1.2: Meaning of communities, hubs and connectors of a brain network, depicted in an artistic view. Adapted from [68].

of the network that maximizes this fitness function. Among all the methods, the most popular approach is Newman's "Modularity maximization" and variations thereof [71]. Following the first demonstration by [36], partitioning of brain networks using Newman's Modularity has been widely applied to assess the brain modular structure. Despite its wide usage and merits, Newman's Modularity is characterized by some important limitations as it may fail to detect modules smaller than a certain scale determined by the network size, resulting in a few, relatively uniform modules of almost the same dimension. This general behaviour, not unique to Newman's approach, is pervasive in community detection and is dubbed as *resolution-limit* [72].

As a consequence, even unambiguously defined modules, like almost complete sub-graphs or even cliques[1], may be unduly merged into larger communities when they are too small compared to the size of the network. This limit hampers the ability to detect small-scale structures of functionally correlated brain areas and to notice alterations in connectivity between subjects or groups of subjects. Hence, it may have prevented detection of alterations in the modular structure of brain connectivity in patients affected by various neuropsychiatric disorders Hence, the resolution limit represents a critical shortcoming for the study of brain networks and is likely to have affected many of the studies reported in the literature.

In the present work I have developed an approach that is valuable in detecting communities in brain networks beyond the resolution limit. The following sections of this introductory chapter guide the reader into the complex networks science, by giving an overview of the mathematical terminology and methodological concepts. In the second chapter, the experimental evidence of the modular structure of brain connectivity is reviewed, together with the methods to disentangle the mesoscopic structure of complex systems. In particular, sections 2.3 and 2.4 consider the problem of the resolution limit and degeneracy of Newman Modularity. A new approach, dubbed Surprise, that overcomes the limitations imposed by the resolution limit and degeneracy, is described in the chapter 3. Throughout chapters, the problems imposed by the resolution limit and degeneracy are analysed for Surprise and its weighted counterpart, Asymptotical Surprise. The characteristics of the two methods are discussed, together with an efficient algorithm aimed at their optimization. The application of Surprise and Asymptotical Surprise to functional connectivity networks constitutes the heart of chapters 3 and 4, whereby the implications of a resolution limit free analysis of the functional modules are

---

1 Graphs where all nodes to connect to all nodes

explored. Throughout the third and fourth chapter, an approach for the creation of benchmark networks is then proposed, whereby both Surprise and Asymptotical Surprise are validated and shown to outperform current resolution-limited methods.

The lack of resolution limit of Surprise motivated an application on real-world resting state networks that revealed the presence of a rich structure of heterogeneously distributed modules, and differences in networks' partitions that are undetectable by resolution-limited methods. Furthermore, after validation in the synthetic networks with simulated noise as from fMRI studies, Asymptotical Surprise optimization is applied to real world functional connectivity networks and is shown to lead to substantial differences in the identification of connector hubs compared to other community detection methods.

Finally, in the last chapter of this thesis I discuss the advantages of using Asymptotical Surprise optimization to detect the reorganization of the modular structure in the diseased brain. In particular, I found that the higher specificity provided by a resolution-limit-free method highlights a modular fragmentation of the primary sensory areas in schizophrenic patients with respect to healthy controls. This is a preliminary finding although in keeping with recently posited theories about the aetiology of schizophrenia, that hypothesize an alteration in early sensory processing as the main cause of the cognitive and behavioural deficits often observed in this disease.

1.3 ELEMENTS OF GRAPH THEORY

A short and self-contained tutorial to the jargon of graph theory is in order before delving into the details of community detection and the illustration of how the resolution limit negatively affected many neuroscientific studies: this section is devoted to the introduction to the mathematical concepts of graph theory. We adopt the typical notation used in other works of network science [73, 74].

• A *graph* $G = (V, E)$ is a representation of a set V of n nodes, also called *vertices*, connected by m links, also called *edges*, in a set E (Figure 1.3a).

• A graph with no multiple links and no self-loops is dubbed *simple graph*. If multiple links exist between two nodes, the graph is called *multigraph*. Additionally, if a node links itself, the graph is called *loopy graph*. Figure 1.3c shows the combination of a multigraph and a loopy graph: a loopy multigraph. If the direction of the links is important, the graph is called *directed*, as shown in Figure 1.3d, otherwise the graph is called *undirected*.

• The adjacency matrix $\mathbf{A} = \{A_{ij}\} \in \{0, 1\}^{n \times n}$ of a binary undirected graph is a square $n \times n$ symmetric matrix with elements $A_{ij} = 1$ if an edge exists between vertex i and j and 0 otherwise.

• An *edge-weighted graph* (also called weighted graph) $G = (V, E, \omega)$ is a graph that has a set of weights ω on the links (Figure 1.3b). Although not true in general, we consider only weighted undirected graphs with a symmetrical real adjacency matrix, meaning that if an edge exists between node i and j the weight from i to j is the same as the weight from j to i: $\omega_{ij} = \omega_{ji}$. The weighted adjacency matrix, indicated as $\mathbf{W} = \{\omega_{ij}\} \in \mathbb{R}^{n \times n}$ is a real, square $n \times n$ symmetric matrix.
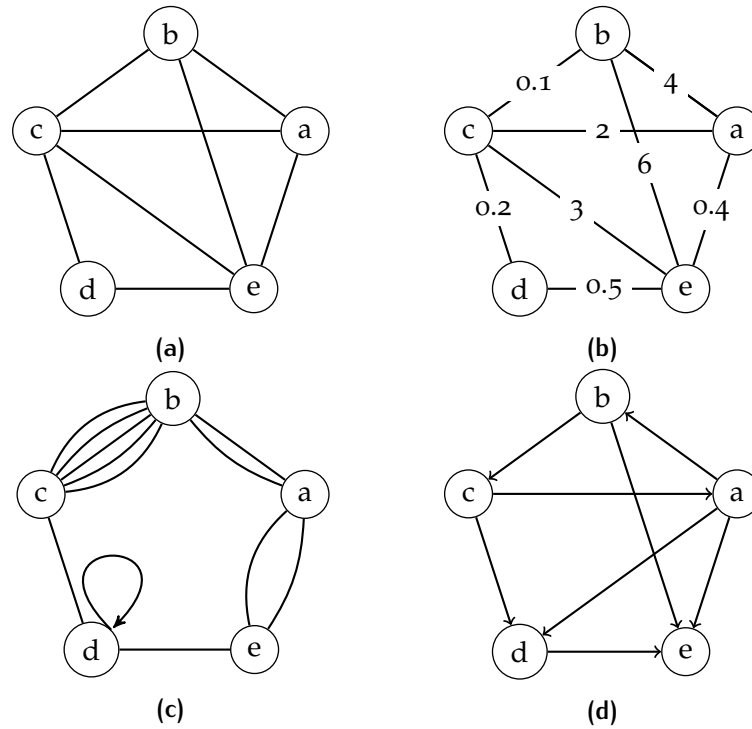
Figure 1.3: Classes of graphs. **(a)** is a simple unweighted undirected graph. **(b)** is a simple weighted undirected graph, **(c)** is an unweighted loopy multigraph, **(d)** is a simple directed graph.

• The *density* of a simple graph ρ is defined as the ratio of the actual number of edges over all possible edges:

$$\rho = \frac{m}{\binom{n}{2}} = \frac{2m}{n(n-1)}. \tag{1.1}$$

• A *dense* graph is a graph where almost every possible pair of nodes is connected with an edge. Conversely, a graph is termed *sparse* if the density is low, meaning that also the adjacency matrix is sparse and the number of edges is in the order of magnitude of the number of nodes $m \approx \mathcal{O}(n)$.

• The number of edges incident to a node in a simple graph is called *degree*, denoted as k_i. Every half edge incident to a node is called *stub*. A vertex of degree k_i has therefore k_i incident stubs.

• On simple weighted graphs, the sum of weights of the edges incident to a vertex i is called *strength* and is denoted as s_i. Degree and strength are the sums over rows of the adjacency matrix $d_i = s_i = \sum_{j=1}^{n} A_{ij}$, they are equal for binary graphs, but different for weighted graphs.

• The *degree sequence* $\{k_i\} \forall i = 1, \ldots, n$ is the sequence of the degrees of the nodes, with these numbers put in ascending order, and repetitions if needed. By the *handshaking lemma* [75], the sum of all node degrees is equal to twice the number of links:

$$\sum_{i=1}^{n} k_i = 2|E|, \tag{1.2}$$

consequently the *average node degree* is given by

$$\langle k \rangle = \frac{2m}{n}. \tag{1.3}$$

• The *neighborhood* of node i is the set $\Gamma_i = \{j \in V | (i,j) \in E\}$. In other words, the neighbor nodes of i are the nodes which share and endpoint edge to i.

• A *path* in a graph is a sequence of edges such that the destination of each edge in the sequence is always the source of the following edge. A *cycle* is a closed path, i.e. a path where the origin and destination nodes are the same.

• A graph that has all the possible links, indicated by $K = (V, V \times V)$ is called *complete graph* or *clique* (Figure 1.4a). Complete graphs on n nodes, indicated as K_n have a total of $m = \binom{n}{2}$ edges and density equal to 1.

• The complementary graph $\bar{G}$ is the graph that has edges where G has no edges and vice-versa, formally $\bar{G} = (V, V \times V - E)$. The *empty graph* is therefore defined as the complementary graph of the clique $\bar{K}$, as it has n nodes and no edges (Figure 1.4b). The *cycle graph* C_n is a graph containing a single cycle through all nodes and is the smallest connected cyclic graph (Figure 1.4c). A *tree graph* is a graph with no cycles and exactly $n - 1$ edges.

(a) Complete graph K_5 **(b)** Empty graph $\bar{K}_5$ **(c)** Cycle graph C_5
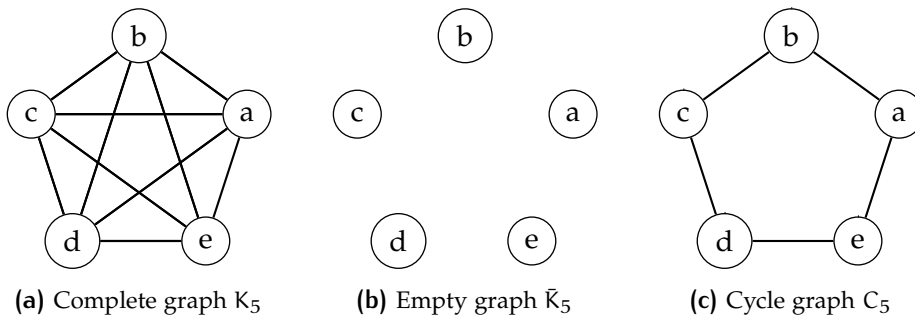
• A subgraph $\mathcal{G}$ of a graph G is said to be *induced* if, for any pair of vertices i and j of $\mathcal{G}$, the pair (i,j) is an edge of $\mathcal{G}$ if and only if (i,j) is also an edge of G. In other words, $\mathcal{G}$ is an induced subgraph of G if it has the most edges that appear in G over the same vertex set. If $\mathcal{G}$ is chosen based on a vertex subset S of $V(G)$, then $\mathcal{G}$ can be written as $G[S]$ and is said to be induced by S.

• A simple graph is said *connected* if there exists a path between any pair of nodes. If the graph itself is not connected, then it is formed by a set of connected subgraphs, also called *weakly connected components*.

1.3.1 Clustering

Grouping nodes according to some criterion of similarity between them is the basis of the identification of clusters in networks.

• A set of mutually disjoint, induced subgraphs that covers all the nodes is called a *clustering*. A clustering $\zeta = \{\zeta_c\}$ of G is a partitioning of the set V into C disjoint sets of nodes, $\zeta_c \subseteq V$, which we call *modules* or *communities*, interchangeably. Each module is a node-induced subgraph $\mathcal{G} := (V[\zeta_c], E[\zeta_c])$, also indicated as $G[\zeta_c]$.

• We denote the number of nodes of the module $G[\zeta_c]$ with n_c, the number of edges with m_c and the total number of pairs of nodes with $p_c = \binom{n_c}{2}$. From

the notational point of view, a clustering can alternatively be defined with a *node assignment vector* $\sigma \in \mathbb{N}^n$ in which every node is assigned to an integer label representing the community index.

• The *Kronecker delta* on the assignment vector $\delta(\sigma_i, \sigma_j) = 1$ indicates when two nodes lie in the same module and $\delta(\sigma_i, \sigma_j) = 0$ when they belongs to different modules.

• The *internal degree* $k_{int}(i)$ of vertex i in the module $\mathcal{G}$ is the number of edges connecting i to other vertices in $\mathcal{G}$. If $k_{ext}(i) = 0$, the vertex has neighbors only within $\mathcal{G}$, on the other hand if $k_{int}(i) = 0$, the vertex is disjoint from $\mathcal{G}$ and should not be part of the same community.

• The *external* degree $k_{ext}(i)$ of vertex i in the module $\mathcal{G}$ is the number of edges connecting i to vertices not in $\mathcal{G}$.

• The *subgraph internal degree* $K_{int}(\mathcal{G})$ is the sum of the internal degrees of its vertices. Likewise, the *subgraph external degree* $K_{ext}(\mathcal{G})$ is the sum of the external degrees of its vertices.

• The *subgraph total degree* $k(\mathcal{G})$ is the sum of the degrees of the vertices of $\mathcal{G}$. By definition, $K(\mathcal{G}) = K_{ext}(\mathcal{G}) + K_{int}(\mathcal{G})$. For convenience of notation we will denote without loss of generality $K_{int}(\mathcal{G}[\zeta_c])$ as K_c, specifying the full notation where is needed.
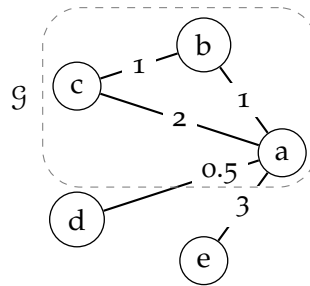
Figure 1.5: Internal degree $k_{int}(a) = 2$, internal strength $s_{int}(a) = 3$, external degree $k_{ext}(a) = 2$, external strength $s_{ext}(a) = 3.5$.

• The *intracluster density* of the subgraph $\mathcal{G}$ is defined as the ratio of the intracommunity edges over the number of pairs of nodes

$$\rho_{int}(G) = \frac{m_c}{p_c} = \frac{2m_c}{n_c(n_c - 1)}. \tag{1.4}$$

• The *inter-cluster density* is instead defined as the ratio of edges with one endpoint in $\mathcal{G}$ and the other in $G - \mathcal{G}$, over the total possible number of edges connecting $\mathcal{G}$ with the rest of the graph:

$$\rho_{ext} = \frac{m - m_c}{n_c(n - n_c)}. \tag{1.5}$$

- The total intracluster edges and total intracluster node pairs numbers are the sum over all modules c of intracommunity edges and node pairs:

$$m_\zeta = \sum_{c=1}^{C} |E\,[\mathcal{G}_c]| = \sum_{c=1}^{C} m_c \tag{1.6}$$

$$p_\zeta = \sum_{c=1}^{C} \binom{|V\,[\mathcal{G}_c]|}{2} = \sum_{c=1}^{C} p_c. \tag{1.7}$$

1.3.2 Communities

The word *community* applied to complex networks originated in sociology as people tend naturally to form groups: family, friends, acquaintances. Social communities can then be thought as subgroups of people connected to each other more strongly than with other external people.

The problem of quantitatively defining a community, although intuitive at first sight, is multifaceted and ambiguous. No universally accepted definition of network community exists and some degree of arbitrariness or common sense is required, a feature that is difficult to describe quantitatively. Importantly, the identification of communities is possible only for relatively sparse graphs. On dense graphs indeed, the community detection problem requires different tools, closer to those of data clustering, with different concepts and definitions.

In the attempt of making the concept of community quantitative, two classes of definitions can be provided: local and global. The *local definitions* involve that a community is viewed as a separate entry or an autonomous module: the community is studied as isolated from the graph as a whole. These definitions imply that the corresponding communities are mostly maximal subgraphs, subsets of nodes that cannot be enlarged with the addition of new nodes or edges without altering the property upon which they are defined. The most important local criterion for communities definition are the *complete-mutuality*, the *reachability*, the comparison of vertex degree and *internal-external cohesion* [76, 77]. Complete mutuality means that the identified community is a clique where every node is mutually connected with all the others, while reachability implies that there exist a path between every pair of nodes in the community. The internal- external cohesion principle is equivalently represented by the Radicchi's definitions of *strong communities* [78]: for a community $\mathcal{G}$ to be defined in *strong* sense, every node in the community must exhibit a larger internal degree than external degree. This concept is quantified by the following definition:

$$\forall i \in V[\mathcal{G}] \qquad k_{int}(i) > k_{ext}(i). \tag{1.8}$$

The strong community criterion is extremely stringent and most real-world networks, even the simplest, do not exhibit such property. It is possible though to weaken this constraint, asking that *on average* the internal degree is greater than the external degree. Upon this idea the concept of *weak communities* is built (Eq. 1.9):

$$\sum_{i \in V[\mathcal{G}]} k_{int}(i) > \sum_{i \in V[\mathcal{G}]} k_{ext}(i) \tag{1.9}$$

Finding strong communities is more difficult than finding weak communities and most algorithms fail in this job. Most methods indeed encode their own concept of community. Nonetheless these two last definitions are almost universally accepted and give a precise idea of the meaning of community. Additionally, the weak criterion of community is the basis of the planted-partition model [79] and its later modifications [80] that we'll encounter in the next sections.

Most of the global definitions of communities encompass properties of the graph that are used by an algorithm to deliver a modular decomposition. In this sense the global definitions are always indirect, often based on *null models* or stationary properties of dynamical processes such as random walkers [81, 82].

Multiple different sub-networks in brain FC networks have been observed consistently with a variety of analysis techniques [83, 84, 85, 35] such as ICA or other multivariate methods. This empirical evidence, shown not only in humans but also in higher vertebrates as the cat [86] and macaque [87], suggests that to some extent the high-level functional organization is based on the balance of two competing principles of segregation and integration. The nervous system is obviously shaped by this balance as the need of fast information processing is satisfied by small-scale neuronal circuits, typically at the first stages of sensory input processing. Typical examples are the very localized functional area of visual processing V1 or the primary motor cortices. At a larger scale the cortex works in concert, connecting all the segregated areas in a widely distributed complex network. It is indeed more and more recognized that complex behaviours arise thanks to the dynamical interplay of the integrated and segregated actions of different regions of the brain [61, 62, 88].

The presence of communities within FC networks and hubs supports the concepts of integration and segregation. Recently, experiments in which computational dynamical processes are simulated over the topology of whole-brain networks have shed light on the role of communities in shaping the cortical activity [88]. Importantly, it was found that the network communities consist in the areas that are devoted to the segregated information processing, while the hubs and rich clubs are deputed to convey that information over the whole brain.

Assessing the role of nodes within a community is an important task that depends on the community assignment. To this end one can design a cartographic classification of nodal roles that describes if a node connects two communities or is buried inside its own community in of help.

1.3.3 Cartographic classification of nodes

Every node in a network is characterized by its role in relation to other nodes. Once a clustering ζ is imposed on a graph, some nodes may act as bridges between two or more communities, having neighbours in one or the other module, others instead may share the majority of neighbours inside their same community, being more central to their module.

To investigate differences in nodal roles as determined by different community detection methods, Guimera and Amaral [89] introduced a cartographic characterization of nodes based on two parameters. The first, dubbed *participation coefficient*,

indicated by P_i and comprised in the range $[0, 1]$ reflects the extent to which a node is connected to nodes in other modules, and is defined in Eq 1.10 as:

$$P_i = 1 - \sum_c \left(\frac{k_{ic}}{k_i} \right)^2 , \qquad (1.10)$$

where k_{ic} is the number of links of node i to nodes in module c and k_i is the total degree of node i. Converselely, the second parameter dubbed *within-module degree score*, indicated by z_i, reflects the degree of connectivity of a node to the nodes in its same community, and is defined as:

$$z_i = \frac{k_i - \langle k_{c_i} \rangle}{\sigma_{k_{c_i}}}, \qquad (1.11)$$

where $\langle k_{c_i} \rangle$ is the average degree of nodes in the same module c of node i and $\sigma_{k_{c_i}}$ is the standard deviation of degrees of nodes in module c.

Hubs classification is critical for the interpretation of the roles played by the highly connected nodes within the network structure. The Guimera and Amaral [89] classification scheme places participation coefficient and within-module degree z-score of every node on a two-dimensional plane. Different regions of the two-dimensional plane are divided empirically into seven regions as shown in Figure 1.6. Depending on the specific cut-off values of participation coefficient and intra-modular connectivity, the nodes are defined as ultra-peripheral (region R1), peripheral (region R2), connector non-hubs (region R3), kinless non-hubs (region R4), provincial hubs (region R5), connector hubs (region R6) and kinless hubs (region R7). Naturally, connector hubs are of great importance for communication between modules (Figure 1.2). Indeed, connector hubs are those with high values of participation coefficient and high values of within-module degree z-score.

In the context of brain networks, these hubs are responsible for the integration of the different modules into a cohesive structure and are thought to be particularly vulnerable in the presence of brain insult or disease, as damage to these nodes is likely to disrupt network connectedness [25, 26, 65]. In contrast, provincial hubs are more importantly involved in functional specialization and local processing; therefore damage to these areas may yield specific clinical deficits, whereas damage to connector hubs results in complex and pervasive neurological dysfunctions [25, 26].

1.3.4 Models of random graphs

A large effort of network science is devoted to the description of the generative processes that lead to the networks observed in real-world settings. The ability to understand such generative models is of primary importance in many fields, as it enables researchers to make predictions about the behaviour of the system under exam and to compare systems, gaining knowledge. In general, a random graph is a model network in which some specific set of parameters take fixed values, but the network is random in other respects. The theory of random graphs studies the intersection between graph theory and probability theory. It is usual to describe a random graph by a sequence of steps to construct it. This sequence is called a *random graph model*.
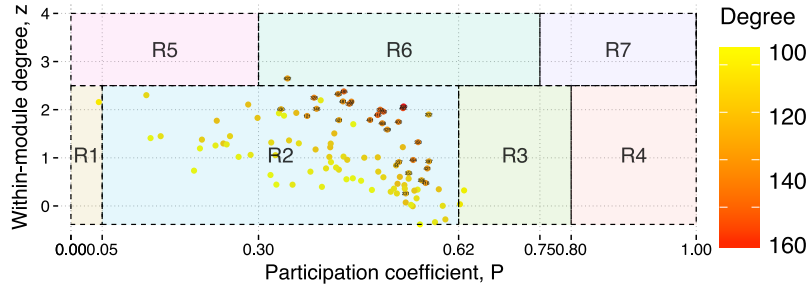
Figure 1.6: Guimera and Amaral classification scheme for the modular structure induced by Modularity maximization of a co-activation network as obtained from [34]. Each node is indicated by specific coordinates (P_i, z_i) and occupies one of the seven different regions. Nodal degrees are color-coded from yellow to red, as shown in the legend.
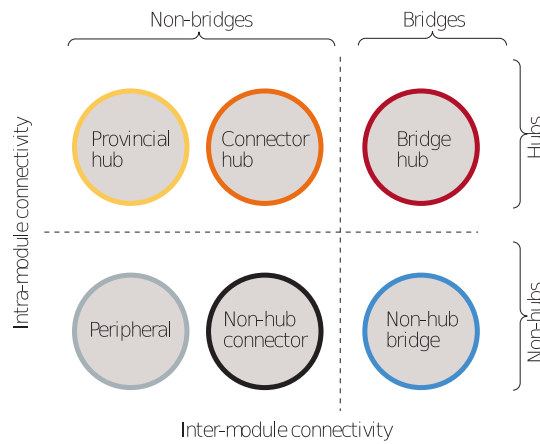
Figure 1.7: Schematic assessment of nodal roles. Interestingly under this convention, bridges represent the most extreme case of nodes, with a relatively equal distribution of connectivity between different modules. Hence, bridges are best detected when overlapping communities are allowed. They bridges hubs and non-hubs occupy the R7 region of the original GA scheme. Adapted from [26].

In the Erdős-Rényi model (also called $G_{n,m}$ model, the sample space Ω is the set of all graphs having n nodes and m edges. One of the simplest models is the random network generated by selecting exactly m edges from the set of all possible edges and placing them at random over the nodes pairs. This model, with the fixed constraint of edge number, is mathematically known as $G(n,m)$ or *Gilbert* random graph. In probabilistic terms, the $G(n,m)$ graph is equivalent to say that the network is sampled choosing uniformly from the set Ω of all $\binom{\binom{n}{2}}{m}$ possible graphs. Intuitively then it is possible to assign a probability score to each instance of a random graph model. In the $G(n,m)$, each instance is equiprobable.

Strictly speaking, a random graph model is better defined as an ensemble of networks, therefore as a probability distribution $\mathcal{P}(\Omega)$ over the space of all graphs G. The Gilbert random graph model is associated with a probability distribution sharply peaked in $\mathcal{P}(G) = 1/\Omega$ for graphs with n nodes and exactly m edges and zero otherwise. Although some properties of the $G(n,m)$ models are easy to compute, like the mean number of edges (trivial), others are more complicated because of the hard constraint imposed on edge number.

A way more useful random graph model is instead due to Erdős and Rényi [90]. To solve the analytical computation problems on the $G(n, m)$ model, the Erdős-Rényi model introduces a parameter p_{er} that specifies the probability for an edge to exist given two nodes, rather than picking exactly m edges. Technically the $G(n, p)$ random graph model or *Erdős and Rényi* (ER) model, is the ensemble of simple graphs with n nodes in which each simple graph G of m edges appears with probability

$$P(G) = p_{er}^m (1 - p_{er})^{\binom{n}{2} - m} \tag{1.12}$$

Differently from the $G(n, m)$ model, because of the Bernoulli distribution of edges, this graph model is also called *Bernoulli random graph* or *Poissonian random graph* and generally intended as *the* random graph.

As every graph obtained from the Erdős-Rényi model is a sample from a probability space, a description in terms of ensemble averages is more appropriate. In the ER model, the expected number of edges in such a graph is the average over $\binom{n}{2}$ independent coin tosses with probability p_{er}, namely $\langle m \rangle = \binom{n}{2} p_{er}$ and the expected degree $\langle k \rangle = p_{er}(n - 1)$ is the same for every node. The probability that a node i, chosen at random, has degree k is called the *degree distribution*. In general terms, such probability is expressed as:

$$\Pr(k) = \frac{\langle |\{i \| k_i = k\}| \rangle}{n}. \tag{1.13}$$

In the ER model, a given node is connected with probability p_{er} to each of the other $n - 1$ nodes. Hence, the probability of being connected to a particular k other nodes and not to any of the others is $p^k(1 - p)^{n-1-k}$, but since there are $\binom{n-1}{k}$ ways to choose those k other nodes, hence the total probability of being connected to exactly k others is:

$$\Pr(k) = \binom{n-1}{k} p_{er}^k (1 - p_{er})^{n-1-k} \tag{1.14}$$

a binomial distribution that can be approximated to a Poisson distribution for large n:

$$\Pr(k) \approx \frac{(n\, p_{er})^k e^{-n\, p_{er}}}{k!}. \tag{1.15}$$

In generalizing the concepts of random networks models, one can define every simple graph on n nodes as the output of a random process where each edge (i, j) is sampled with a probability $\Pr(i, j)_\theta$ with θ a vector of hyperparameters. In the simplest case of the ER graph, this generative model is a uniform probability distribution with $\theta = p_{er}$ as every pair of nodes is linked by an edge with constant probability.

Importantly, both the ER and Gilbert models should not exhibit any modular structure as every substructure that appears is an effect of random fluctuations due to finite sampling effects. Nonetheless, some algorithms may find communities in ER graphs, due to overfitting [91]. Additionally, the ER model does not realistically represent empirical networks, even if it is very useful to simplify analytical calculation of network properties. Its Poissonian degree distribution (Figure 1.8) is far from the degree distributions seen in most of the real-world networks. In fact,
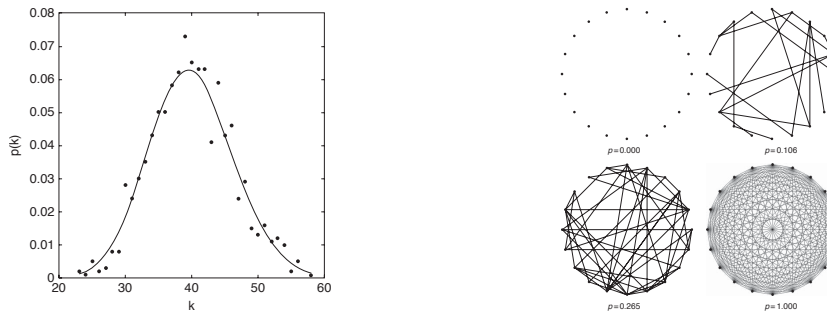
Figure 1.8: Examples of ER graphs. On the left the Poisson degree distribution of an ER network, $n = 1000, p = 0.04$. On the right ER model instances at different values of p. Adapted from [74].

the empirical behaviour of the degree distribution of real networks is one of the main reasons why they attracted so much interest from the scientific community.

The world-wide-web, as well as the connectome of the worm *C. Elegans*[2] and many other examples, shows that the degree distribution of complex networks is not centered around a mean value, as predicted by ER model, instead highly connected nodes exist within a bulk of low connected nodes and the degree distribution is not decaying exponentially, thus producing the observed phenomenon of *heavy tails*. Many networks display a degree distribution that follows a *power-law* form:

$$\Pr(k) \propto k^{-\gamma} \tag{1.16}$$

where γ is usually included in the $[2, 3]$ interval. Power-law distributions, also dubbed *scale-free*, lack a specific scale. Indeed, $P(\alpha k) = f(\alpha)P(k)$ for some multiplicative constant α and f a function of α. Therefore, the functional form of the power-law remains unchanged.

Hence, in understanding the limitation of the Erdős-Rényi model, researchers developed a multitude of generative models of graphs that tried to better model properties like the heavy-tails saw in power-laws or also the *small-worldness* phenomenon, i.e. the observation that most nodes in a network are reachable from every other in a small number of hops. To keep into account the features of empirical networks, other models of networks have been developed like the models of Watts-Strogatz [37] model and Barabasi-Albert [92].

For example, in the Watts-Strogatz model a graph with groups of densely connected nodes is generated as follows:

1. Construct a ring with n nodes and connect each node to the l nearest nodes ($l/2$ nodes on each side of the ring).

2. Choose a node i and the edge e that connects i to its nearest neighbour in a clockwise manner.

3. With probability p_r replace the edge e by the edge that connects i to a node taken at random according to a uniform distribution over the entire ring.

4. Repeat the steps 2 and 3 for each node in a clockwise manner.

---

2 *Caenorbiditis elegans*, a small nematode worm, famous as the first living organism whose connectome has been completely mapped.

5. Choose a node i and the edge e that connects i to its second neighbour in a clockwise sense and repeat the steps 2-4.

6. Repeat the process considering the third nearest neighbour and so on until each edge of the original lattice has been considered.

The resulting graph shows the property of having short average path length, although the degree distribution is not power-law. To prevent this issue, the Barabasi-Albert graph model aims to construct graphs with a dynamic attachment process. In the BA model, one considers a small number n_0 of initial nodes. Then at each step, one adds a new node and connect it to a fixed number ($\leqslant n_0$) of nodes that already exist in the network. The probability that the new node will be connected to a node i is then proportional to $k_i^{p_s}$, where p_s is a parameter called the *scaling exponent*. That algorithm generates graphs in which the frequency of nodes with degree k is asymptotically proportional to k^{-3}. This power relationship between frequencies and degrees indicates that this model accounts for the power-law distribution of degrees seen in many real networks. Although very well studied, both models are still lacking a strong community structure and more complex models have been studied to explicitly embed modules.

1.3.5 Stochastic block models

None of the previously reported generative models explicitly accounts for the modular structure. The modules often observed in the ER model are due to statistical fluctuations. The BA model, too, does not take into account a block structure, when generating a power-law network and clusters in the BA model are due to local structures around nodes with high degree. To overcome these limitations, a variable encoding the block structure must be taken into account explicitly.

The most commonly used generative model for random modular networks is called the stochastic block model (SBM) [93], which generalizes the Erdős-Rényi random graph by giving each pair of nodes a connection probability depending only on their module affiliation σ. The SBM model tends to produce graphs containing communities, subsets characterized by being connected with one another with specific edge densities.

A stochastic block model for undirected graphs is defined by three variables: the number of nodes n, the set of communities (represented by a membership vector σ) and a symmetric square matrix $\mathbf{B} \in \mathbb{R}^{|C| \times |C|}$ of edge probabilities between modules. In the case the matrix $\mathbf{B}$ is constant, the SBM recovers exactly the ER model. If the matrix $\mathbf{B}$ takes only two different values, a model known as the *planted partition model*, each pair of nodes in the same block are connected with probability $\mathbf{B}_{ij} = p_{in}$ and pairs of nodes in different modules are linked with probability $\mathbf{B}_{ij} = p_{out}$, as shown in Figure 1.9. Specific edge probability in the planted partition model is simply described as $Pr_{ij} = p_{in}\delta(\sigma_i, \sigma_j) + p_{out}(1 - \delta(\sigma_i, \sigma_j))$. In building the planted partition model, though one has to set in advance the number of modules $|C|$ (or blocks) and to define an equal number of nodes in every module.

Although largely used, the planted-partition model has several drawbacks. Indeed, as a generalization of the ER model, each node has the same expected degree
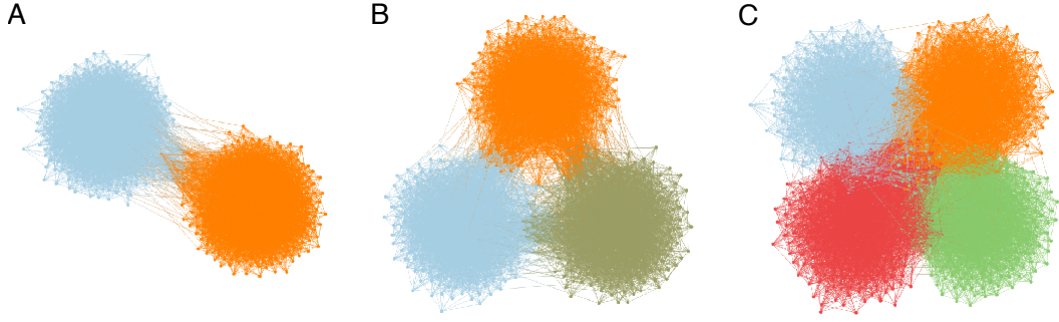
Figure 1.9: Realizations of the planted partition model with different intra- and extra-cluster linking probabilities and number of modules. A., $|C| = 2, p_{in} = 0.99, p_{out} = 0.025$, B. $|C| = 3, p_{in} = 0.98, p_{out} = 0.06$, C. $|C| = 4 p_{in} = 0.97, p_{out} = 0.12$. Adapted from [94].

and all communities have the same size. While these properties are of great help when performing analytical calculations, they also make it badly suited for practical applications, as degree homogeneity is seldom realized in real networks.

To overcome the unrealistic settings of the planted partition model, the Lancichinetti-Fortunato-Radicchi (LFR) model, first described in [80], modified the planted partition model to exhibit power-law distribution of both node degrees and community size, with tunable exponents τ_d and τ_c respectively, to respect what is observed in real networks. As each realization is accompanied by the membership affiliation vector σ, the LFR model has been used primarily as a benchmark for evaluating the effectiveness of different community detection methods [54, 69] at different levels of difficulty. Furthermore, the LFR model was extended to weighted networks with possibly overlapping modules in [95]. The major advantage of the LFR model is the tunable level of community mixing. Indeed, each node share a fraction $1 - \mu_t$ of edges with nodes in its same community and a fraction μ_t with nodes in other communities: $0 \leqslant \mu_t \leqslant 1$ is dubbed the *topological mixing parameter*. Additionally on weighted networks, a parameter β is used to assign strength to each node: $s_i = k_i^\beta$. This power-law form of node strengths is chosen to match observations on real graphs, as indicated by [96]. A weights-mixing coefficient μ_w is used to divide the node strength into two components, intra- and inter community, such that:

$$s_i = (1 - \mu_w)s_i^{(intra)} + (\mu_w)s_i^{(inter)}.$$

The actual weights of the edges are then chosen to minimize the variance between the expected and the observed strengths by means of an iterative optimization step. Hence is possible to obtain a theoretical estimate of the average edge weights that reads:

$$\langle w^{int} \rangle = \frac{(1 - \mu_w)\langle s \rangle}{(1 - \mu_t)}\langle k \rangle = \frac{(1 - \mu_w)}{(1 - \mu_t)}\langle k \rangle^{\beta - 1}, \tag{1.17}$$

$$\langle w^{ext} \rangle = \frac{(\mu_w)\langle s \rangle}{(\mu_t)}\langle k \rangle = \frac{(\mu_w)}{(\mu_t)}\langle k \rangle^{\beta - 1}. \tag{1.18}$$

For the planted partition model, communities were defined as long as $p_{in} > p_{out}$. For the weighted LFR model instead the condition to have well defined communities is:

$$\langle w^{int} \rangle \geqslant \langle w^{ext} \rangle \tag{1.19}$$

resulting in the equivalent formulation $\mu_w \geqslant \mu_t$. If condition 1.19 is not met, then each node has on average a strength relatively higher than its degree, an unrealistic circumstance in real-networks, as highest degree nodes are typically nodes with high strength [96].

The modularity of the LFR network is controlled primarily by the topological mixing coefficient. In the case μ_t is exactly zero, the network consists of isolated components and the stochastic block matrix is completely assortative or block-diagonal. Conversely, for $\mu_t = 1$, the block matrix is null on the diagonal and completely disassortative, exhibiting no links inside communities and only links between them. The LFR model has a number of other parameters that control all the network parameters with great precision [80, 95] as described in Figure 1.10. The LFR model is of great importance as it served as a benchmark test for community detection algorithms and as a model of heterogeneity of modules size when studying the performance of community detection methods. The original c++ code to generate LFR weighted networks is available at the Santo Fortunato's homepage `https://sites.google.com/site/santofortunato/package3.tar.gz`. Bindings for `Matlab`,`Octave` and `Python` the Fortunato's code are freely available at `https://github.com/carlonicolini/lfrwmx`.

$\tau_d \geqslant 1$	Power-law exponent of nodes degrees.
$\tau_c \geqslant 1$	Power-law exponent of community size.
$\beta \geqslant 1$	Exponential parameter for strengths distribution.
$C \in [0,1]$	Desidered average clustering coefficient.
min_c	Minimum number of nodes in each community.
max_c	Maximum number of nodes in each community.
$\mu_t \in [0,1]$	Topological mixing coefficient.
$\mu_w \in [0,1]$	Weights mixing coefficient.
$\langle k \rangle$	Desidered average degree.
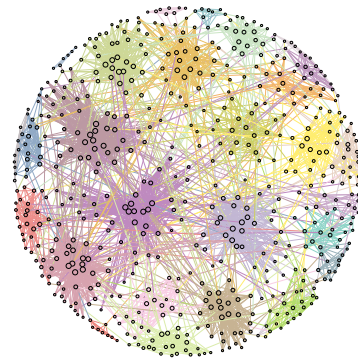
Figure 1.10: LFR parameters and a single instance with $n = 600$, $\langle k \rangle = 12$, $\tau_d = 2$, $\tau_c = 1$, $\mu_t = 0.1$, $\mu_w = 0.1$, $min_c = 5$, $max_c = 50$.

1.4 A SYNTHETIC MODEL OF A RESTING STATE NETWORK

Among the synthetic generative models for complex networks, no model tries to embed the typical circumstances that take place when measuring human brain resting state activity with fMRI. To this end, here I introduce a theoretically sound method for the generation of synthetic FC networks that mimic properties of resting state fMRI networks, including noise and intersubject variability while presenting a pre-determined ground-truth modular structure against which the performance of community detection algorithms can be tested [97].

The general idea is that, starting from an adjacency matrix with a given modular structure, we can generate time-courses for each of the nodes whose pairwise correlations reproduce the edge structure of the original matrix. Noise can be added to the time-courses, and the resulting correlation matrix will provide a noisy representation of the original one. This procedure can be repeated multiple times to produce different datasets that represent different subjects in the study.

In practical terms, given an undirected weighted graph $\mathbf{C} \in \mathbb{R}^{n \times n}$ whose community structure is known a-priori, its nearest positive definite matrix [98] is obtained with its Cholesky decomposition, i.e. an upper triangular matrix $\mathbf{L} \in \mathbb{R}^{n \times n}$ such that $\mathbf{LL}^{\mathsf{T}} = \mathbf{C}$. Starting from uncorrelated variables $\mathbf{X} \in \mathbb{R}^{n \times l}$, one can then generate correlated random variables $\mathbf{Y} = \mathbf{LX}$ such that $\mathbb{E}(\mathbf{YY}^{\mathsf{T}}) = \mathbf{C}$ [3]. Different levels of noise can then be injected into $\mathbf{Y}$ prior to the computation of the correlation matrix. Schematic of this procedure is shown in Figure 1.11. As described in [97] I tested this idea on two different models of planted partition: a variant of the ring of cliques [72] and the Lancichinetti-Fortunato-Radicchi (LFR) network [80], whose degree distribution and modular structure can be tuned to replicate topological features of real-world networks, including scale freeness [99] and the presence of densely interconnected cores [44].
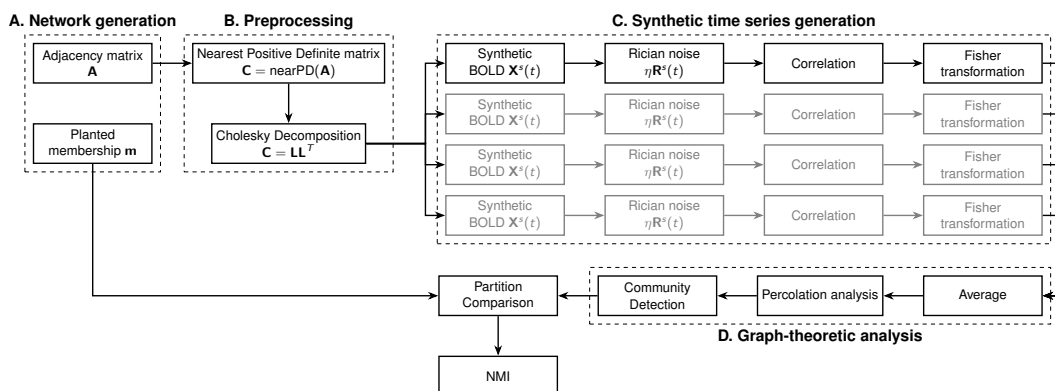
Figure 1.11: Flowchart of the generation and analysis of the synthetic datasets. In A the network with a pre-defined community structure is generated. The adjacency matrix is then processed in block B to obtain the nearest positive definite matrix for the Cholesky decomposition. This enables the generation of node-wise time-courses into which different levels of noise can be injected. The procedure is repeated multiple times to generate different instances (mimicking different subjects in the sample). Finally, correlation matrices are calculated for each instance (block C), and Fisher transformed to calculate the average adjacency matrix for analysis by community detection algorithms (block D). Lastly, resulting partitions are compared with the original, planted one in terms of NMI.

---

3 Since the covariance matrix of a zero-mean, unit-variance random variable X is given by $\mathbb{E}(XX^{\mathsf{T}})$, it is easy to verify that $\mathbb{E}(YY^{\mathsf{T}}) = \mathbb{E}(LX)(LX)^{\mathsf{T}} = \mathbb{E}(LX)X^{\mathsf{T}}L^{\mathsf{T}} = L\mathbb{E}(XX^{\mathsf{T}})L^{\mathsf{T}} = L^{\mathsf{T}}IL = C$

1.5 COMPARING COMMUNITY STRUCTURE IN NETWORKS

Over the years a large number of community detection methods appeared and a quantitative way to assess their performance on benchmark networks was needed. In most generative models of networks, the community structure is determined a-priori, like for example in the stochastic block model, the planted partition model and the LFR model. Conversely, community structures of real-world networks are often annotated by experts in the field: it is the case of the karate club network of Zachary [100] or the political blog network [101].

A large number of functions for comparing similarities and differences between partitions of a network have been proposed in the past [102, 103]. They are used to quantitatively provide a number that tells to which extent two partitions are similar. Typically the result is normalized in the $[0, 1]$ range, being close to 1 for very similar clusterings and close to 0 when mostly dissimilar.

The most well-grounded and performing metrics for the comparison of two community structures are rooted in information theory [102, 103, 104]. *Normalized Mutual Information* [102] and *Variation of Information* [105] are the most widely used, despite it was recently found that both of them suffer from systematic errors due to the finite size of the network [106]. In the rest of the discussions, we ignore the limitation of finite size effects, by only working with relatively large networks.

Normalized Mutual Information (NMI) assumes that the clustering comparison is a problem of message decoding. Implicit in this interpretation is the idea that if two partitions are similar, inferring one partition from the other needs very little information. Let us denote two generic partitions of the same graph by their clusterings ζ^x, ζ^y with c_x and c_y communities respectively: $\zeta^x = \{\zeta^x_1, \ldots, \zeta^x_{c_x}\}$ and $\zeta^y = \{\zeta^y_1, \ldots, \zeta^y_{c_y}\}$. The number of nodes in the i-th community of ζ^x is $n_i := |\zeta^x_i|$ and the number of nodes in the j-th community of ζ^y is $n_j := |\zeta^y_j|$. The two communities share a number of nodes $n_{ij} = |\zeta^x_i \cap \zeta^y_j|$. Let us also consider the community assignments σ^x_i and σ^y_i for partitions ζ^x and ζ^y respectively; we treat the labels as values of two random variables X and Y with joint distribution $P(x, y) = P(X = x, Y = y) = n_{xy}/n$, which implies that $P(x) = P(X = x) = n^X_x/n$ and $P(y) = P(Y = y) = n^Y_y/n$.

The *mutual information* between the two clusterings ζ^x, ζ^y is then defined as $I(X, Y) = H(X) - H(X|Y)$ where $H(X) = -\sum_x p(x) \log p(x)$ is the Shannon entropy of X [104]. The mutual information itself is not very useful, because hierarchically splitting the clusters in ζ^x would produce no change in the prior $H(X|Y)$ and partitions with different hierarchies of the same clusters, would go unnoticed. This observation led Danon [102] to introduce normalized mutual information for clustering comparison as:

$$\text{NMI}(X, Y) = \frac{2I(X, Y)}{H(X) + H(Y)} = \frac{2 \sum_{i=1}^{c_x} \sum_{j=1}^{c_y} \frac{n_{ij}}{n} \log\left(\frac{n_{ij} n}{n_i n_j}\right)}{\left(-\sum_{k=1}^{c_x} \frac{n_k}{n} \log\left(\frac{n_k}{n}\right)\right) + \left(-\sum_{k=1}^{c_y} \frac{n_k}{n} \log\left(\frac{n_k}{n}\right)\right)}.$$

(1.20)

Another measure based on entropy is Variation of Information (VI), proposed by Meila [105]:

$$
\begin{aligned}
VI(X,Y) &= H(X) + H(Y) - 2I(X;Y) \\
&= (H(X) - I(X;Y)) + (H(Y) - I(X;Y))
\end{aligned} \tag{1.21}
$$

Informally, VI weighs the contribution of two terms. The first term in brackets measures the amount of information about X that one looses, while the second measures the amount of information about Y that one gains when going from clustering X to clustering Y. The situation is captured in Figure 1.12. VI has the desirable property that, being a proper distance, it defines a metric on the space of the partitions. Variation of information is also a local measure, i.e. the similarity of partitions differing only in a small portion of a graph depends on the differences of the clusters in that region, and not on the partition of the rest of the graph. As noted by Karrer [107] VI is upper-bounded by a $\log(n)$ factor, so a simple normalization brings it in the $[0,1]$ range. Importantly, VI is zero for maximally equal partitions and 1 for mostly dissimilar, inversely to NMI.
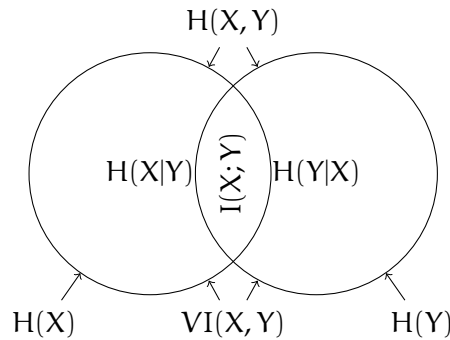
Figure 1.12: Venn diagram representation of the important quantities used to treat the clustering comparison problem in terms of information theory. The two circles are the entropies of variables X and Y. The mutual information $H(X;Y)$ is the intersection of the information in X with the information in Y. The Variation of information is equal to $VI(X;Y) = H(X|Y) + H(Y|X) = H(X) + H(Y) - 2I(X;Y)$ because $I(X;Y)$ is counted twice.

1.5.1 Confusion matrix

It is also possible to quantify the confusion matrix **C** between the detected and planted modules. Each element C_{ij} is the number of nodes in the planted community i that appear in the detected community j. For each planted community one scores as true positives (TP) the nodes correctly identified as belonging to the ground-truth community, and as false positives (FP) the nodes wrongly assigned to a community; similarly false negatives (FN) are nodes wrongly classified in different communities and true negatives (TN) the nodes correctly classified as out of the community. Sensitivity, defined as $TP/(TP+FN)$, decreases with increasing number of False Negatives. Specificity instead is defined as $TN/(TN+FP)$ and decreases when many nodes are wrongly assigned in the same community.

Accuracy (Acc) and Matthew Correlation Coefficient (MCC) are defined on the basis of the confusion matrix as:

$$\text{Acc} = \frac{(TP + TN)}{TP + FP + TN + FN} \quad \text{MCC} = \frac{(TP \times TN - FP \times FN)}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

where TP, TN, FP, FN are the number of true positives, true negatives, false positives and false negatives identified in the detected partition with respect to the planted partition. Accuracy takes into account the proportion of correctly classified samples and can present relatively high values even in the case of poorly performing detection methods when the classes have very different size. The Matthew Correlation Coefficient takes into account true and false positives and negatives. It is a balanced coefficient, to use especially when classes are very imbalanced.

1.6 THRESHOLDING AND PERCOLATION ANALYSIS

In weighted networks, sparsification procedures are often applied to remove the weakest edges, which are the most affected by experimental noise, and to reduce the density of the graph, thus making it theoretically and computationally more tractable. Indeed, weak links may contain significant structural information [108], and procedures to identify the optimal trade-off are the subject of active investigations [109]. After sparsification, the network can be binarized, by setting all edge weights to unity. Whether this procedure may help to obtain a better view of the underlying structure of networks, is a hot debate in the literature [110, 108].

Ideally, the obtained network should retain structural relations and ignore spurious links, but how is it possible to determine a proper value of threshold? No well-accepted method to choose the best threshold has been adopted pragmatically in the brain networks literature and different scholars motivated their choice by different needs [111, 112, 113].

Percolation analysis, a method originally grounded in statistical physics, was suggested by Gallos [111] to identify the optimal sparsification threshold for community detection in brain connectivity networks. Percolation is a model to describe phase transitions of connected subgraphs in random networks [114, 115]. In the ER random graph, the size of the largest connected component shows a sharp transition at some threshold value [114]. Contrarily, brain networks exhibit multiple percolation thresholds, revealing a hierarchy of clusters as shown in Figure 1.13. The jumps corresponding to percolation transitions are indicative of non-trivial correlations that constitute well-defined modules of brain activity. Hence, the multiple percolation thresholds are representative of a set of weak ties that guarantee appropriate communication (in terms of information transfer) between functionally specialized modules. Following these observations, the definition of the optimal threshold is the one just above the fragmentation of the largest connected component. This choice maximizes the information extracted by subsequent applications of community detection algorithms, and has been applied and validated in human [111] and animal [116] studies.

The most common choice when comparing networks from different groups is to adopt a fixed threshold that keeps the same number of edges in the two groups.
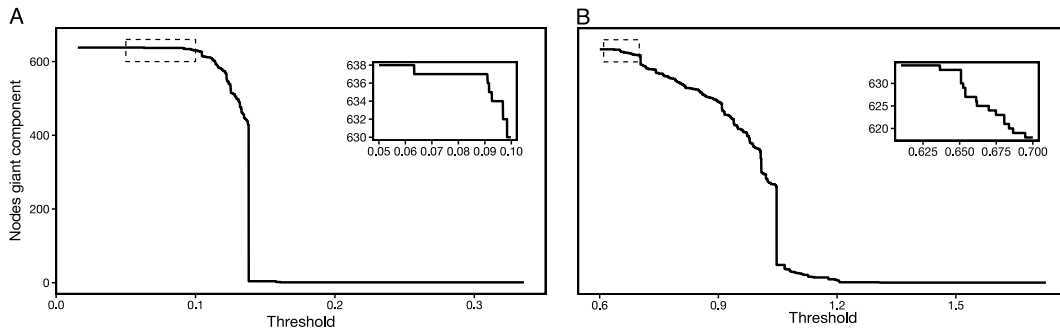
Figure 1.13: Percolation analysis for two brain networks available from the BrainConnec-
tivityToolbox [110], available at `https://sites.google.com/site/bctnet/`.
(A) is a coactivation network, (B) a resting state network, both with 638 nodes
and 18625 edges. The number of nodes in the giant component has a step-
wise behaviour with respect to the threshold.

This does not take into account the particularities of the analyzed group and
datasets. It is indeed important to stress that the distribution of the weights as
well as the density or connectivity of the population are not exploited when us-
ing fixed thresholds and the final result does not mirror this information. To this
end, it has been demonstrated, especially in patients groups such as in schizophre-
nia, that the brain presents weaker functional connectivity than in healthy control
persons [65].

For this reason, percolation analysis should enable data-driven determination
of the optimal sparsification threshold that preserves network structure and con-
nectedness while removing potentially spurious correlations.

2 | MODULAR ORGANIZATION OF BRAIN CONNECTIVITY: THE RESOLUTION LIMIT

2.1 MODULARITY OF BRAIN NETWORKS

A systematic analysis of the structure of neural systems in different organisms points out that modularity is a factor that shapes neural networks from the small nematode worm *Caenorhabditis Elegans* to the most complex human brain [59, 117, 118]. Topological network analyses suggest indeed that modular and hierarchical structural networks are particularly well-suited for the functional integration of locally specialized neural operations that underlie cognition [119, 39, 120, 59, 121]. In this respect, brain function or cognition can be described as global integration of local integrators [118].

Among the many advantages of a modular organization of the brain, three are of greatest importance and strictly intertwined: adaptability, robustness and wiring cost economy. Firstly, from an evolutionary perspective, a modular architecture is highly adaptable to an ever changing environment and variable loads of cognitive demands. From generation to generation, modules that were already well-fit to the environmental conditions were selected and passed, while others were only slightly rewired through the natural selection mechanisms. This pressure, together with the imperative of a reduced wiring cost have been crucial to select an architecture that converged into tightly connected sub-networks loosely linked to each others [122]. Computational studies have shown that a modular organization emerges spontaneously in complex brain networks, promoting flexibility to an ever-changing outside environment [123, 124] and maximizing bidirectional information exchange between neural populations [125].

A second benefit of modularity is robustness as it facilitates evolvability and robust traits like the multiscale modular architecture are often selected by evolution [126, 127]. Indeed, keeping a system compartmentalized can limit the influence of external sources of perturbation, thus bounding the interdependence of biological processes going on in each unit [128], and promoting greater resilience in the context of continuous genetic and developmental changes. In addition, swapping and rewiring maladaptive modules is less demanding that restructuring the whole system. A general problem in the acquisition of new skills is the phenomenon of catastrophic forgetting, by which newly acquired skills come at the expense of others [129]. In the case of neural networks it has been shown that a modular organization helps to avoid catastrophic forgetting of newly learned skills, whereby new skills are learned by slightly rewiring existing modules, rather than changing the overall organization [130]. Local disruptive modification may only affect one module, whereas the others are preserved from the insult [25]. In the case of externally or internally generated fluctuations, a modular architecture

prevents functional disruption of the network leading to the possible death of the organism.

Thirdly, the spatial and metabolic constraints of neuronal wiring are factors that deeply shaped the brain architecture [120, 131]. Computational and empirical studies converged on the result that a multiscale organization of modules inside modules is the one that satisfies the constraints imposed by minimization of energetic cost and spatial embedding [132, 120, 133, 134, 135]. Here, wiring cost must not only be intended as the physical volume of axons and synapses but also be considered in terms of energetic demand for signal transmission, additional processing cost for noise correction over long distance signalling and sustenance of the necessary neuroglia that supports neuronal activities [120]. Regarding wiring cost reduction, modular networks emerge naturally when minimizing the average path length and the total number of links, while maximizing robustness against perturbations in node activity. By way of example, in the field of robotics, spontaneous structural modularity of neural networks controllers emerge from evolutionary optimization when successful robot behaviour is encoded as fitness function [136].

Application of resting state fMRI has allowed the identification of various modules in human functional connectivity. These spatially distinct modes that demonstrate synchronous BOLD fluctuations consist in strongly correlated, anatomically separated, possibly remote regions of the brain [9, 15, 83, 137]. Various methods exist for analyzing resting-state data, including seed-based approaches [9, 15], independent component analysis [138, 139] and graph-based methods [140, 27]. They have generally shown that the human brain is functionally organized in a number of interconnected regions with boundaries that reflect specific aspects of cognitive and behavioural aspects [141, 142, 34]. The first analysis method ever used was the seed-based approach [9], which has been applied in numerous studies [15]. This technique is based on the correlation of voxels within selected regions of interest (ROIs), with the average BOLD signal from all other voxels in the brain. An accurate thresholding is then needed to identify voxels that are significantly correlated with the region of interest.

Another widely used approach is the decomposition of the endogenous BOLD fluctuations by means of Independent Components Analysis, a technique that aims at maximizing statistical independence between time courses of single voxels and that can identify separate spatial modes [139, 138, 143]. Yet, ICA with its variants still relies on user experience to discern important components from noise [144] and despite some notable attempts to identify noise [145, 146], ICA, like seed-based approaches, does not yield any information on the hierarchy of modes, nor on their inter-relations. Moreover, largely overlapping results can be obtained by ICA and seed-based approaches [147].

The evidence of modules in human functional networks is based on their robustness and consistency within a variety of MRI protocols, different analysis methods, and MRI scanners. Differently from structural networks, since functional networks do not measure physical relations with associated metabolic or material cost, they may feature long-distance connections [148, 149] and ever changing patterns. The majority of analysis techniques are concordant in identifying a set of robust functionally linked regions, or *modules* in resting state networks. These modules, roughly highlighted with coloured ellipses in Figure 2.1 span the

primary somatosensory network, the primary visual and extra-striate network, a network consisting of bilateral temporal/insular areas and anterior cingulate cortex regions, left and right lateralized networks consisting of superior parietal and superior frontal regions (often reported as one single network) and the so-called default-mode network spanning precuneus, medial frontal, inferior parietal cortical regions and medial temporal lobe. The Default Mode Network, or DMN, firstly
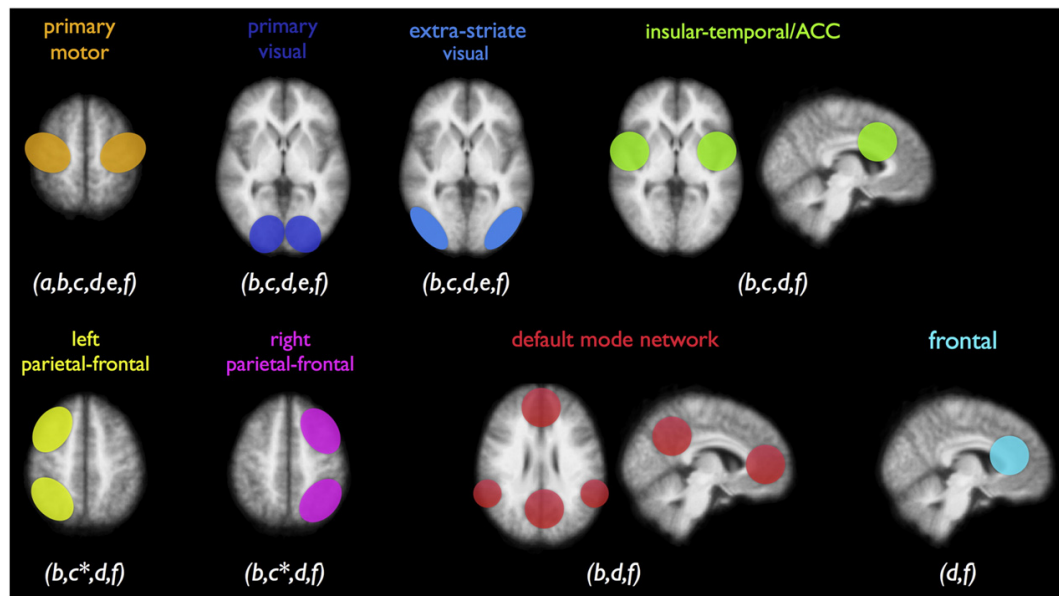
Figure 2.1: Functional modules of the resting state brain, averaged over different studies and protocols, and visualized roughly on a brain template to indicate their anatomical position. Figure taken from [30].

identified by Raichle [15] with positron emission tomography and then confirmed by Greicius et al using fMRI [150], is one of the most consistent networks among a variety of studies. It comprises a group of spatially remote brain regions that show higher levels of correlation when the brain is not involved in any particular cognitive or behavioural task. For this reason, it is hypothesized that there are two opposing concurrent large-scale modes, one including the DMN, indicated as "task-negative" or "intrinsic" system, and the other comprising regions involved in attentional and task-evoked activities, dubbed "task-positive" or "extrinsic" system [151, 10, 152]. The Default Mode Network or DMN (Figure 2.1) is thought to be connected to high-level cognitive functions, given its specific traits of structural links to associative brain regions and that the DMN may support internal mental processing like mind-wandering, abstract thinking and planning of events in the future [16, 153]. Additionally, evidence has pointed to disruptions in the DMN with people with Alzheimer's disease [153], autism spectrum disorders [154] and schizophrenia [155].

Apart from the DMN and networks for the primary sensory areas, other networks have different roles, many of which are partly speculative. Subcortical communities have also been identified with rs-fMRI: an appropriate ICA seeding makes possible to observe functional modules located at the thalamus and hippocampus [156]. Modules in functional brain networks have been explained as the possible basis of a large variety of neural processes. By way of example, cer-

tain functional processes, like colour vision, have been described as anatomically localized [157], while others, like working memory, have been proposed to involve more globally integrated processing systems [148, 158]. Interestingly modules of functional brain networks are non-static as recent studies found that they tend to become less and less modular with ageing [159, 160]. In all respects, the experimental indication of the modular organization is a powerful and robust marker against which to measure the health of the brain. Thus, modularity is not only a biologically plausible feature but it is very likely necessary to maintain the rich repertoire of cognitive tasks that our brain engages in every day of our life.

Modeling of resting state networks using graph-theory provides a compelling alternative to seed-based and ICA approaches. Within this framework, ROI are nodes of a network, connected by edges of weight proportional to the strength of correlation between ROIs. Graph-theory offers a wide amount of measures to characterize functional connectivity networks and in particular a wealth of methods to identify communities, subsets of nodes which are more strongly correlated between each other, than with external nodes [27, 25, 68].

Interestingly, the size distribution of the functional modules has never been analyzed in details, especially when graph-theoretical methods were applied. This is the case of one of the most fortunate and widely adopted methods to identify modules in functional networks, Newman's Modularity [71]. Despite its popularity and merits, Newman's approach presents some important limitations. Already at an early stage, Modularity-based methods were shown to suffer from a resolution limit, as they fail to identify modules that are smaller than a scale that depends on the size of the overall network [72]. As a consequence, even unambiguously defined modules, like complete sub-graphs or cliques, may be unduly merged into larger communities when they are too small compared to the size of the network. Subsequent work by various groups has shown that the resolution limit is quite pervasive [69, 161, 162, 163, 164], and affects, to a different extent, many other methods, including Reichardt and Bornholdt's [165], Arenas and Gomez' [166], Ronhovde and Nussinov's [167], Rosvall and Bergstrom's (*Infomap*) [82, 164] and others.

Fixes have been proposed to circumvent the resolution limit, including the introduction of a tunable parameter that enables analysis of the network at an adjustable resolution level [165, 168, 169]. However, this requires prior knowledge of the expected size of the communities for the tuning of the resolution parameter. Moreover, it has been shown that an adjustable resolution parameter may reduce the tendency to merge small clusters, but only at the cost of unduly splitting large clusters [163]. Adjustment of the resolution parameter is an attempt to balance these two biases, but multiresolution methods fail to recover community structures comprising heterogeneous distributions of cluster sizes [163]. However, real-world networks are characterized by the coexistence of clusters of very different sizes, and no single parameter can adapt to the variety of network topologies observed in nature. Hence, the resolution limit may represent a critical shortcoming for the study of brain networks and is likely to have affected many of the studies reported in the literature.

In the following sections, I'm going to illustrate the current graph-theoretical approaches for community detection in networks and to discuss the theory of com-

munity detection, highlighting the practical problems of Modularity maximization.

2.2 COMMUNITY DETECTION IN NETWORKS

The process of grouping nodes in a graph to establish their common behavioral properties is a popular technique that goes under the name of *community detection*. Following the initial work by Hilgetag [170] dubbed OSA (Optimal Set Analysis), several graph theoretical methods have been deployed to investigate the modularity of structural and functional brain networks [67, 59, 151, 140, 25]. Typically, these methods rely on the optimization of a fitness function that measures the quality of a network partition against that of an ensemble of randomized networks with similar statistical properties, called the "null model". Optimization of the fitness function of choice is often computationally demanding and scales steeply with increasing network size. Hence, heuristics are needed to calculate nearly optimal partitions of large networks, like those derived from neuroimaging data, within reasonable computation time [171, 82, 172]. As the definition of community is not unambiguously stated, it is therefore, important to have a quantitative way to evaluate the goodness of such clusterings.

A global criterion for the definition of communities in networks can be thought in terms of a fitness function defined over the clustering. A *quality function* is a function Q that given a clustering of a graph ζ returns a scalar number. Usually one identifies "good" clusterings with high scores of the quality function and "bad" clusterings with low scores. In this sense, it is possible to rank partitions from bad to good, although is important to stress that the definition of good or bad clusterings is an *ill-posed problem* as every quality function puts emphasis on some features and perhaps not on others. Here and in the following sections, it must be stressed that the concept of quality function and community detection methods are separate as the first is a way to assess the goodness of a partition while the second relies on the definition of a quality function to design efficient algorithms and heuristics to find such good partitions.

One of the most important properties of a quality function is additivity, i.e. its expression as a sum of terms each pertaining to a single community. For a generic function of a cluster, or subgraph, $f(\zeta_i)$, an additive quality function specifies the goodness of a clustering as the sum of f over the distinct communities as follows:

$$Q = \sum_{\zeta_c \in \zeta} f(\zeta_c). \tag{2.1}$$

The majority of quality functions are additive, even if this requirement is not strictly required. In the next section, we explore the properties of some of the most important additive quality functions that emerged from the literature in this decade.

2.2.1 Spin glass based quality functions

The simplest formulation of a quality function is one that puts emphasis on intracluster edges and penalizes intercluster edges. In this terms, local optima of the quality function should correspond to partitions where the communities emerge as dense areas in the network loosely connected among them. A framework grounded in statistical mechanics for the definition of suitable quality functions for community detection meeting these requirements has been introduced by Reichardt and Bornholdt (RB) [165].

In the RB model the problem of community detection is cast in terms of finding the *ground state* of a spin glass, a model describing the behavior of large sets of interacting microscopic magnets. Actually, the properties of spin glass models are subject of intensive research in the last decades as their applications range from condensed matter and nuclear physics to neural networks. Here we present only their salient application to community detection and refer the reader for the details of the model to more specialized books [173].

A spin glass model is based on the definition of an Hamiltonian, a multi-variable scalar function that describes the total energy of the physical system with the configuration of its internal components. In our case, the internal components of the system are the nodes. The configuration of the system is then expressed by the community affiliation vector σ, meaning that node i stays in the community σ_i. The Hamiltonian used by Reichardt and Bornholdt (RB) includes four different contributions. The first two contributions act at intracluster level, positively weighing intracluster edges and negatively weighing intracluster non-edges with coefficients a_{ij} and b_{ij} respectively. The third and fourth contributions work on intercluster edges and non-edges, weighing them with factors c_{ij} and d_{ij}. The general form of the RB model is then expressed by the following Hamiltonian:

$$\mathcal{H}^{RB}(\sigma) = - \sum_{(i,j)\in V^2} \left[a_{ij}A_{ij} - b_{ij}(1-A_{ij}) \right] \delta(\sigma_i,\sigma_j) +$$
$$\left[(c_{ij}A_{ij} - d_{ij}(1-A_{ij}) \right] (1-\delta(\sigma_i,\sigma_j)), \tag{2.2}$$

with the convention that lowest energy states correspond to best community assignments. Rearranging Eq.2.2 and collecting the terms independent of the partition into a constant H_0, one then gets a simpler expression for $\mathcal{H}^{RB}(\sigma)$:

$$\mathcal{H}^{RB}(\sigma) = -H_0 - \sum_{(i,j)\in V^2} \left[\alpha_{ij}A_{ij} - \beta_{ij} \right] \delta(\sigma_i,\sigma_j), \tag{2.3}$$

where the two parameters $\alpha_{ij} = a_{ij} + b_{ij} + c_{ij} + d_{ij}$ and $\beta_{ij} = b_{ij} + d_{ij}$ depend on the *null model*, i.e. the probability that an edge exists between i and j after random edge rewiring. Hence, a null model provides a mean to compare a specific set of features of a graph, with its randomized version that should specifically lack those features.

Setting $\alpha_{ij} = 1$, $H_0 = 0$ and moving the summation indexes over the communities rather than over the pairs of nodes, one can express the resulting Hamiltonian, as an additive quality function expressed here as $\mathcal{H}^{RB}_{reduced}(\sigma)$, that reads:

$$\mathcal{H}^{RB}_{reduced}(\sigma) = - \sum_{(i,j)\in V^2} \left[A_{ij} - \beta_{ij} \right] \delta(\sigma_i,\sigma_j) = - \sum_c^C \left[m_c - \langle m_c \rangle \right]. \tag{2.4}$$

where $m_c = \sum_{ij} A_{ij} \delta(\sigma_i, c) \delta(c, \sigma_j)$ represents the number of links inside community labeled by c and $\langle m_c \rangle = \sum_{ij} \beta_{ij} \delta(\sigma_i, c) \delta(c, \sigma_j)$ is the expected number of links in community c as prescribed by the null model β_{ij}[1]. Among the additive quality functions that show up in the form of 2.4, the most important and popular is *Newman-Girvan's Modularity*.

2.2.2 Newman–Girvan Modularity

Newman-Girvan Modularity (or simply Modularity) [71], denoted here and for the rest of the work by Q^N, is based on the idea that a network obtained by randomly reshuffling the original graph edges while keeping the same degrees sequence, should not display any community structure. An important consequence of such randomization is that any stub in this null model, dubbed "configuration model" is equally likely to be connected to any other [73]. Thus, in the absence of correlations, the probability that two nodes are connected is expressed by:

$$P_{ij} = \frac{k_i k_j}{2m}, \tag{2.5}$$

where k_i and k_j are the degrees of node i and j. The "configuration model" is of great importance in network science as it assigns a higher probability of linking to nodes with high degrees, a feature that is compatible with most real world networks [73]. The motivations of the configuration model are addressed in details in section 2.2.3.

In terms of a spin glass model, Modularity measures the deviation from the observed intracluster density with respect to the expected intracluster density specified by the configuration model. Modularity is described in the form of 2.4 but normalized by the number of edges in the graph, taking the form described in Eq. 2.6.

$$Q^N = \frac{1}{2m} \sum_{(i,j) \in V^2} \left[A_{ij} - \frac{k_i k_j}{2m} \right] \delta(\sigma_i, \sigma_j), \tag{2.6}$$

whereby optimal partitions have high values of Q^N. As shown in Eq. 2.4, Modularity can be expressed as sum over communities of the difference of two terms:

$$Q^N = \sum_c^{|C|} \left[\frac{m_c}{m} - \left(\frac{K_{int}(\mathcal{G}_c)}{2m} \right)^2 \right]. \tag{2.7}$$

Modularity takes values in the range $[-0.5, 1]$, then a good partition should have Q^N values close to unity, identifying groups with much more internal connections than expected at random. In contrast, a bad partition with Q^N close to zero should identify groups with no more internal connections than we expect at random. In the next sections, we will challenge this idea and show that this observation has led to false statements, as a general phenomenon, dubbed *resolution limit*, heavily affects any quality function based on comparison with global null models. Specifically, in the case of Newman's Modularity, the resolution limit hampers its ability to detect modules smaller than a scale determined by the size of the graph.

---

1 Here and for the rest of the work, we set $P_{ij} := \beta_{ij}$ for agreement with more conventional notation of null models

2.2.3 The configuration model

A central property of the configuration model is the probability P_{ij} of the occurrence of an edge between two specified vertices i and j. In absence of correlations, as prescribed by the random reshuffling imposed by the configuration model, the probability that a stub emerging from vertex i is connected by an edge to any of the stubs of vertex j is $k_j/(2m-1)$ as there are $2m-1$ stubs in total. Therefore the total probability that vertex i and vertex j are connected by an edge is the product of the stub probability times the number of stubs from vertex i, namely:

$$P_{ij} = \frac{k_i k_j}{2m-1} \approx \frac{k_i k_j}{2m}. \tag{2.8}$$

Although the configuration model is most frequently applied to *simple graphs* (Figure 1.3a), the configuration model implies that the randomization is carried over the space of *loopy multigraphs* (Figure 1.3c). The rewiring probability considers indeed that the stubs of nodes can be reconnected both to the same source vertex (self-loops) or added to already existing edges (multi-edges). Consider for example the set of matchings on six stubs that form a triangle graph as illustrated in Figure 2.2. The configuration model chooses each distinct edge stubs labeling with equal probability. However, not only the first eight distinct reshuffled labelings are possible under configuration model (Fig.2.2a) but many other distinct matchings producing non-simple networks as shown in Figure 2.2b.

In practice, though, the fraction of edges involved in either self-loops or multi-edges is vanishingly small in the large n limit, and thus we may generally ignore them without much impact. Hence, the estimate of the probability of rewiring in Eq. 2.8 is only valid in large, sufficiently sparse graphs with a sufficiently bounded degree sequence. Under these hypotheses the expected number of edges between two vertices in the space of simple graphs is asymptotically the same as the expectation in the space of stub-labeled loopy multigraphs, i.e. the one previously introduced for the definition of Modularity: $\mathbb{E}_s[a_{ij}|k] \approx k_i k_j/(2m)$, where the subscript s denotes the space of simple graphs.

(a)   (b)
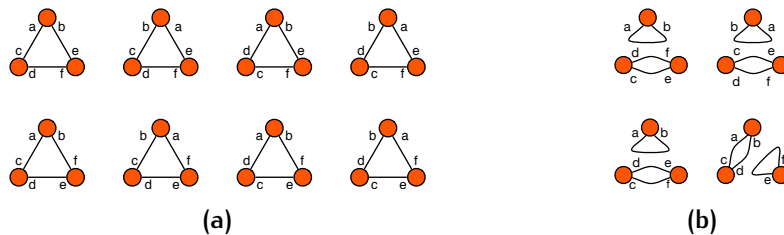
Figure 2.2: Twelve of the ninety different possible rewirings of a triangle graph in the configuration model. In (a) only rewirings leading to simple graphs are considered. In (b) just four rewirings leading to loopy-multigraphs are shown.

An approach that allows to get the right configuration model depending on the class of graph under exam, relies on the computational simulation of the correct rewiring probability by means of Markov Chain Monte Carlo algorithms, as proposed in [174]. In the remaining paragraphs, although flawed, we will use the classic configuration model from Eq. 2.8 to be adherent to most of the brain networks literature, justified also by the vanishing effects of self-loops in large sparse

networks. Nonetheless, we will illustrate many of the problems that the noncritical use of Modularity with this null model has introduced.

2.2.4 Other null models for Modularity

Modularity identifies communities as the subset of nodes whose internal fraction of edges deviates from the null configuration model on the same subset with the term $m_c/m > (K_c/2m)^2$. Despite measuring deviation from a null model, Modularity does not take into account the statistical evidence associated with this deviation. For this reason, Modularity is not able to separate actual communities from those arising only from statistical fluctuations of the null model. Even worse, Modularity can find high-scoring partitions in fully random graphs [91] and in artificially built graphs with no community structure [175].

The configuration model is not the only possible null-model to use in spin-glass based quality functions. Different authors proposed several variants of Modularity [168, 167, 161] with different null models. The simplest variation of Modularity is the so-called ER Modularity [176] that instead of the configuration model uses an Erdős-Rényi random graph in which every edge appears with the same probability p_{ER}. The number of expected edges $\langle m_c \rangle$ in a community of size n_c is thus (in the space of simple graphs):

$$\langle m_c \rangle = p_{ER} \binom{n_c}{2}. \tag{2.9}$$

and plugging this null model into the RB model of Eq. 2.4 we obtain the model:

$$\mathcal{H}^{ER} = -\sum_c^C \left[\frac{m_c}{m} - p_{ER} \binom{n_c}{2} \right]. \tag{2.10}$$

As the most similar ER graph of a given graph is the one that matches its density, the probability parameter p_{ER} must be set equal to the original empirical graph density ρ. The so-called ER Modularity is derived by plugging the graph density into Eq.2.10 and inverting its sign:

$$Q^{ER} = \sum_c^C \left[\frac{m_c}{m} - \rho \binom{n_c}{2} \right]. \tag{2.11}$$

Under this model, a group of nodes forms a community if its internal density is greater than the graph density ρ, on average. Interestingly, from a machine learning perspective, all spin glass models quantify the discrepancy between observed and expected intramodular fraction of edges using a linear loss function. Among all the loss function, the linear is not the only possible one, and loss functions that may take into account the relative size of clusters are desirable.

2.3 RESOLUTION LIMIT

Modularity attracted a lot of attention over the years as it became the tool of election to inspect the community structure of networks. On one side, the vast use of

Modularity led many researchers to gain interest in complex networks, with applications in sociology [177], bioinformatics [178] and ICT [179, 180], just to mention a few. On the other hand, it offered a fallacious view on the community detection problem. Indeed, although simple in many senses, Modularity optimization hid a problem that heavily limits its use in real world networks.

In 2007, a seminal article by Fortunato and Barthelemy [72] did a thorough analysis of Modularity. Their work showed the inability of Modularity to correctly identify communities that are smaller than a certain scale, determined by the square root of the total number of edges. They dubbed this general phenomenon *resolution limit*. To illustrate what is meant by the resolution limit, here we follow the example of Fortunato and Barthelemy, with some obvious notational change.

Let us consider a toy network, $G = (V, E)$ that is composed of three subnetworks, as shown in 2.3A. The first subnetwork, a subgraph G_0 with n_0 nodes and m_0 edges is connected to two cliques, G_1 and G_2 by m_{01} and m_{02} links respectively. The two cliques are also connected by a number of m_{12} links as shown in Figure 2.3. While G_1 and G_2, are complete subgraphs and modules by construction, G_0 may consist of many communities. Maximum Modularity partitions then, should identify G_1 and G_2 as communities, independently from G_0. To be more specific, let the partition where the two cliques are separated be denoted by A, with Modularity value Q_A. On the other hand, the partition where the two cliques are merged into a single community is denoted by B, with Modularity value Q_B. To ease the calculations, we indicate the number of links m_{12} as functions of m_1 and m_2, such that $m_{12} = a_1$, $m_1 = a_2 m_2$, $m_{01} = b_1 m_1$ and $m_{02} = b_2 m_2$ with $a_1, a_2, b_1, b_2 \geqslant 0$.

As Modularity is a sum over the modules, and the module G_0 has the same Modularity Q_0 in both partitions, we are interested in studying the difference $\Delta Q = Q_A - Q_B$. After some simple algebraic manipulations it results:

$$\Delta Q = \frac{2m a_1 m_1 - (a_1 + b_1 + 2)(a_2 + b_2 + 2)m_1 m_2}{2m^2} \quad (2.12)$$

Specifically, we want to find the conditions whereby $\Delta Q > 0$, i.e. when the partition A has a higher Modularity than partition B. This inequality is verified as long as

$$m_2 < \frac{2m a_1}{(a_1 + b_1 + 2)(a_2 + b_2 + 2)}. \quad (2.13)$$

In the case where $a_1 = a_2 = 0$, there are no links between G_1 and G_2 and the condition is satisfied. When the two subgraphs are connected ($m_{12} \neq 0$), it happens that at some values of m_1 and m_2 the partition where the two modules are merged is preferred, i.e. $\Delta Q < 0$. This means that when maximizing Q^N on a network, is possible to miss some important structures, if they are too small. More specifically, if the two modules have the same size and one sets $a_1 = a_2 = b_1 = b_2 = 1/m$, is easy to verify that Eq. 2.12 is not satisfied if the number of links in the modules m_c, is lower than the square root of the total number of links in the network:

$$m_c < \sqrt{\frac{m}{2}}. \quad (2.14)$$

To make this example more concrete, we studied numerically a toy network where the three subgraphs take a precise form. For the sake of illustration, we have

defined G_1 and G_2 as two identical cliques of 5 nodes connected to G_0 by a single edge ($m_{01} = m_{02} = 1$) and to each other by m_{12} edges. The module G_0 was defined as a clique of variable size with a number of edges ranging from 45 to 2775. We then computed the numerical difference ΔQ and plotted it as a function of the number of edges m_0 in the G_0 clique.
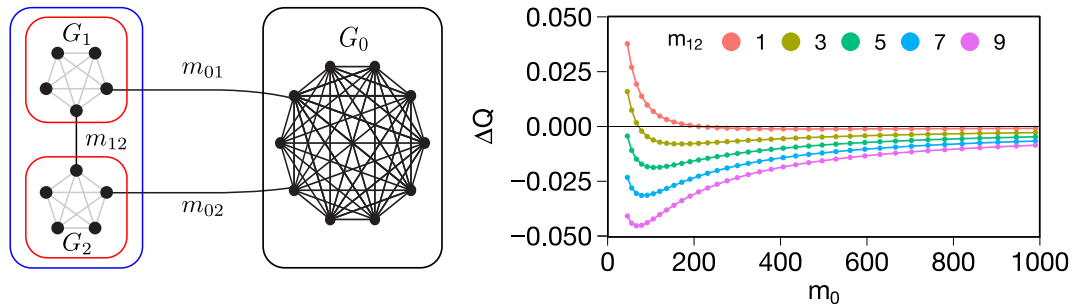
Figure 2.3: Analysis of the onset of the resolution limit for Modularity in a model graph (A) consisting of two cliques, G_1 and G_2, and a size-varying components G_0. The red line indicates the partition A, with G_1 and G_2 as different modules, and the blue line the partition B, with G_1 and G_2 merged into a single module. The graph (B) shows the difference in Modularity for increasing number of edges in G_0.

The onset of the resolution limit occurs when ΔQ changes sign and becomes negative for increasing values of m_0. For $m_{12} = 1$, i.e. when the two cliques G_1 ad G_2 were connected by only one edge (red curve), Q showed this sign inversion for $m_0 \approx 200$ (Figure 2.3B). With an increasing number of intercluster edges m_{12}, the resolution limit appeared for smaller and smaller values of m_0, eventually leading to ΔQ values that were always negative, i.e. the two cliques G_1 and G_2 were always merged by Modularity optimization. An even more striking example of how the resolution limit affects Modularity is when looking at the optimal partition of a synthetic lattice graph, known as "ring of cliques", a network made out of identical cliques, connected in a ring-like structure by single links, as shown in Figure 2.4. In this toy network, the intracommunity density is the highest possible, while the intercommunity density is the lowest as to keep the network connected. If the number of cliques is large enough (they are more than $\sqrt{m}$), Modularity optimization leads to a partition where the cliques are combined into groups of two or more. This phenomenon can be observed already when considering a ring of 30 cliques made of 5 nodes, when the optimal Modularity partition combines cliques in pairs. Yet, extending the number of cliques, Modularity may merge cliques into groups of three, fours etc.

2.3.1 Resolution parameter

Different authors proposed solutions that tried to circumvent the resolution limit: from the introduction of a tunable parameter that enables analysis of the network at an adjustable resolution level [165, 168, 169] to an adequate edge re-weighting to decrease its adverse effects [181].

A

B
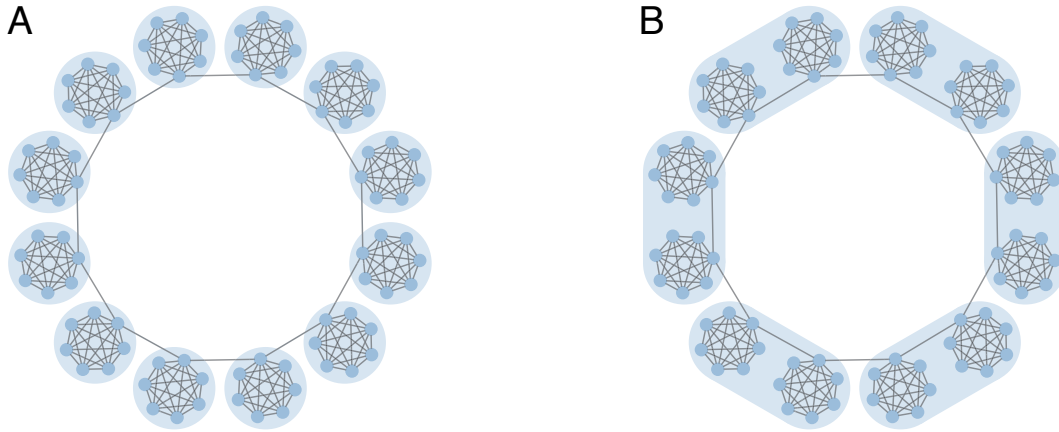
Figure 2.4: The ring of cliques toy model. The left panel (A) shows the ground truth partition with maximal intracluster density and minimal intercluster density. The right panel (B) shows the maximum Modularity partition, whereby cliques have coalesced in pairs.

In this sense, the most common tuning to the Modularity quality function is the one that takes into consideration a *resolution parameter*. From the definition of Q^N is evident that is possible to tune the size of the detected modules by multiplication of the configuration null model by a constant factor $\gamma_N \in [0, 1]$, resulting in a modified modularity $Q^N(\gamma)$:

$$Q^N(\sigma, \gamma) = \sum_c^C \left[\frac{m_c}{m} - \gamma_N \left(\frac{K_c}{2m} \right)^2 \right]. \tag{2.15}$$

However, exact specification of γ_N requires prior knowledge of the expected size of the communities. Moreover, it has been shown [163] that an adjustable resolution parameter may reduce the tendency to merge small clusters, but only at the cost of unduly splitting large clusters. Indeed this is due to two opposite co-existing effects: the tendency to merge small subgraphs, which dominates when the resolution is low and the tendency to split large subgraphs, which dominates when the resolution is high.

In benchmark networks with heterogeneous distributions of cluster sizes, the simultaneous elimination of both biases is not possible and multi-resolution modularity is not capable of recovering the planted community structure even in graphs where the ground-truth structure is evident. Therefore, adjustment of the resolution parameter is an attempt to balance these two biases, but multi-resolution methods fail to recover community structures comprising heterogeneous distributions of cluster sizes [163].

In another attempt to better tolerate the resolution limit, scholars applied appropriate techniques of edge re-weighting [181] to enhance communities prior to Modularity maximization. Although this technique proved able to better tolerate the detrimental effect of the resolution limit, it only shifted the scale of the problem. After all, real-world networks are characterized by the coexistence of clusters of very different sizes, therefore no single parameter can adapt to the variety of network topologies observed in nature.

2.3.2 Resolution-limit free quality functions

In an attempt to quantitatively characterize the resolution limit, Traag et al. [161] proposed a rigorous definition of *resolution-limit-free* graph partitioning.

A quality function is resolution-limit-free if, given an optimal partition ζ of a graph G, any module ζ_i is also optimal for the graph induced by the nodes in ζ_i. In other words, each community of the optimal partition is not split by optimization of the quality function applied to the subgraph induced by the nodes in the community. Hence, each community does not depend on the rest of the network and is both locally and globally optimal. Then, in order to design such resolution-limit-free quality function, Traag took a spin glass model, with a null model specified by constant quantity γ_{CPM} [161] and dubbed it *Constant Potts Model*.

The *Constant Potts Model* (CPM) identifies community as subset of nodes whose internal density ρ_c is bigger than the overall graph density multiplied by a factor γ_{CPM} that defines the typical scale of the communities. In the framework of Reichardt and Bornholdt, the CPM model has the following Hamiltonian:

$$H(\sigma)^{CPM} = - \sum_{(i,j) \in V^2} \left[a_{ij} - \rho\gamma_{CPM} \right] \delta(\sigma_i, \sigma_j), \qquad (2.16)$$

that once reworked in an additive quality function, results in the form of Eq. 2.17

$$Q^{CPM} = \sum_c^{|C|} \left[m_c - \gamma_{CPM} n_c^2 \right] \qquad (2.17)$$

In other words, the model tries to maximize the number of internal edges while at the same time keeping relatively small communities. The parameter γ_{CPM} balances these two imperatives acting as the inner and outer threshold of edge density. Hence, in the CPM settings is better to split two communities r and s if γ_{CPM} exceeds the inter-community density $m_{r,s}/(2n_r n_s)$.

Unfortunately, the operation of tuning the resolution parameter, both in the CPM model as well as in other similar models, is difficult and no widely accepted method exists, making all the models based on a resolution parameter, scarcely used in practical applications.

2.3.3 Effects of resolution limit

In the context of brain networks, the resolution limit first highlighted by Fortunato and Barthelemy may be particularly critical for the analysis of brain connectivity networks, as it may unduly merge modules that are too small, therefore hampering the ability to highlight regions of functional segregation from the rest. Indeed, there are some functional processes in the brain that are thought to be better represented as anatomically localized like the first stages of sensory processing, while others are seen as more globally integrated processing system, in particular attentional processes [142]. Hence, we may expect the brain modular structure to comprise communities of heterogeneous dimensions. Whether the relatively uniform modular structure of brain connectivity, highlighted by Newman's

Modularity and other community detection methods in many studies, reflects the true architecture of the brain organization or is the result of the resolution limit is still unclear [182]. Hierarchical approaches have shown that large modules can be further subdivided, indicating that connectivity networks show structure at different spatial scales [67]. However, these findings do not provide information on the optimal partition of the network, i.e. the optimal cut through the dendrogram representing connectivity at the different scales.

Thus, the resolution limit is a critical shortcoming for the study of brain networks and likely influenced many studies in the literature. The limitation on the number of detected modules in brain networks is evident in studies where typically four or five communities of the same size are detected in humans, as in Crossley et al. [34], Meunier et al. [159, 59], Fair et al. [183] or also in mouse models [36]. To this end, an optimization method that does not suffer from the resolution limit would be needed.

It is also very important to stress that comparing the Modularity value Q^N across different networks obtained in different studies, with possibly a different number of nodes or different densities, is an erroneous practice. As shown by a number of studies [184, 175, 185], even in graphs which do have a natural community structure, high modularity values can be achieved by partitions which do not respect this natural structure. Hence, the Modularity value is only a numerical indication of the current status of optimization of the detected community structure on a specific graph and should never be confronted across different networks.

2.4 DEGENERACY

The differences in Modularity between optimal and suboptimal partitions can be very small, as observed in Figure 2.3B, where ΔQ between the two partitions with split or merged cliques, remains close to zero for a large range of m_0 especially at $m_{12} = 1$. Indeed, even in the case where it would not be beneficial for Modularity to merge two modules, i.e $\Delta Q < 0$, this difference can be made arbitrarily close to zero.

The total number of different partitions in a graph is the Bell number B_n [186] and it grows faster than exponentially in n, therefore a combinatorially large number of sub-optimal partitions exist around the global optimum. These partitions may be very close in terms of Modularity to the optimum but radically different in terms of similarity. Thus, counter-intuitively, when the network becomes more modular, the globally optimal partitions becomes harder to find among the growing number of suboptimal but competitive alternatives.

This consideration explains the empirical observation that nearly-optimal solution tend to group into high-modularity plateaus, although they may differ substantially [184]. This phenomenon, dubbed *degeneracy* affects Newman's Modularity and probably many other spin-glass based quality functions. As a matter of example, the variation in Modularity for merging a pair of adjacent cliques in the ring of cliques graph shown in Figure 2.4, is given by:

$$\Delta Q = \frac{1}{r\left(\binom{n_r}{2} + 1\right)} - 2r^{-2} \tag{2.18}$$

where r is the number of cliques and n_r the number of nodes in each clique. In the large r limit, the difference in Eq. 2.18 tends to a small negative value, indicating that a solution where the cliques are merged, may have a Modularity very close to the optimum. Indeed, already for $r = 20$ cliques, $\Delta Q \approx 5 \times 10^{-3}$, making all the suboptimal partitions where cliques are merged, very close to the optimal solution.

To make this argument more intuitive, I extrapolated a visually interpretable form of the complex landscape of partitions' Modularity for a ring of cliques network. I sampled the configuration space of the partitions of a ring of clique network through a Montecarlo procedure and annotated the corresponding values of Modularity for every partition as in [184]. I then built a similarity matrix between all sampled partitions and embedded it into a three-dimensional space maintaining similarity relations between partition with the help of Curvilinear Components Analysis (CCA). In the embedded manifold, two partitions are close if they are similar, and the z-axis encodes the quality function. A highly degenerate plateau can be observed in Figure 2.5 whereby different solutions with high values of Modularity lay in the same neighborhood.
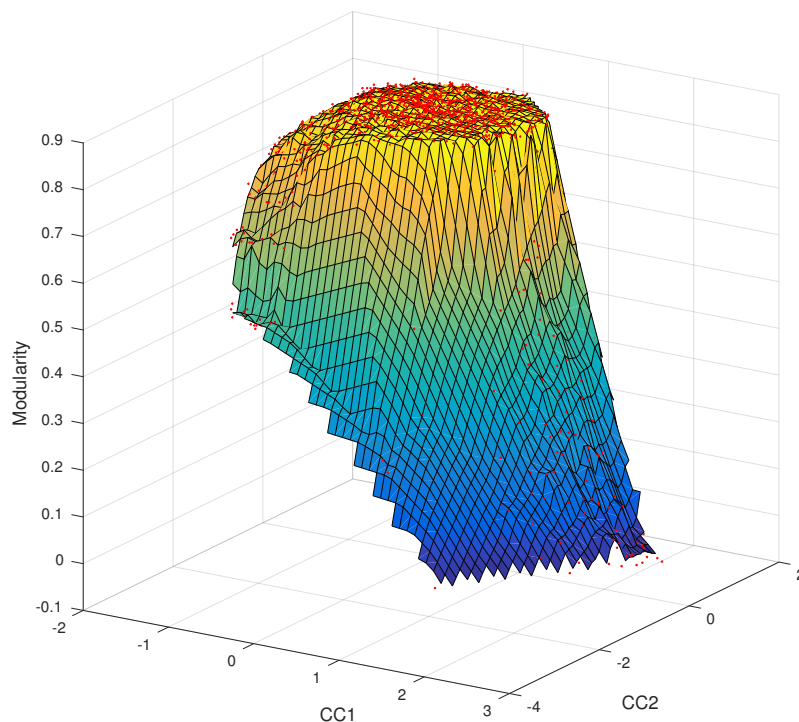
Figure 2.5: Degeneracy landscape for Newman's Modularity in a ring of cliques with $r = 24$ cliques of $n_r = 5$ nodes. The axes CC1 and CC2 are complicated functions of the original partition space computed as to maintain the distance relation between points and their scale is unimportant.

Like for Modularity, any other quality function afflicted from the degeneracy problem, displays a landscape with a plateau of optimal partitions, like the one shown in Figure 2.5, whereas in the case of non-degenerate quality functions, a sharp and distinct peak exists in the embedded landscape.

Several approaches have been proposed to mitigate the problems that are apparent in Modularity. These include the use of multiple near-optimal solutions to

avoid the pitfalls of degeneracy. Moreover, a proper choice of the null model may help overcome the fundamental resolution limit.

The issue of degeneracy affects Newman's Modularity Q^N and other quality functions as well. In this respect, it is clear that choosing the partition with the highest quality value is not meaningful and a choice that considers all the good partitions at the plateau may be more desirable. An optimal partition of the network can then be expressed as a *median* of the solutions over the optimal plateau. This approach, known as *consensus clustering*, averages local effects of noise during optimization over a large set of solutions. Consensus clustering is typically based on a meta-algorithm rather than a proper community detection method. Indeed, given a community detection method, consensus clustering forms an ensemble of optimal solutions from which an association matrix $\mathcal{A}$ is computed, where edges are proportional to the probability that two nodes are connected in the same community. Consensus communities are then obtained iteratively by clustering the thresholded association matrix. The choice of the threshold that makes the association matrix sparser [187] is dictated by the properties of the chosen method of community detection itself. Typically with a judicious choice of the threshold, consensus clustering converges in a few iterations. The pseudo-code reported in Algorithm 1 illustrates all the steps of consensus clustering.

ConsensusClustering(a graph G, a community detection method)

1 Apply community detection on G for n_p times, yielding n_p partitions.
2 Compute the consensus (association) matrix $\mathcal{A}$ where $\mathcal{A}_{ij}$ is the probability that vertex i and j are assigned in the same community over n_p partitions.
3 Threshold the association matrix $\mathcal{A}$ with a parameter τ.
4 Apply community detection on $\mathcal{A}$ for n_p times to yield n_p partitions.
5 If $\mathcal{A}$ is block diagonal (all partitions are equal) stop, else return to 1.

Algorithm 1: Pseudocode for the implementation of consensus clustering.

Although useful to address the degeneracy issue, this approach still suffers from the resolution limit, as the near-optimal optimal solutions which the consensus partition is made of, are resolution-limited per se.

2.5 MODULARITY OPTIMIZATION: THE LOUVAIN METHOD

Optimization of fitness function in the additive form of Eq. 2.4 is typically performed using the Louvain method [171], a greedy stochastic agglomerative clustering algorithm that works on hierarchical refinements of the network's partitions. The inspiration for this method of community detection is the optimization of Modularity as the algorithm progresses. In the Louvain method of community detection, first small communities are found by optimizing modularity locally on all nodes, then each small community is grouped into one node and the first step is repeated.

This algorithm has become extremely popular in recent years because it is fast, allowing to analyze huge networks with billions of edges [69]. Due to its intrinsic stochastic nature, the Louvain method returns different near-optimal partitions for every run and must be typically run thousands of times to identify the best solution.

The optimization method is divided in two phases applied recursively as depicted in Figure 2.6. The first phase is the optimization phase which looks for a locally optimal partition by considering only individual vertex swaps. As in many other greedy algorithms, the community partition is initialized with a single node per community and as many communities as nodes in the input network. Then, based on the chosen cost function, individual nodes are removed from their current community and swapped to the neighboring community which produces the largest positive gain of the cost function. This procedure is repeated iteratively at different levels, until no further movement produces improvements in the quality function.
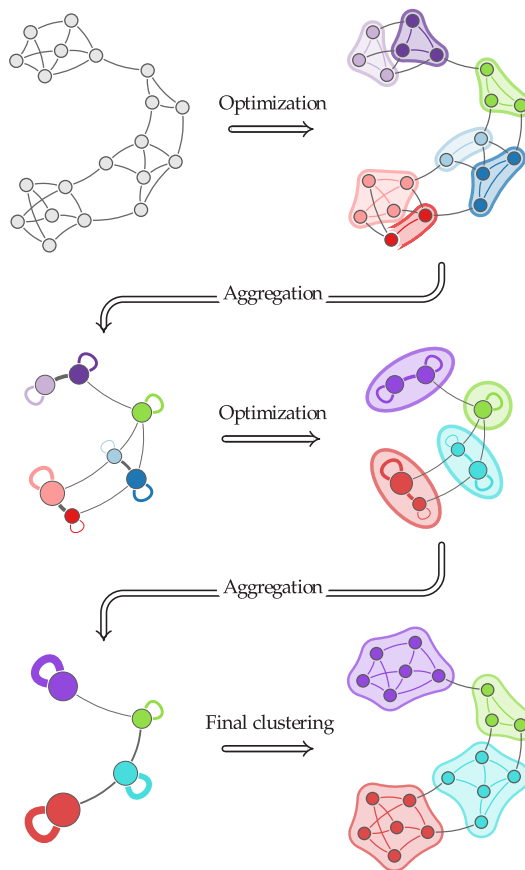
Figure 2.6: Optimization and aggregation steps with the Louvain method. Figure taken from [188].

2.5.1 Spectral optimization

Another (and earlier) method to optimize Modularity is the spectral Modularity by Newman and Girvan [71]. Given the form of Modularity as in Eq. 2.6, one calls

the quantity $\mathbf{B} = A_{ij} - k_i k_j / 2m$ *modularity matrix*. From it, it is possible to obtain an equivalent linear algebra formulation of Modularity that reads:

$$Q^N = \frac{1}{4m} \mathbf{s}^\top \mathbf{B} \mathbf{s} \qquad (2.19)$$

where $\mathbf{s}$ is a numerical vector that considers a division of the network in two communities, with values $s_i = +1$ if vertex i belongs to group 1 and $s_i = -1$ if vertex i belongs to group 2. With some manipulations [71, 73], the Modularity maximization is turned into an symmetric eigenvalue problem of the form

$$\mathbf{B}\mathbf{s} = \beta \mathbf{s}. \qquad (2.20)$$

The group assignment of each node is obtained as the signs of the eigenvector $\mathbf{s}$ corresponding to the greatest eigenvalue β. Finally, by repeatedly subdividing the network, the optimal Modularity assignment is recovered [71, 73].

2.6 CONCLUSIONS

The modular organization of brain functional connectivity networks has been investigated by a variety of methods and specifically by graph theoretical approaches. The rich repertoire of behavioural and cognitive aspects of everyday life is supported by integration of local processing that is reflected in the functional modules, as identified by a variety of protocols. Unfortunately, Newman's Modularity, the most widely used graph-based method for community detection in FC networks is flawed as it cannot detect modules that are smaller than a certain scale, making impossible to ascertain important details of the architectural structure. The two main drawbacks of Modularity have been introduced namely the resolution limit and its extreme degeneracy, together with analyses that specifically address their origin. In the following chapter, I will illustrate how to overcome the limitations imposed by Newman's Modularity with an approach that has its roots in probability theory: Surprise.

3 | BEYOND THE RESOLUTION LIMIT IN BINARY NETWORKS

In the last chapter I have introduced the framework of network science used to study brain networks. I discussed the modular organization of brain networks and the graph theoretical approaches that enabled its study. I have also shown that graph partitioning methods based on the maximization of global fitness functions, like Newman's Modularity, suffer from a fundamental resolution limit as they fail to detect modules that are smaller than a scale determined by the size of the entire graph. Furthermore, I explored the effects of this limitation on the study of brain connectivity networks, demonstrating that the resolution limit prevents detection of significant details of the modular brain structure, thus hampering the ability to appreciate differences between networks and to assess the topological roles of nodes.

In this chapter I'll show that Surprise, a recently proposed fitness function based on probability theory, does not suffer from these limitations. I'll introduce Surprise from its theoretical foundations and discuss its properties in details. Moreover, I'll describe an algorithm for Surprise optimization in binary graphs with an assessment of its performance on benchmark networks. Surprise maximization in brain co-activation and functional connectivity resting state networks reveals the presence of a rich structure of heterogeneously distributed modules, and differences in networks' partitions that are undetectable by resolution-limited methods. Moreover, Surprise leads to a more accurate identification of the network's connector hubs, the elements that integrate the brain modules into a cohesive structure.

3.1 PROBABILISTIC APPROACH TO CLUSTERING IN BINARY NETWORKS

The spin-glass based additive quality functions described in the previous chapter, including Newman's Modularity, measure the deviation between the fraction of edges falling within modules and the null model. None of these, though, assesses the statistical significance of this deviation. In other terms, no spin-glass based quality function is telling us the level of statistical confidence at which one should discard the null hypothesis that the observed fraction of edges inside a community is the same as the one expected from a given null model.
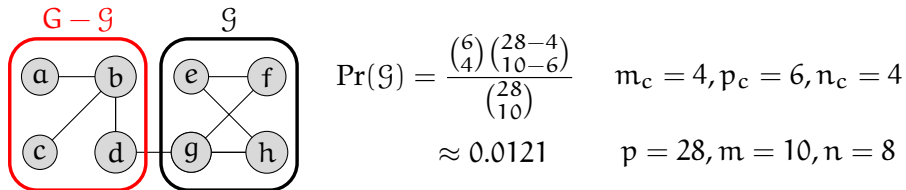
To address this question, one should instead compute probabilities, as they are the natural way to measure statistical significance. Specifically, given a node induced subgraph (a community), one is interested in computing the probability to find another subgraph having more internal edges than the observed one: the lesser the value, the more significant is the considered subgraph.

More precisely, the problem can be expressed as in the following: given a subgraph, what is the probability of observing another subgraph with a larger number

of internal edges drawn from a random graph with a given density? This question can be answered by the urn model in classical probability theory. The reason is clear if one imagines that each pair of nodes is a marble, which is one color if the nodes are connected by a link and the other color if they are not. A number of marbles is drawn from an urn without replacement, and the probability of observing a given number of marbles of one specific color is calculated using the hypergeometric law. When applied to clustering, the problem corresponds to the calculation of the probability to observe a fixed number of internal edges in a randomly drawn set of nodes defining a community. For example, suppose that one draws a subgraph with n_c nodes, p_c pairs of nodes and m_c edges from a graph with n nodes, p pairs of nodes and m edges (see Section for the notation 1.3.1). As indicated by the urn model, the probability to observe exactly m_c internal edges is given by:

$$\Pr[i = m_c] = \frac{\binom{m}{i}\binom{p-m}{p_c-i}}{\binom{p}{p_c}} = \frac{\binom{p_c}{i}\binom{p-p_c}{m-i}}{\binom{p}{m}} \tag{3.1}$$

where the last equality is because of the Vandermonde identity [189]. The probability in Eq. 3.1 is simply understood in terms of urn model as there are $\binom{p_c}{i}$ ways of choosing exactly i black marbles from a population of p_c black marbles, $\binom{p-p_c}{m-i}$ ways of choosing $(m-i)$ white marbles from a population of $(p-p_c)$ white marbles, and $\binom{p}{m}$ total possible number of combinations of m marbles taken from a population of p of them. Figure 3.1 shows an example where the probability of randomly picking the subgraph $\mathcal{G}$ from the graph G is computed by means of the urn model. As the probability of observing i or more internal edges, is given by

$$\Pr(\mathcal{G}) = \frac{\binom{6}{4}\binom{28-4}{10-6}}{\binom{28}{10}} \qquad m_c = 4, p_c = 6, n_c = 4$$

$$\approx 0.0121 \qquad p = 28, m = 10, n = 8$$

Figure 3.1: Probability for the subgraph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ where $\mathcal{V} = \{e, f, g, h\}$ and $\mathcal{E} = \{(e, f), (f, g), (g, h), (e, h)\}$ to be randomly drawn from the graph G.

the sum of the probabilities of observing exactly $i, i+1, i+2$ etc. internal edges, summing on i yields the probability of getting *at least* m_c white marbles:

$$\Pr[m_c \geqslant i] = \sum_{i=m_c}^{m} \frac{\binom{p_c}{i}\binom{p-p_c}{m-i}}{\binom{p}{m}}. \tag{3.2}$$

In this last equation we are considering the probability of randomly drawing a subgraph with m_c or more internal edges over the set of all random subgraphs with n nodes (p pairs of nodes) and exactly m edges as in the G_{nm} model described in section 1.3.4. Indeed, the denominator $\binom{p}{m}$ in Equation 3.2 represents the possible number of graphs with p pairs of nodes and m edges.

3.2 SURPRISE

The intuition captured by Eq. 3.2 is very close to a quality function dubbed *Surprise*, firstly introduced in a seminal paper by Aldecoa and Marin [190]. To the best of my knowledge, this clearer theoretical motivation did not appear in any other works. If one considers as the random subgraph that is drawn from the $G_{n,m}$ set, a subgraph with at least $m_\zeta = \sum_c m_c$ internal edges and $p_\zeta = \sum_c p_c$ pairs of edges, then this subgraph represents a whole clustering. In this case, the definition given in Eq. 3.2 and the definition given in [190] are perfectly corresponding. Indeed, for a partition ζ, the probability that a subgraph $\mathcal{G}$ randomly drawn from the set G_{nm} has at least m_ζ intracluster edges and p_ζ intracluster pairs is modeled after the inverse cumulative hypergeometric distribution, exactly as in 3.2. Here and in the rest of the work, the quality function *Surprise* is therefore expressed by the following definition:

$$S(\zeta) := \sum_{i=m_\zeta}^{m} \frac{\binom{p_\zeta}{i}\binom{p-p_\zeta}{m-i}}{\binom{p}{m}}. \tag{3.3}$$

Surprise computes the probability to (surprisingly) observe at least as many internal edges as within the proposed partition in a uniform random graph. As mentioned, model in Eq. 3.3 corresponds to an urn model without reinsertion, where S is the probability of extracting at least m_ζ white balls out of m trials from an urn containing p_ζ white balls and $p - p_\zeta$ black balls. Intuitively, the lower $S(\zeta)$, the better the clustering. Optimal partitions with respect to S are those with the highest number of intracluster edges and the smallest number of intracluster pairs.

Differently from Modularity, $S(\zeta)$ equals one both for the partition where every node is in a separated community ($|\zeta| = n$) and for the partition that entails all nodes into a single community ($|\zeta| = 1$), as is evident from its formulation in terms of urn model. Contrarily, Newman's Modularity is zero when $|\zeta| = 1$, but different from zero in the case $|\zeta| = n$. In terms of Surprise, these two extreme partitions are thus equally uninformative. This can be understood as for a single community comprising all nodes, the probability of observing a number of intracluster edges greater or equal than the number of edges is one. On the other hand, if the partitioning has zero intracluster edges (the case of n separate communities), the probability to observe more than zero intracluster edges still equals one.

It should be noted that, due to numerical precision problems in the evaluation of large binomial coefficients, $\hat{S}(\zeta) := -\log_{10} S(\zeta)$ is often taken as measure of quality of the partition, which is totally equivalent with respect to optimum solutions, whereas higher values correspond to better clustering. Different authors [191, 192] refer to S as Surprise, whereas others [190, 193] use $\hat{S}$. Hereafter I stick to the notation of [192] where Surprise is indicated as S defined in Eq. 3.3 and indicate $\hat{S}$ where needed.

3.2.1 Statistical test interpretation

Surprise considers the problem of community detection as the one of making the intracluster density as further as possible from the global density in statistical

terms. In this sense, it is worth noting that S is a *p*-value of a one-tailed Fisher exact-test where one is asking how confidently should reject the null hypothesis that the intracluster density is greater than the graph density. It turns indeed out that this problem has an equivalent description in statistics, where one seeks to maximize the *odds-ratio* of the 2×2 contingency table defined in Table 3.1.

The Fisher exact test implemented by Surprise is, as its name states, exact as long as the contingency table keeps the row and column totals fixed, and it can be used regardless of the sample characteristics. A simpler χ^2 statistic can be used when the elements of the contingency table are large enough, although only an approximation of the *p*-value can be obtained. In this case, it is possible to tackle the problem of computation of S by means of odds-ratio. Precisely, the *normalized log odds-ratio* is computed as:

$$\log(\text{OR}) = \log\left(\frac{m_\zeta(p - m - p_\zeta + m_\zeta)}{(m - m_\zeta)(p_\zeta - m_\zeta)}\right), \tag{3.4}$$

and asymptotically Surprise is equal to the probability:

$$\Pr\left(z < -\frac{|\log \text{OR}|}{\text{SE}}\right) \tag{3.5}$$

where z is a random variable with standard normal distribution $z \approx \mathcal{N}(0, 1)$ and SE is the standard error of the odds-ratio, approximately computed as

$$\text{SE} = \sqrt{\frac{1}{m_\zeta} + \frac{1}{(p_\zeta - m_\zeta)} + \frac{1}{(m - m_\zeta)} + \frac{1}{(p - m - p_\zeta + m_\zeta)}}. \tag{3.6}$$

Although not interesting in the case of binary graphs, Equation 3.5 is telling us that in developing a version of Surprise that will keep into account weighted graphs, we should in some way rely on its asymptotic distribution.

	Drawn	Not drawn	**Total**
Intracluster	m_ζ	$p_\zeta - m_\zeta$	p_ζ
Intercluster	$m - m_\zeta$	$p - m - p_\zeta + m_\zeta$	$p - p_\zeta$
Total	m	$p - m$	p

Table 3.1: Contingency table for the urn model.

3.2.2 General properties of Surprise

As noted by [192], for a given graph, m and p are fixed so $S(\zeta)$ is depending only on the number of intracluster edges m_ζ and intracluster pairs p_ζ, namely $S := S(m_\zeta, p_\zeta)$. In establishing the domain of validity of (m, p, m_ζ, p_ζ) the urn model is of help. A valid clustering ζ of mutually disjoint communities automatically satisfies intuitive but important requirements:

Observation 1. *it is impossible to draw more white balls than the white balls contained in the urn, therefore* $p_\zeta \geqslant m_\zeta$.

Observation 2. *it is impossible to have more white balls than total number of drawn balls, therefore* $m \geqslant m_\zeta$.

Observation 3. *it is impossible to draw more black balls than the black balls contained in the urn, therefore* $p - p_\zeta \geqslant m - m_\zeta$.

All valid clusterings are enclosed inside the domain $\mathcal{L}$ defined as from these last three observation:

$$\mathcal{L} := \{(m_\zeta, p_\zeta) \, | m_\zeta > 0 \wedge p_\zeta \geqslant m_\zeta \wedge m \geqslant m_\zeta \wedge p - m > p_\zeta - m_\zeta\} \qquad (3.7)$$

The urn models gives other three important properties, indicating that Surprise is a convex function inside the domain of validity of its variables. Specifically as shown in [192] three inequalities apply for Surprise and for the urn model in general.

Proposition 1. *it is less probable to draw at least* $m_\zeta + 1$ *than* m_ζ *white balls if the urn contains the same number of* p_ζ *white balls, therefore* $S(m_\zeta + 1, p_\zeta) < S(m_\zeta, p_\zeta)$.

Proposition 2. *it is less probable to draw at least* m_ζ *white balls if the urn has one white balls less, therefore* $S(m_\zeta, p_\zeta - 1) < S(m_\zeta, p_\zeta)$.

Proposition 3. *it is less probable to draw at least* $m_\zeta + 1$ *white balls if the urn has* $p_\zeta + 1$ *white balls, than drawing at least* m_ζ *white balls if the urn has* p_ζ *white balls, therefore* $S(m_\zeta + 1, p_\zeta + 1) < S(m_\zeta, p_\zeta)$.

The scheme in Figure 3.2 shows the order relation between the elements of the three aforementioned inequalities.
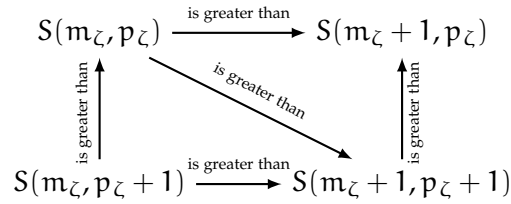
Figure 3.2: Behaviour of the landscape of Surprise S.

A fourth inequality $S(m_\zeta, p_\zeta + 1) > S(m_\zeta + 1, p_\zeta)$ is implicit by looking at the behaviour of $S(\zeta)$ for a given graph. The scheme in Figure 3.2 allows to rank the values of Surprise in relation to changes in m_ζ and p_ζ. it is to easy verify that S satisfies the following strict order relation:

$$S(m_\zeta + 1, p_\zeta) < S(m_\zeta + 1, p_\zeta + 1) < S(m_\zeta, p_\zeta) < S(m_\zeta, p_\zeta + 1). \qquad (3.8)$$

This property led me to the observation that Surprise S is a monotonically decreasing function (increasing $\hat{S}$) of m_ζ and monotonically increasing function (decreasing $\hat{S}$) of p_ζ in the convex interval $\mathcal{L}$. Moreover, from Figure 3.2 follows that optimal solutions with respect to S are Pareto-optimal with respect to maximizing m_ζ and minimizing p_ζ. In other words, a partition that is optimum with respect to Surprise is such that no further increment in m_ζ leads to a decrement in p_ζ that has lower S (higher $\hat{S}$). The Pareto optimality of optimum Surprise partitions, implies that a perturbation of $\delta m_\zeta, \delta p_\zeta$ leads to a change in Surprise such that it monotonically decreases if and only if $\delta m_\zeta > \delta p_\zeta$, precisely:

$$S(m_\zeta + \delta_{m_\zeta}, p_\zeta + \delta_{p_\zeta}) < S(m_\zeta, p_\zeta) \iff \delta_{m_\zeta} \geqslant \delta_{p_\zeta}. \qquad (3.9)$$

The result in Eq.3.9 is of great help in designing optimization algorithms as every move that increases the number of intracluster edges more than intracluster pairs is good, while moves that increase intracluster pairs more than intracluster edges must be ignored, therefore sparing time for the computation of Surprise. Additionally, Eq. 3.9 is useful when analyzing the onset of the resolution limit for different models. In the next sections I will show to what extent Surprise is affected by the resolution limit for Surprise, to show with convincing theoretical arguments that S is nearly resolution-limit-free in Traag's sense [176].

3.3 RESOLUTION LIMIT AND SURPRISE

Unfortunately, the resolution limit is an intrinsic feature of many quality functions, and there appears to be a "narrow scope to resolution-limit-free methods" [176]. Surprise has been shown to outperform other network partitioning methods in the detection of small features within large graphs, but the extent to which it suffered from the resolution limit was unknown [190, 193], until our work [182].

Aldecoa [190] pointed out that while Modularity-based methods define a community as a region with an unexpectedly high density of links with respect to the global characteristics of the network, Surprise weights the number of actual intracluster edges against the maximum number of links given the nodes in the clusters. Hence, Surprise can discriminate local subnetworks whose internal density is close to that of a clique independently of their size.

To assess the extent to which Surprise is affected by the resolution limit, I directly compared Newman's Modularity and Surprise in the example of Fortunato and Barthelemy illustrated in the previous chapter in Figure 2.3. Already for $m_{12} = 1$, i.e. when the two cliques G_1 and G_2 were connected by only one edge, Q^N showed sign inversion for $m_0 \approx 200$, meaning that it was beneficial for Modularity to merge two cliques. In this same example, Surprise never merges cliques, as shown in Figure 3.3.
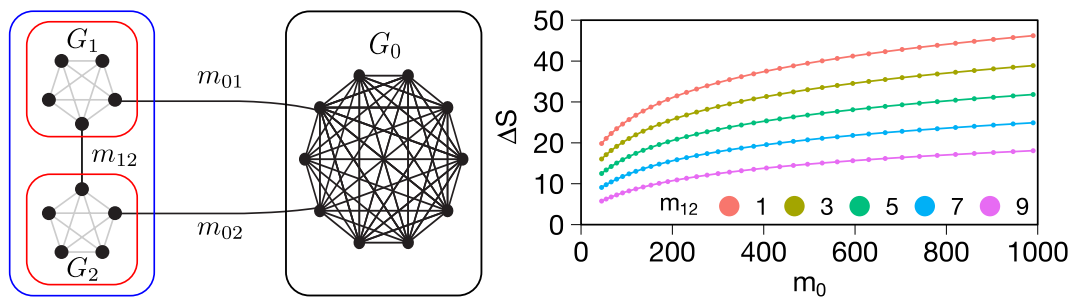
Figure 3.3: Difference in $\hat{S}$ between the partition where the two cliques are separated (red) and the partition where the two cliques are merged (blue), as in the benchmark network of Fortunato and Barthelemy. As the quantity $\Delta\hat{S}$ is always positive, Surprise is never going to merge cliques, independently from other global parameters. Adapted from Nicolini and Bifone [182].

Figure 3.3 shows that Surprise does not suffer from the resolution limit at least in this specific case. Indeed, $\Delta\hat{S}$ is always positive and grows monotonically with increasing m_0. Hence, the two cliques G_1 and G_2 are always resolved by Surprise

as separate communities independently of the network size, and also in the presence of some "fuzziness", i.e. when $m_{12} > 1$ and the two cliques are connected by more than one edge. In order to assess whether this behavior reflects a general property of Surprise, or is incidental to this particular example, I have also studied a generalization of Fortunato and Barthelemy's model.

A consequence of the definition of a resolution limit free quality function in Traag's sense, is that such method will never depend on the size of the network to merge cliques in a graph comprising r cliques of n_r nodes connected in a ring structure as in Figure 2.4 (often called the "ring of cliques" model). This observation prompted me to explore the behavior of $\hat{S}$ in the ring of cliques model graph, as an extension of Fortunato and Barthelemy's model. Interestingly, given its two-variables formulation, Surprise optimization can be seen as a multiobjective optimization problem where one seeks to minimize the intracluster pairs while maximizing the number of intracluster edges. With increasing graph size, the computational problem of calculating $\hat{S}$ for every possible partition becomes rapidly intractable (maximization of S is NP hard) [192]. However, as explained before and pointed out by Fleck et al. [192], the Surprise optimal clustering must be Pareto optimal with respect to minimizing p_ζ and maximizing m_ζ, i.e. any further improvement in one of the two variables must occur at the expense of the other.

In this sense, the problem of Surprise optimization is shown to be equivalent to a linear program [192], where one seeks to minimize p_ζ while keeping m_ζ equal to a constant h from 1 to m, choosing then among the resulting m clusterings, the one that maximizes $\hat{S}$. Hence, to delineate the Pareto frontier in the (m_ζ, p_ζ) space, we need to solve m integer linear programs (ILP) in the form:

$$\text{minimize} \sum_{\{i,j\} \in \binom{V}{2}} x_{ij}$$
$$\text{s.t.} \quad x_{ij} \in \{0, 1\}$$
$$x_{ik} + x_{ki} - x_{ij} \leqslant 1$$
$$\sum_{\{i,j\} \in E} x_{ij} = h$$

where x_{ij}, equivalent to $\delta(\sigma_i, \sigma_j)$, are a set of $\binom{n}{2}$ binary variables corresponding to vertex pairs, with the interpretation that $x_{ij} = 1$ if vertex i and vertex j are in the same community. The alternative objective function $\sum_{ij} x_{ij}$ measures the number of intracluster pairs p_ζ, while the number of intracluster edges $m_\zeta = \sum_{ij \in E} x_{ij}$ was set equal to a fixed $h \in [0, m]$. The remaining constraints are necessary in order to ensure transitivity, i.e. if nodes i and j are in the same community, nodes i and k are in the same community, then nodes j and k share the same community too. Figure 3.4 shows the Pareto frontier for a ring of cliques where I independently varied the number of cliques r and the number of nodes n in every clique[1]. As $\hat{S}$ reaches its minimum (zero) both in the case where all nodes are separated communities ($p_\zeta = 0$) or when a single community entails all nodes ($p_\zeta = p$), Figure 3.5 shows monotonically increasing Surprise along the frontier

---

1 Linear programs were solved using the Python interfaces of Gurobi 5.7.3 on Linux (Gurobi Optimizer Version 5.7, Gurobi Optimization, Inc., Houston, Texas, United States).

with increasing p_ζ up to the Surprise optimum, indicated by black circles in the Pareto front of Figure 3.4, whereas the corresponding partition identified each clique as a separate community. Importantly, in the range of parameters I have investigated, Surprise optimization never merged cliques in the ring of cliques, independently of the size of the graph, and behaved as a Traag's resolution-limit free method.
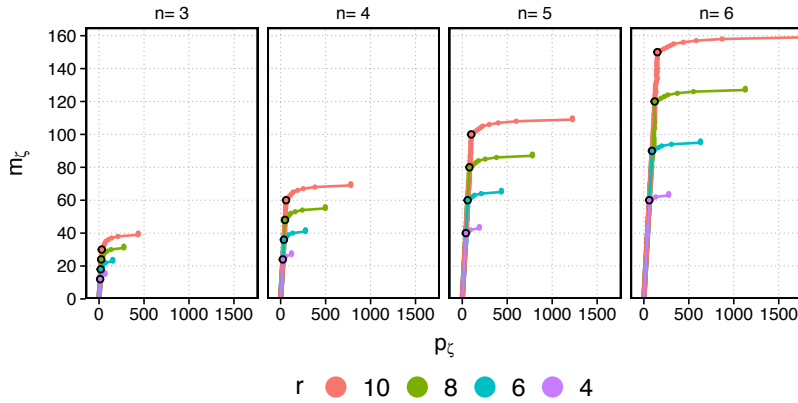
Figure 3.4: Pareto front for a ring of cliques graph. Optimal solutions with respect to Surprise are indicated as small black dots. Adapted from Nicolini and Bifone [182].
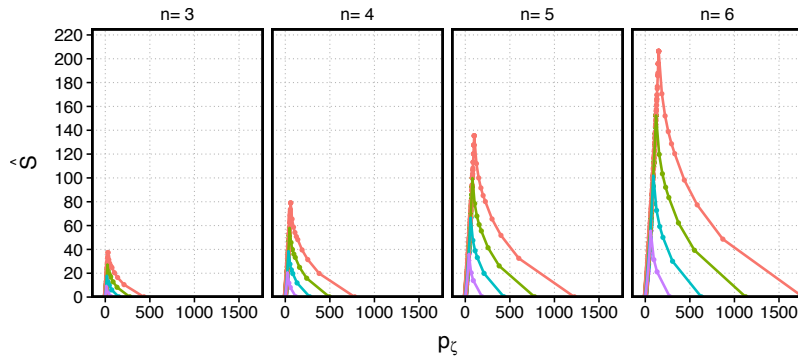
Figure 3.5: Surprise $\hat{S}$ for a ring of cliques graph at different levels of p_ζ for corresponding solutions at fixed m_ζ. The peak of $\hat{S}$ is always reached at p_ζ when the modules entail single cliques separately. Adapted from Nicolini and Bifone [182].

While it is likely that this property is quite general and can be extended to every ring of cliques, an analytical demonstration is hampered by the non-additivity of the Surprise function. Nonetheless, the size of the graphs explored numerically is quite typical of brain connectivity networks and I felt encouraged to apply Surprise maximization to the study of the community structure of the brain.

3.4 DEGENERACY OF SURPRISE

Interestingly, the maximum value of Surprise for given (m_ζ, p_ζ) is sharply peaked and very different from other partitions, indicating that at least in this case Surprise shows no degeneracy of optimal solutions, as instead shown for Modularity

in Section 2.4. This observation led me to investigate to what extent Surprise is affected by the degeneracy problem (previously described in section 2.4). I repeated the procedure indicated in section 2.4, embedding the complex landscape of partitions of a ring of cliques graph in a three-dimensional space to show whether optimal Surprise solutions group in a broad plateau of high Surprise values or not. As evident from Figure 3.6 instead, the global optimum sits on a sharply distinct point, at the highest value of Surprise. This observation makes evident that the Surprise does not show the problematic degeneracy of optimal solutions, but instead has a global optimum that is very different from other suboptimal partitions, at least in these ring of cliques example.
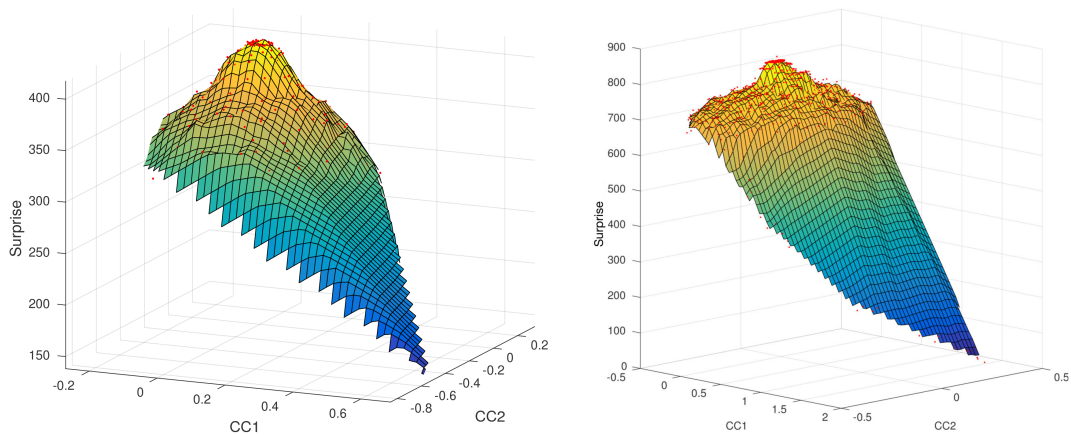
Figure 3.6: Degeneracy of Surprise landscape in a small (left, 24 cliques with 5 nodes) and large (right, 30 cliques with 6 nodes) ring of cliques graph. No plateaus exist, indicating the uniqueness of the global maximum that emerges from the other local optima.

3.5 MAXIMIZATION OF SURPRISE: FAGSO

Community detection is a NP-hard problem, and heuristics have to be developed for the optimization of quality functions for relatively large networks. In their original paper, Aldecoa et al. [190] applied metaheuristics, involving the evaluation of Surprise for partitions resulting from seven different community detection methods, each of those maximizing different quality functions. Here, I describe direct maximization of Surprise by exploiting FAGSO [194] (Fast AGlomerative Surprise Optimization), an agglomerative optimization algorithm that builds on a variation of the Kruskal algorithm for minimum spanning tree [75]. The detailed pseudocode of this algorithm is reported in Algorithm 2 and illustrated step by step on an example network in Figure 3.7.

The first step of FAGSO consists in ranking the edges in the graph in decreasing order by the Jaccard index of the neighbors of their two endpoints nodes. An union-find data structure is used to hold the community structure throughout the computation. In the beginning, each community consists only of one vertex. Then, starting from the edge with the highest Jaccard index at the top of the list, the endpoints are attributed to the same community by disjoint-set union if this

operation leads to a strictly better Surprise and if they do not belong already in the same community. This step is repeated for all edges and the final community structure is returned in the disjoint-set. FAGSO finds partitions with high Surprise and it is deterministic, unless two edges with the same Jaccard index are found. In this case, ties are broken at random. The implementation of FAGSO in `C++`, `Python` and `GNU Octave` is freely available at `https://github.com/carlonicolini/fagso`.
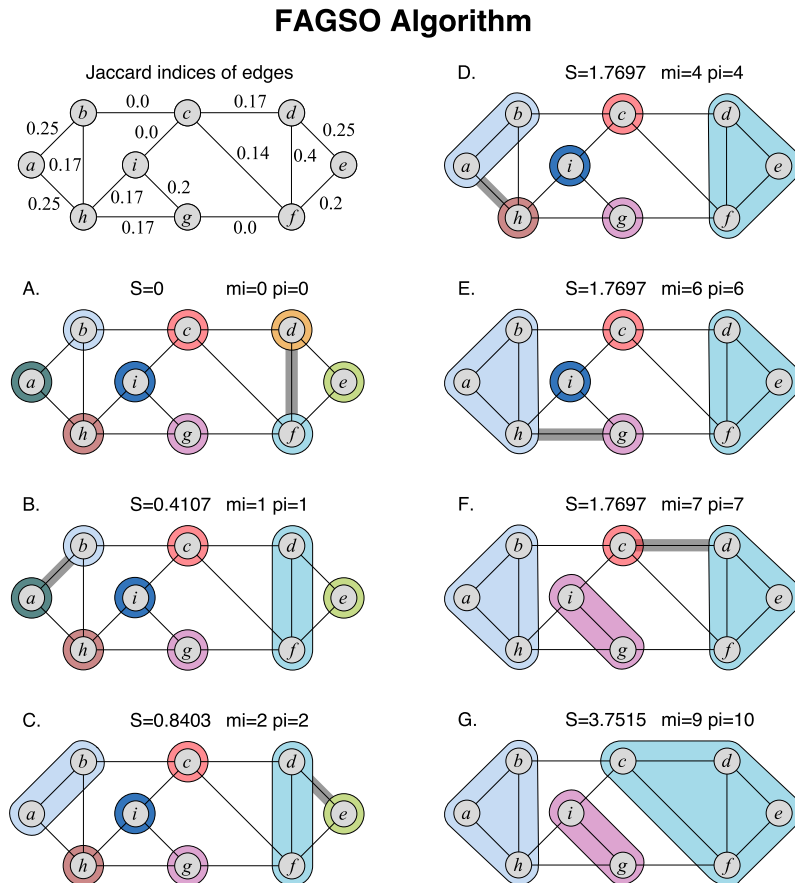
FAGSO Algorithm

Figure 3.7: FAGSO algorithm steps illustrated. Firstly, Jaccard indices are computed. In box A., every node is associated to a separate community and FAGSO evaluates whether to merge nodes d and f in the same community as the edge they form has the highest Jaccard index. Since this merge results in an increased value of Surprise, d and f are merged in box B. then FAGSO proceeds to evaluate the merges of nodes a and b as their Jaccard index is the second greatest. FAGSO merges the two nodes in a second community and proceeds updating the Surprise value and analyzing the subsequent merges in boxes D to F, until no moves are available and FAGSO terminates with the community structure in box G.

FAGSO(G)

1   $S \leftarrow 0 \triangleright$ *Initialize Surprise to 0*
2   $D \leftarrow \emptyset \triangleright$ *Initialize disjoint set forest*
3   **for** each vertex v in $V[G]$
4   **do** MAKE-SET(v)
5   $E' \leftarrow$ SORT-JACCARD(E) $\triangleright$ *Sort edges in decreasing order by Jaccard index*
6   **for** each edge$(u,v) \in E'$, $\triangleright$ *Taken in decreasing order by Jaccard index*
7   **do if** FIND-SET(u) $\neq$ FIND-SET(v)
8   **then if** SURPRISE($G, D \cup \{(u,v)\}$) $> S$
9               $D \leftarrow D \cup \{(u,v)\}$
10   UNION(u,v) $\triangleright$ *Merge the communities u and v belong*
11              $S =$ SURPRISE(G, D) $\triangleright$ *Update current Surprise*
12 **return** D

MAKE-SET(x)

1   $p[x] \leftarrow x$
2   $rank[x] \leftarrow 0$

LINK(x, y)

1   **if** $rank[x] > rank[y]$
2   **then** $p[y] \leftarrow x$
3   **else** $p[x] \leftarrow y$
4   **if** $rank[x] = rank[y]$
5   **then** $rank[y] \leftarrow rank[y] + 1$

UNION(x, y)

1   LINK(FIND-SET(x),FIND-SET(y))

FIND-SET(x)

1   **if** $x \neq p[x]$
2   **then** $p[x] \leftarrow$ FIND-SET($p[x]$)
3   **return** $p[x]$

SURPRISE(G, D)

1   $m_\xi \leftarrow 0 \triangleright$ *Number of intracluster edges*
2   $p_\xi \leftarrow 0 \triangleright$ *Number of intracluster pairs of vertices*
3   $m \leftarrow |E[G]| \triangleright$ *Number of edges*
4   $p \leftarrow \binom{|V[G]|}{2} \triangleright$ *Number of pairs of vertices*
5   **for** each g in CONNECTED-COMPONENTS-SUBGRAPHS(D, G)
6   **do** $m_\xi \leftarrow m_\xi + |E[g]|$
7           $p_\xi \leftarrow p_\xi + \binom{|V[g]|}{2}$
8   **return** $-\log_{10}\left(\sum_{i=m_\zeta}^{m} \frac{\binom{p_\xi}{i}\binom{p-p_\xi}{m-i}}{\binom{p}{m}}\right)$

Algorithm 2: Pseudocode of FAGSO, with description of the implementation of union-find data structure.

3.6 SURPRISE OPTIMIZATION ON FUNCTIONAL CONNECTIV- ITY NETWORKS

I assessed the performance of Surprise maximization through of FAGSO in the detection of the community structure of two benchmark brain networks. All coordinate data and functional metadata were taken from the BrainMap database [10, 195], processed by Crossley et al. [34] and made available to the scientific community as reference networks through the public Brain Connectivity Toolbox [110]. I used BrainNetViewer as a tool for the visualization of the communities on brain templates [196].

3.6.1 Human resting state dataset

The first network represents the coactivation of brain regions as obtained from a meta-analysis of 1641 task-related fMRI or PET studies [34]. Meta-analyses have been useful in estimating the frequency with which two brain regions are consistently activated across different tasks and are an indication of the behavior of the brain during activity. Jaccard similarity, i.e. the number of studies activating both regions divided by the number of studies activating either one of them, was used as an index to evaluate the strength of the coactivation of 638 parcellated brain regions. More details on the construction of the network are available in [34].

The second network that I considered is a resting state functional connectivity network obtained from correlations between time series of fMRI signals, from a group study of 27 healthy subjects. In short, fMRI data were acquired from 27 healthy volunteers at 3T. Gradient echo-planar imaging data were collected for 5 min with 2s TR and 13 and 31 ms echo-times. Thirty six interleaved 3mm slices with in-plane resolution of 3.5×3.5 mm were acquired. Time series were extracted from 638 brain regions defined by a template [34], corrected for motion and band-passed (0.01–0.1Hz). Functional connectivity was defined in terms of pairwise Pearson correlations at a subject's level. A group-level functional connectivity matrix was calculated by averaging individuals' matrices after Fisher-transform[2], and thresholded to retain 18625 edges, as described in Crossley et al. [34]. Ethical statements are present in the original references by the groups who performed the experiments. The resting state network was built using the same set of 638 regions and thresholded to have the same number of edges as in the coactivation study. Both networks have been previously studied using Modularity-based algorithms and node-classification methods [34].

Due to its definition in terms of binomial coefficients, Surprise can be computed for integer values of its parameters. I have therefore binarized the two adjacency matrices retaining an equal number of edges for both networks. While the binarization process discards information contained in the edge weights, a judicious choice of the threshold can ensure robust decomposition of the network [59, 197]. I have checked this statement by percolation analysis, a natural

---

2 A mathematical transformation needed to correctly compute averages of correlation coefficients. Fisher transformation of a correlation coefficient $-1 \leqslant r \leqslant 1$ is computed as $\operatorname{arctanh}(r)$ and ranges from $-\infty$ to $+\infty$.

and non-arbitrary method to derive binary graphs from continuous adjacency matrices (see section 1.6). Specifically, I have studied the size of the largest connected component of the coactivation and resting state networks iteratively removing the smallest weight edges.

This analysis, shown in Figure 3.8, revealed the presence of percolation-like transitions, whereby the largest component of the network drops in jumps with increasing binarization threshold. For the coactivation and the resting state networks I found that the thresholds adopted by [34] of 0.015576 and 0.600, respectively, are above the first jump in the size of the largest connected components and maintain network connectedness while ensuring that the networks are sufficiently sparse and possess the same number of edges. Hence, I adopted these thresholds for network's binarization. Analysis of the structures of networks obtained by a range of thresholds around these values showed stable solutions, with Normalized Mutual Information between partitions close to 1, and a stable number of communities (Figures 3.9 and 3.10).
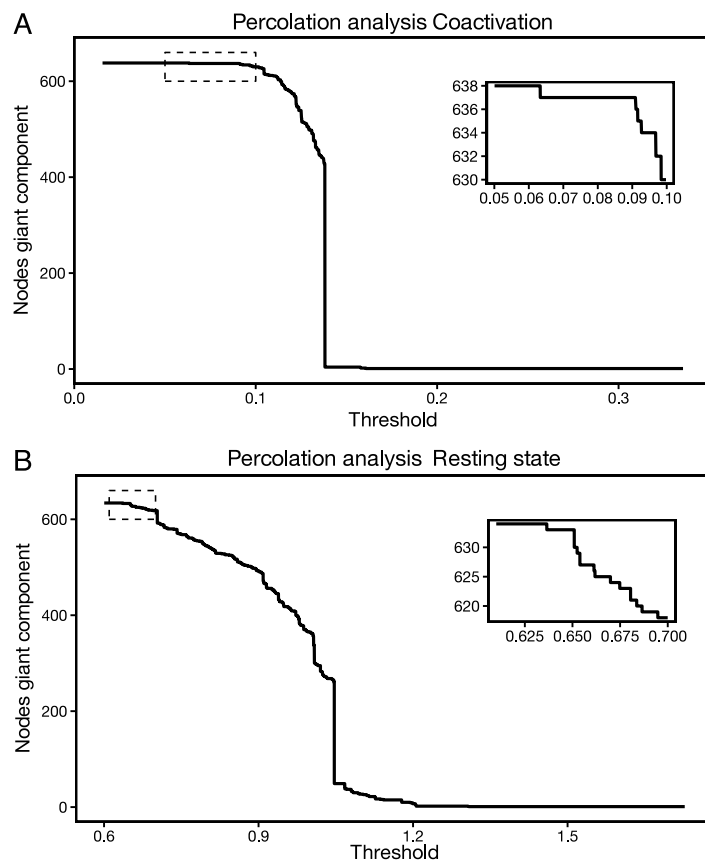
Figure 3.8: Percolation analysis for the coactivation matrix (A) and resting state matrix (B).

3.6.2 Community detection results by FAGSO

Figure 3.11 shows a direct comparison of the partitions obtained by Modularity and by Surprise maximization for the coactivation and resting state networks. The
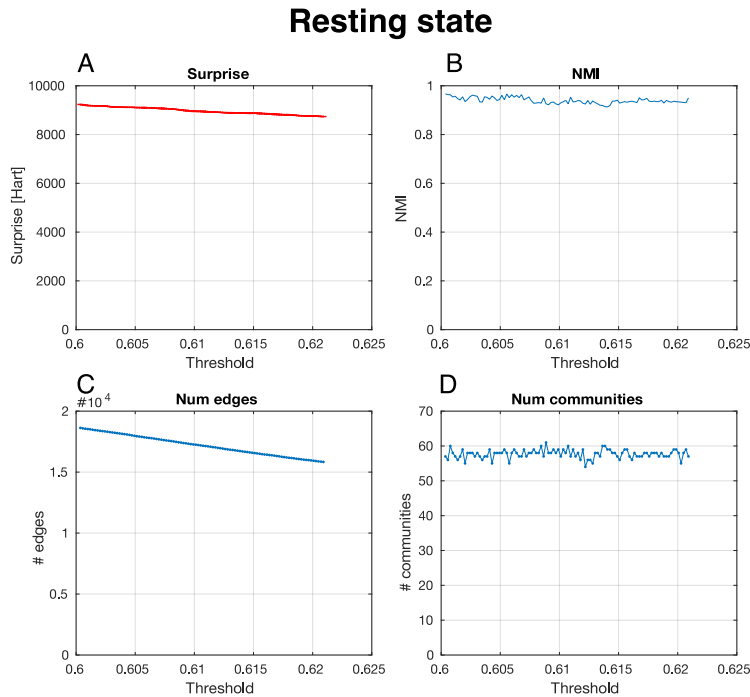
Figure 3.9: Different resting state networks were generated varying the binarization threshold in a range that changed the number of edges up to the 15th quantile of the edge weight distribution. This corresponds to threshold ranges of 0.015-0.03 for the co-activation network, and 0.60-0.62 for the resting state network. The resulting networks were partitioned by Surprise maximization. The value of Surprise, the Normalized Mutual Information between the resulting optimal partitions, the number of edges and the number of communities are reported in panel A, B, C and D, respectively.

four panels display the adjacency matrices of the two networks, with their vertices rearranged by their module membership.

By Newman's decomposition, the resting state and coactivation brain networks present a modular structure with four large modules that have been anatomically labeled as occipital, central, frontoparietal and Default Mode networks [34] (demarcated by a red line in Figure 3.11). These partitions are highly similar (Rand Index 0.78), despite the different neurofunctional bases of the two networks [198] and comprise modules that are relatively uniform in terms of number of nodes and number of edges within each module.

The partitions obtained by Surprise maximization for the two networks are shown in Figure 3.11B, 3.11D. Surprise found 51 communities, $\hat{S} = 8969.24$, for the resting state network, and 28 communities, $\hat{S} = 5725.65$ for the coactivation network. These modules are delimited by blue lines that show the wide distribution in size of the components, ranging from communities with 119 nodes and 4586 edges down to singletons. The size distributions of the modules are different for the two networks, with a more rapid drop and a fatter tail in the coactivation network compared with the resting state network.

The complete list of communities, with anatomical labels and stereotaxic coordinates for all nodes [199, 200, 201], as well as the density and number of nodes of

Figure 3.10: Different coactivation networks were generated varying the binarization threshold in a range that changed the number of edges up to the 15th quantile of the edge weight distribution. This corresponds to threshold ranges of 0.015-0.03 for the co-activation network, and 0.60-62 for the resting state network. The resulting networks were partitioned by Surprise maximization. The value of Surprise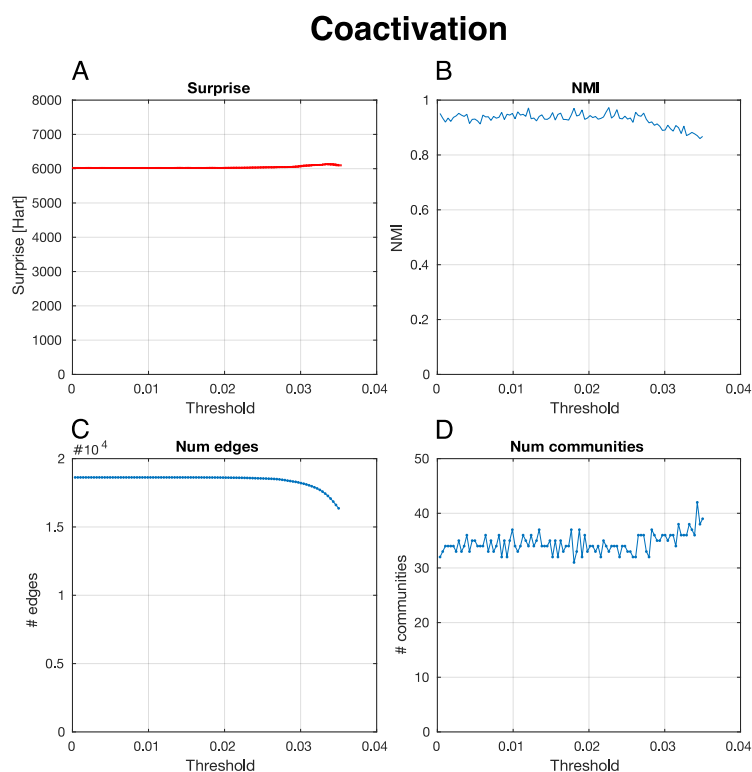, the Normalized Mutual Information between the resulting optimal partitions, the number of edges and the number of communities are reported in panel A, B, C and D, respectively.

each community found by Modularity and Surprise optimization, are reported in tabular form in the appendix (Tables A.2, A.3, A.4, A.5).

Analysis by Surprise suggests that the modular structure of resting state functional connectivity brain networks comprises modules of very different sizes, in sharp contrast with previous studies that have used resolution-limited functions like Newman's Modularity (see [59] for a review). To emphasize this point, I have also partitioned the coactivation and resting state networks using Infomap [82] and a multiscale version of Modularity with an adjustable resolution parameter [165] provided by the Brain Connectivity Toolbox [110]. Interestingly, increasing the resolution parameter results in a larger number of smaller communities that are however characterized by a relatively homogeneous size distribution, a result of the intrinsic scale built into these methods (Figure 3.14). Additionally, I have made a quantitative comparison between the partitions obtained by Surprise, Infomap and the Reichardt and Bornholdt's method [165] by calculating the Normalized Mutual Information between the resulting community structures (Tables 3.2 and 3.3). Despite the fact that these methods retrieve a few more modules that Newman's Modularity, they fail to capture the heterogeneous distribution of clusters revealed by Surprise.

Figure 3.11: Modular structure of the coactivation and resting state networks under Modularity and Surprise maximization. The node indexes have been reordered by membership to highlight the modules, which are demarcated by a red line, for Modularity, or a blue line, for Surprise. Modularity maximization identifies only four, large modules, consistent with previous analysis of these data-sets. Surprise reveals a much finer and complex modular structure.

In order to assess the significance in neurofunctional terms of the finer partitions obtained by Surprise, I show the node distribution as an overlay of the MNI brain atlas template for the 10 largest modules of the resting state network in Figure 3.12. The communities highlighted by Surprise show a correspondence with some well-known functional networks previously identified by multivariate analysis (e.g. Independent Component Analysis) of functional MRI data [15, 85, 139, 84], and with well-defined, segregated anatomical or functional districts.

The largest communities of the resting state network correspond to the primary sensorimotor cortex [9], primary visual and extra-striate visual network, fronto-parietal lateralized networks [198] as well as the so-called default mode network

Coactivation

NMI	Modularity	Surprise	Infomap	RB $\gamma = 1.75$
Modularity	**1.00**	0.61	0.56	0.61
Surprise	0.61	**1.00**	0.54	0.62
Infomap	0.56	0.54	**1.00**	0.59
RB $\gamma = 1.75$	0.59	0.61	0.59	**1.00**

Table 3.2: Normalized Mutual Information between partitions of the coactivation network with Newman's Modularity, Infomap, Reichardt and Bornholdt and Surprise. We used the Fagso algorithm for Surprise maximization, the `igraph` implementation of Infomap, the Brain Connectivity Toolbox implementation for RB (`community_louvain.m` function) and `modularity_und.m` for Modularity. For each method the best solution over 100 repetitions was used to calculate NMI.

Resting State

NMI	Modularity	Surprise	Infomap	RB $\gamma = 1.75$
Modularity	**1.00**	0.52	0.70	0.66
Surprise	0.52	**1.00**	0.53	0.58
Infomap	0.70	0.53	**1.00**	0.68
RB $\gamma = 1.75$	0.66	0.58	0.68	**1.00**

Table 3.3: Normalized Mutual Information between partitions of the Resting state network with Newman's Modularity, Infomap, Reichardt and Bornholdt and Surprise. We used the Fagso algorithm for Surprise maximization, the `igraph` implementation of Infomap, the Brain Connectivity Toolbox implementation for RB (`community_louvain.m` function) and `modularity_und.m` for Modularity. For each method the best solution over 100 repetitions was used to calculate NMI.
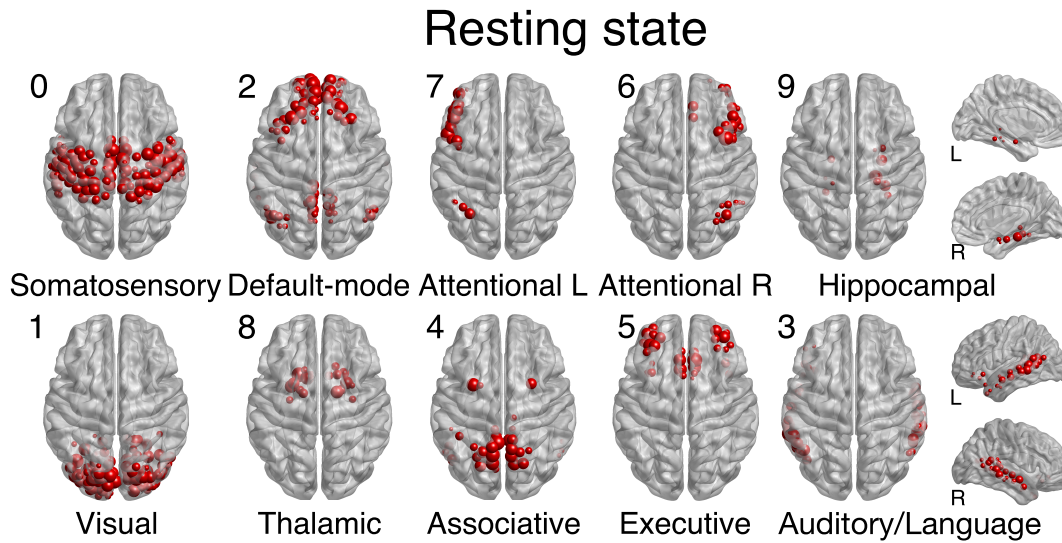
Resting state

Figure 3.12: The ten largest modules found by Surprise in the resting state network overlaid on an MRI brain template. The module indexes are ordered by decreasing size. The modules are named after corresponding functional networks previously identified by multivariate analysis of resting state fMRI data.

(DMN) [15, 17]. The attentional frontoparietal networks (FPAN) [138] were detected as two separate, lateralized subnetworks, in agreement with [84] although other studies have identified a single, bilateral FPAN [202].

Smaller networks, like the executive control and auditory networks [85, 30] were also resolved by Surprise, as well as subcortical structures, like the hippocampal and thalamic formations [203, 204]. Interestingly, the thalamic nuclei appear as one tight community, despite the fact that they are structurally unconnected, in keeping with the idea that functional connectivity does not necessarily require the presence of strong structural links.

The more accurate partition afforded by Surprise may enable identification of differences in the modular structures of networks that cannot be appreciated with a resolution limited method. By way of example, I have compared the partitions of the resting state and coactivation networks (Figure 3.13). Indeed, these networks are of a different nature, the former representing intrasubject baseline fluctuations in the brain's resting state, and the latter the responses to a variety of different tasks across subjects. However, Newman's Modularity finds similar partitions for these two networks, with 4 large modules each.

Conversely, under Surprise maximization, the partition of the resting state network shows many more small communities comprising less than 5 nodes (32 in total) compared with the coactivation one (only 11). Moreover, certain communities of the resting state network appeared to be split into smaller modules in the coactivation matrix. By way of example, the cuneus and the lingual and pericalcarine gyri were part of the occipital visual module in the resting state, but not in the coactivation network, where they formed a separate community (first row of Figure 3.13). Similarly, the precuneus and medial parts of the postcentral gyrus were identified as an independent community in the coactivation network, while they were part of the broad somatosensory network in the resting state connectiv-
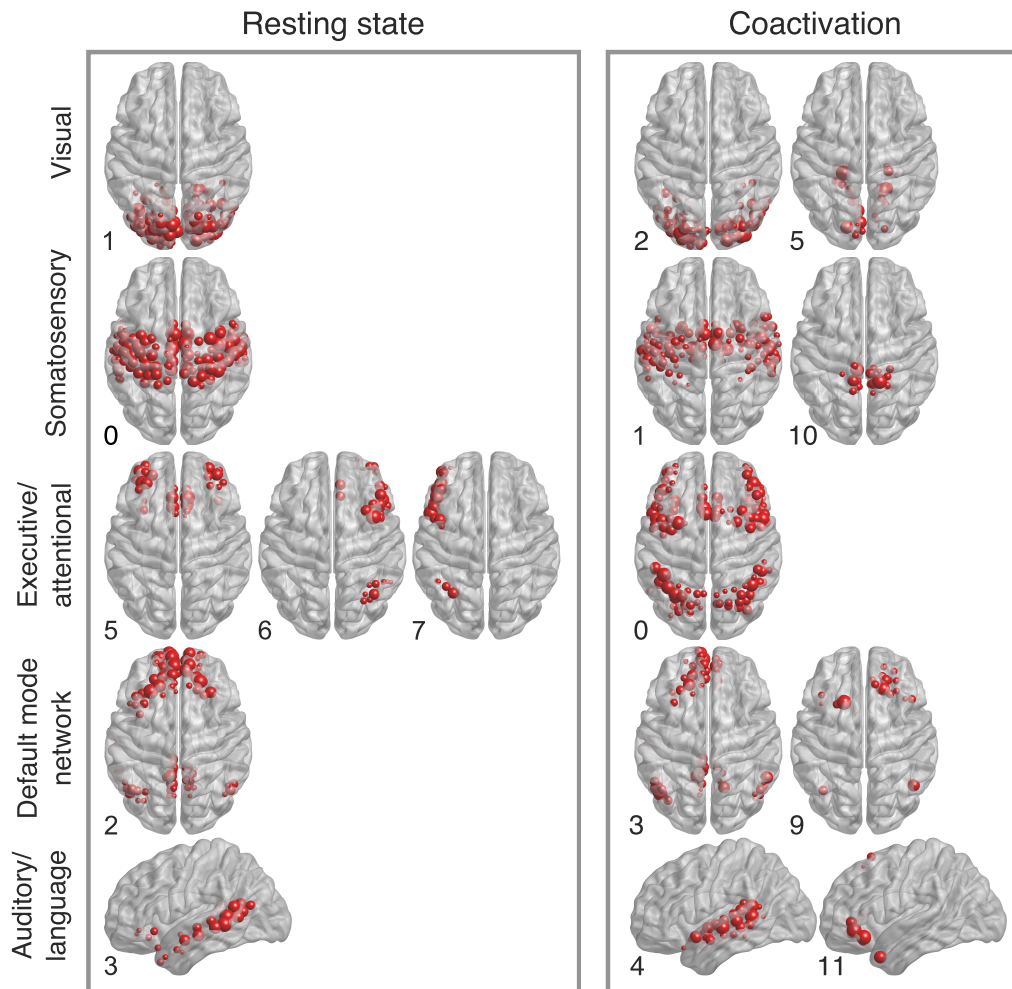
Figure 3.13: Comparison of selected modules in the partition obtained by Surprise in the resting state and coactivation networks. The indexes are inversely ranked according to the size of the modules in their respective networks.

ity graph [205] (second row of Figure 3.13). Interestingly, the Broca area, indicated as Module 11 in Figure 3.13, was separated from the auditory network in the coactivation network, and identified as a small, but anatomically and functionally distinct, community. Conversely, other communities were split in the resting state but not in the coactivation network. The executive and attentional control networks were merged into a large community in the coactivation network, while they were separated under resting-state conditions, including a subdivision of the left and right fronto-parietal networks (third row of Figure 3.13).

While the resting state and coactivation networks appeared to possess virtually identical modular structures under Newman's analysis, they showed functionally and anatomically relevant differences when analyzed by Surprise maximization, with a Normalized Mutual Information between the partitions of the two networks of 0.5922. Indeed, Modularity tends to assign small communities to larger structures even when they correspond to tightly-knit modules, thus concealing differences in the graphs' modular structures that involve aggregation or disag-

Resting State   Coactivation
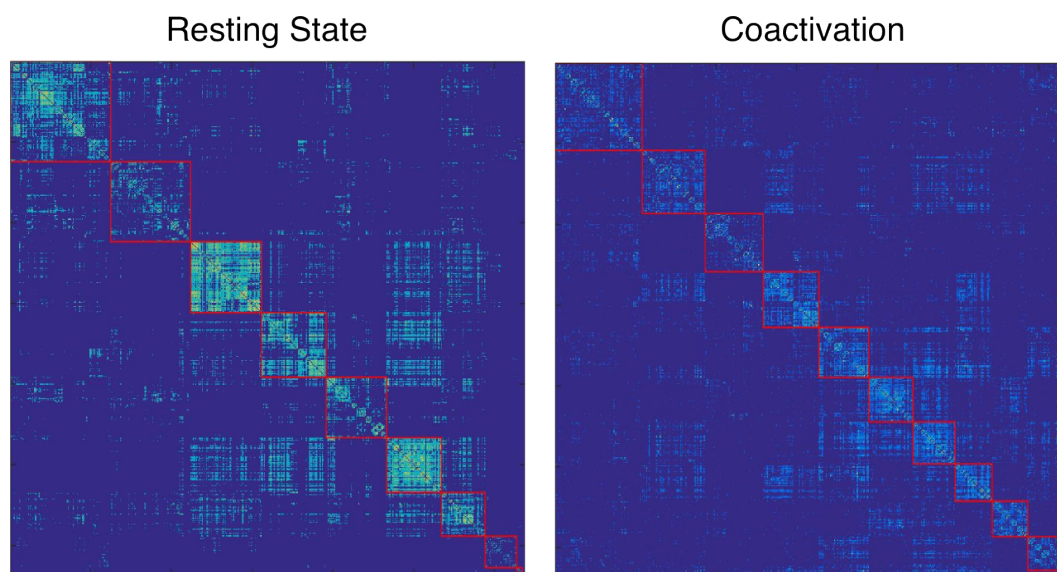
Figure 3.14: Partitions obtained with multiresolution Modularity (Reichardt and Bornholdt method) for a value of the resolution parameter $\gamma = 1.75$ for the coactivation and resting state networks. Increasing the resolution parameter improves detection of smaller modules, but breaks up larger ones, thus resulting in relatively homogeneous size distributions.

gregation of smaller clusters. It is conceivable that the detrimental effects arising from the resolution limit may have affected previous studies comparing different populations [59]. Surprise may offer a sharper tool to detect alterations of brain connectivity induced, for example, by psychiatric or neurological conditions, thus enabling the exploration of novel markers of brain disease.

3.6.3 Hubs classification

Besides the exploration of functional and anatomical segregation, understanding the modular structure of brain networks is critical for the interpretation and classification of the roles played by the nodes within the network structure [47]. Highly connected nodes, or hub nodes, are particularly important for their topological centrality, and function as integrators. Hubs that primarily connect to nodes within the same module are dubbed "provincial hubs", and nodes that connect different modules are referred to as "connector hubs". The former are thought to be responsible for the formation and stability of the modules, while the latter ensure integration of the activity of the different modules. Obviously, interpretation of the hub's role relies on the correct identification of the optimal network partition, and may be strongly affected by the resolution limit.

Here, I have performed a hub's role analysis for the resting state and coactivation networks under Modularity and Surprise maximization. To this end, I have adopted Guimera' and Amaral classification scheme [89], whereby nodes are classified by their within-community degree z (a measure of how well-connected a node is to other nodes in the same community) and their participation coefficient

P, a parameter that is 0 for nodes with purely intra-module connections and 1 for nodes whose links project primarily to other modules.

Figures 3.15A and 3.15B show the different positioning of high-degree nodes in the Guimera' and Amaral plot for the coactivation and resting state networks partitioned using Newman's approach and Surprise maximization. In this scheme, provincial hubs are high-degree nodes that score high z and low P values ($R5$ region); conversely, connector hubs are characterized by larger P values (regions on the right-hand side of the plot).
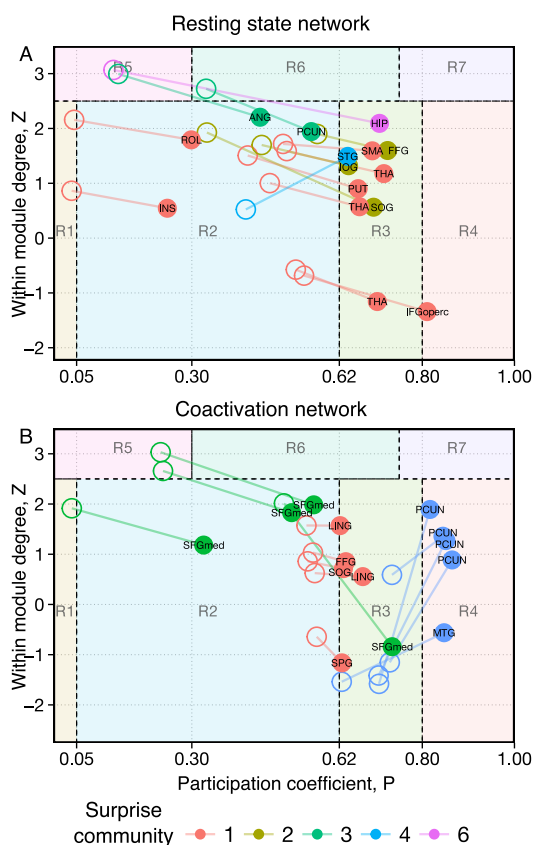
Figure 3.15: Classification of representative nodes according to their intra- and intermodule connections for the resting state (A) and coactivation (B) networks. Empty circles and full circles indicate the position of each node in the Guimera' and Amaral's plot after partition by Modularity or Surprise, respectively. An overall increase in the participation coefficient, a measure of the intermodular connectivity, is observed for the Surprise partition relative to the Modularity partition. To avoid cluttering of the graph, I only reported those nodes with a degree higher than the average within a Standard Deviation, and whose classification is different in the two partitions. The abbreviations of the brain regions corresponding to the nodes are reported in the appendix, Table A.1.

A finer partition in smaller communities may be expected to determine an overall increase in participation coefficient and a decrease in within-community degree. However, the heterogeneous partitions obtained by Surprise maximization resulted in non-trivial changes in the Guimera' and Amaral classification. By way of example, I discuss in greater detail two regions whose roles are very different in the two partitions, to highlight the effects of the resolution limit.

Nodes that belong in the hippocampal formation show a large within-module coefficient, and appear as provincial hubs (region R5 of Figure 3.15) under Modularity optimization. However, their participation coefficient increases 6-fold in the partition by Surprise, which reveals a role as connectors for these nodes, with widespread projections to many other modules across the brain, including the module 3, 6, 8, 5, 2, corresponding to the DMN, the amygdala and parahippocampal formation, the temporal inferior gyrus, the cuneus and lingual gyrus, and the visual cortex, respectively. This finding is in agreement with the idea that the hippocampus acts as "network convergence zone", as it receives polysensory input from distributed association areas in the neocortex [206].

Interestingly, the right shift in the Guimera-Amaral plot is less pronounced for the nodes in the posterior part of the hippocampus (Figure 3.16). A differential classification of the anterior and posterior hippocampus is consistent with the hypothesis of a functional differentiation of this structure [207], with the posterior hippocampus mostly involved in memory and cognition, and the anterior hippocampus playing a role in the processing of stress, emotion and affect [208]. Moreover, studies in animal models have shown differential organization of the efferent connections of the hippocampal formation [209], consistent with different functions for the anterior and posterior hippocampus.
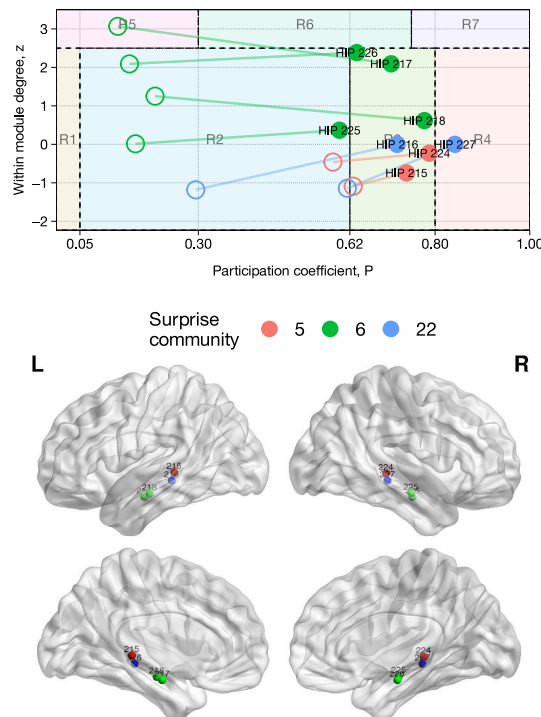
Figure 3.16: Top panel: classification of all the hippocampal nodes according to the Guimera' and Amaral's scheme for the coactivation network. Empty circles and full circles indicate the position of each node after partition by Modularity or Surprise, respectively. Bottom panel: anatomical positions of the nodes in the hippocampal formation, colored by Surprise community membership. The increase in participation coefficient upon partition by Surprise is more pronounced for nodes in the anterior part of hippocampus, with an antero-posterior gradient.

Similar rightward shifts for nodes and hubs were observed in the resting state network, and reported in Figure 3.15. However, increases in participation coefficients are by no means the only differences in the classification of nodes obtained by Surprise maximization. A prominent example is the precuneus (PC) (Figure 3.15B, blue dots) that shows a high participation coefficient in both partitions by Modularity and Surprise, but a much higher within-module degree under Surprise maximization.

Indeed, the nodes comprised in the precuneus intrinsically possess high inter-cluster connectivity, but are distributed among the four modules found by Modularity. In the partition by Surprise they are grouped together, and this precuneal module as a whole plays a connector's role integrating different communities (Figure 3.17), a hypothesis that is consistent with the precuneus supporting a wide spectrum of highly integrated tasks, from visuospatial imagery to episodic memory retrieval and self-processing operations [210].

In summary, partition by Surprise maximization results in very different distributions of participation coefficients and within-module degree compared to Modularity. These differences are not uniform across nodes, and arise from the limited ability of Modularity to identify small modules. Finer partition by Surprise reveals very different roles for some key brain areas, and suggests that a systematic reanalysis of the topological roles of brain nodes and hubs may be in order.
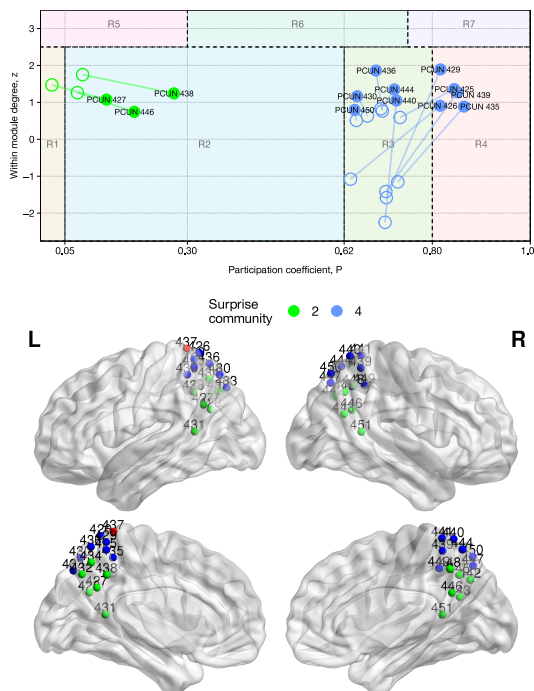
Figure 3.17: Top panel: classification of the precuneal nodes according to the Guimerá' and Amaral's scheme for the resting state network. Empty circles and full circles indicate the position of each node after partition by Modularity or Surprise, respectively. Bottom panel: anatomical positions of the nodes in the precuneus, colored by Surprise community membership. The nodes in the dorsal part of the precuneus exhibit a sharp increase in within-module degree.

3.7 LIMITATIONS OF BINARY SURPRISE

A potential limitation of Surprise is related to its definition in terms of discrete probability distributions. This makes Surprise suited for the study of binarized networks. While the topological backbones of the networks I have investigated appear quite robust against removal of lowest-weight edges and binarization, as shown by our percolation and stability analyses (Figures 3.8, 3.9, 3.10), the application of its weighted counterpart, Asymptotical Surprise, will help overcome the need of binarization.

The superior resolution afforded by Surprise may make it more vulnerable than other methods to noise and experimental errors. Indeed, occasional misassignment of nodes due to noise-induced changes in edge distribution are likely to affect small modules, comprising only a few nodes, more than large ones. Hence, experimental uncertainty also limits resolution, and a resolution-limit-free method would not necessarily improve the quality of the partition in a scenario dominated by noise.

To ascertain whether this is the case for the co-activation and resting state networks investigated here, I have simulated the effects of experimental errors by injecting noise into the distributions of weights prior to the binarization procedure, thus introducing variability in the connectivity structure of the resulting binary networks. I set levels of noise sufficient to perturb up to 10% of the edges of the final binary network. Using this procedure, for each level of noise I generated ten different graphs, and applied the Surprise Maximization algorithm to each of them.

The ten different graphs were generated by adding noise to the off-diagonal elements of the adjacency matrix prior to the binarization procedure. The amplitude of noise was chosen to randomly perturb 10% of the edges after binarization. Surprise maximization was applied to the resulting graphs. In the left panel of Figure 3.18, I reported the Normalized Mutual Information (NMI) between the optimal partitions of each of the ten perturbed networks and that of the original resting state network. The number of communities retrieved by FAGSO in each of the perturbed networks is reported in the right panel of Figure 3.18. The graphs show that the partitions of the "noisy" graphs are consistent and similar to those of the unperturbed network, with NMI scores close to 1 and almost constant numbers of communities. These results demonstrate that Surprise maximization can retrieve the network's community structure even in the presence of substantial noise in the data. I found that the partitions of these graphs were highly consistent with those of the original networks (Figure 3.18).

We should also stress that there is no constraint in the FAGSO algorithm imposing inter-hemispherical symmetry of the partition. Nevertheless, I observed homotopic correspondence in the community structure, and a close resemblance with established neurofunctional circuits as in Figure 3.12 and 3.13. Taken together, simulations of the effects of noise and qualitative considerations on the neurofunctional significance of the modules identified by Surprise corroborate the idea that experimental error is not the predominant factor in the networks investigated.

A final and important point I should highlight is that Surprise maximization, in the implementation I have used here, does not allow for overlapping commu-
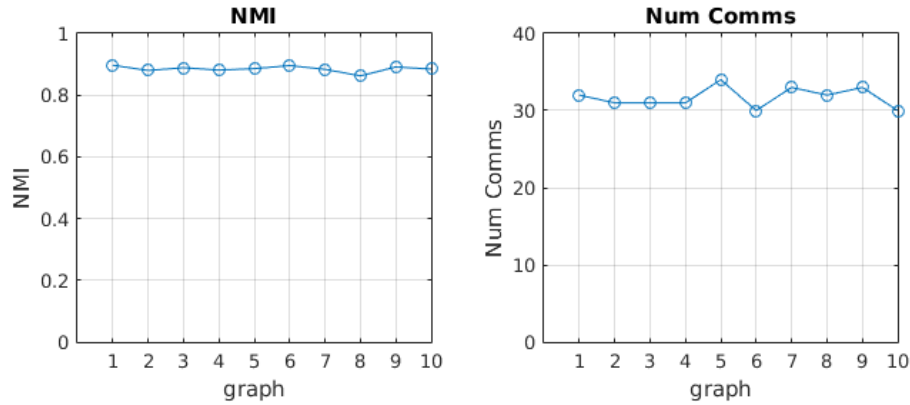
Figure 3.18: Assessment of the effects of experimental error on the community structure of the resting state network.

nities. Other methods have been applied to investigate this aspect in brain networks [211, 212]. However, a recent comparative analysis of graph partitioning algorithms on a variety of benchmark networks [69] has shown that these methods are also prone to merging overlapping communities, with relatively modest performance in recovering heterogeneous cluster distributions planted in model networks. Despite these potential limitations, the resolution-limit-free behavior of Surprise makes it an excellent tool to explore and to overcome the effects of the resolution limit in the modular structure of brain connectivity networks.

3.8 CONCLUSIONS

In conclusion, I have shown that Surprise, a recently proposed fitness function for graph partitioning, behaves like a resolution-limit-free function. I have applied Surprise maximization to study the modular structures of two different brain networks. Surprise maximization resulted in partitions comprising communities of widely distributed sizes, consistent with the idea that small and large modules coexist in brain networks. Moreover, the finer partition afforded by Surprise made it possible to appreciate differences in the modular structures of diverse brain networks that were undetected by resolution limited methods like Newman's Modularity. Finally, the use of Surprise revealed the deleterious effects of the resolution limit in the classification of nodal roles. Altogether, these results indicate that the resolution limit may have substantially affected many of the analyses of brain connectivity networks reported in the literature, and call for a revisitation of some of the conclusions and models that have relied on Modularity maximization or similarly resolution-limited algorithms. Surprise appears as a promising alternative method that appeals to the intuition that tightly-knit clusters of nodes represent legitimate structural or functional modules independently of their size.

4 | BEYOND THE RESOLUTION LIMIT IN WEIGHTED NETWORKS

Brain networks are intrinsically weighted with weights reflecting a continuous distribution of connectivity strengths between different regions. A fundamental limitation of Surprise lies in its definition in terms of discrete probability and binomial coefficients that make it applicable only to binary networks, i.e. graphs with edge values 1 or 0. For this reason, Surprise requires binarization of brain connectivity networks, a process that may discard potentially relevant information contained in the edge weights distribution. Moreover, different binarization procedures may lead to different network representations for the same connectivity dataset. This represents a substantial drawback of the binary Surprise approach.

The results obtained from binary Surprise optimization on real functional connectivity networks encouraged me to extend the Surprise optimization approach to weighted networks. Capitalizing on recent developments in the field of statistical physics of complex networks, here I show that binary Surprise can be extended to weighted networks. This recently introduced approach dubbed Asymptotical Surprise provides a new and valuable tool to study the modular organization of brain connectivity beyond the resolution limit. Furthermore, I will present a new algorithm for the direct optimization of Asymptotical Surprise that is based on the previously described FAGSO method.

Improved resolution afforded by Asymptotical Surprise may imply increased vulnerability to spurious modules resulting from noisy correlations. It is therefore important to assess the benefits of increased resolution against the limitations arising from intrinsic data variability. To this end, I will demonstrate this new approach in networks derived from synthetic datasets that mimic different structures, levels of noise and variability, like those observed in functional connectivity experimental data.

Finally, I will apply Asymptotical Surprise optimization to weighted functional connectivity networks from real-world resting state fMRI data, revealing a heterogeneous, multiscale community structure. The finer modular subdivision of resting state functional connectivity networks obtained by Asymptotical Surprise optimization leads to substantial differences in the identification of connector hubs compared to other community detection methods.

4.1 FROM SURPRISE TO ASYMPTOTICAL SURPRISE

In developing a quality function inspired by the principle of Surprise that works for weighted networks, it is convenient to consider the asymptotical expansion of the hypergeometric distribution. Hereafter I will stick to the derivation obtained by Traag [176]. One can introduce $q = m_\zeta/m$ and $\langle q \rangle = p_\zeta/p$ as the observed and expected fraction of intracluster edges. By only taking into account the dominant

term of the sum in Eq. 3.3 (the one with $i = m_\zeta$), after some manipulations one gets an approximate expression for the logarithmic Surprise[1]:

$$\log(S) \approx \log\left(\frac{\binom{\langle q \rangle p}{m_\zeta}\binom{(1-\langle q \rangle)p}{m(1-q)}}{\binom{p}{m}}\right) \tag{4.1}$$

which corresponds to the probability of observing exactly m_ζ internal links, given the clustering ζ. As the denominator in Eq. 4.1 is independent of the partition, it can be discarded, and, thanks to the Stirling approximation of the binomial coefficients, which reads

$$\log\binom{n}{k} \approx k \log\left(\frac{n}{k}\right) \tag{4.2}$$

one can write the dominant term 4.1 as:

$$\log(S) = -\log\left(\frac{m}{p}\right)^{-m}\left[\left(\frac{\langle q \rangle}{q}\right)^q \left(\frac{1-\langle q \rangle}{1-q}\right)^{1-q}\right]^m \tag{4.3}$$

The term $(m/p)^{-m}$ is also independent of the partition and can be ignored. Hence, the asymptotic expansion of Surprise reads:

$$\log(S) = -m\left[q\log\frac{\langle q \rangle}{q} + (1-q)\log\frac{1-\langle q \rangle}{1-q}\right] \tag{4.4}$$

Interestingly, this last equation corresponds to the binary Kullback-Leibler divergence $mD_{KL}(q\|\langle q \rangle)$, which is interpretable as the distance between the two probability distributions q and $\langle q \rangle$. More precisely, it is the information lost when one encodes the distribution q with the distribution $\langle q \rangle$. Thus, in the limit of large networks, Surprise $\hat{S}$ can be approximated by a binomial distribution. This observation led to definition of Asymptotical Surprise S_a [176]:

$$S_a = mD_{KL}(q\|\langle q \rangle) \tag{4.5}$$

where the binary Kullback-Leibler divergence [213, 104] is

$$D_{KL}(x\|y) = x\log\left(\frac{x}{y}\right) + (1-x)\log\left(\frac{1-x}{1-y}\right).$$

In the framework of information theory [104], Asymptotical Surprise represents the Kullback-Leibler (KL) divergence between the observed and expected fraction of intra-cluster edges; it encodes the information lost when the prior distribution $\langle q \rangle$ is used to approximate the posterior distribution q. Kullback-Leibler divergence is a quasi-distance on probability distributions as it is always non-negative, non-symmetric and zero only when $q = \langle q \rangle$, exactly like binary Surprise.

Asymptotical Surprise has a simpler formulation than binary Surprise as there are no binomial coefficients to evaluate and it has been shown to be resolution-limit-free in the limit of large networks [176]. As a side effect of its definition in terms of an information-theoretic quantity, Asymptotical Surprise transparently allows the extension to weighted networks, when the intracluster number of edges is

---

1 If not specified, starting from here we use natural base logarithms.

replaced by the sum of intracluster weights. This powerful property made Asymptotical Surprise suited for community detection in weighted networks and in particular to brain functional connectivity networks, thus avoiding the need of binarization procedures prior to the community detection, as shown in [97].

Given its information-theoretic formulation as a KL divergence, Asymptotical Surprise clearly features *convexity*, a property that is more difficult to assess for binary Surprise. Optimization of convex functions is typically simpler due to some regularities featured by such functions and the availability of practical algorithms. Furthermore, a convex function guarantees a landscape with a global optimum and no degeneracy, a property that is extremely important for community detection.

Numerical evaluation of Asymptotical Surprise is much faster than binary Surprise. Furthermore, it approximates very well Surprise already for networks with more than 50 nodes, as shown in Figure 4.1.
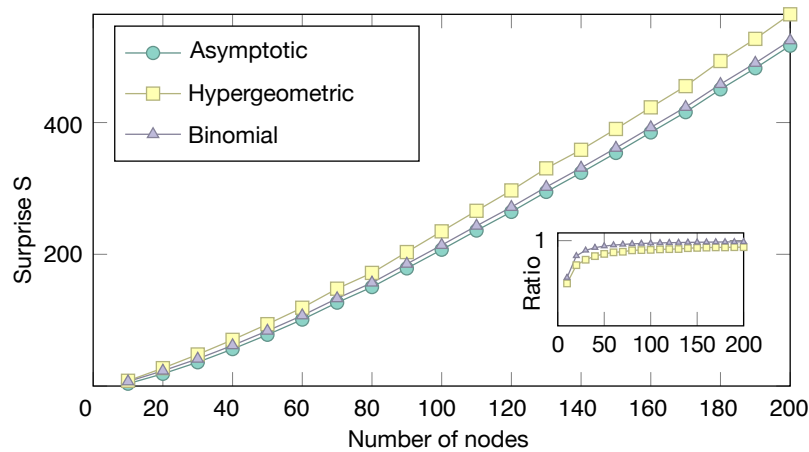
Figure 4.1: Approximation of binary Surprise with a binomial formulation and the Asymptotical formulation based on the Kullback-Leibler divergence. In the inset, the approximation ratio of binomial and Asymptotical Surprise to hypergeometric Surprise tends to 1 for graphs larger than 50 nodes. Adapted from [176].

4.2 MAXIMIZATION OF ASYMPTOTICAL SURPRISE: PACO

Finding the optimal partition of a graph is an NP-hard problem [54] and practical implementations of community detection rely on heuristic approaches that enable finding nearly optimal solutions in a reasonable computation time.

Here I introduce a powerful and general method for the optimization of Asymptotical Surprise dubbed PACO (PArtitioning Cost Optimization). PACO is a non-deterministic agglomerative algorithm based on FAGSO (described in chapter 3.5) and, like the Louvain method, has an element of randomness that enables a more efficient exploration of the partition landscape.

The operating principle of PACO is based on the triadic closure property, i.e. the fact that, in real-world networks, nodes with many common neighbors are more likely to be neighbors. This transitive neighborhood property underlies the formation of communities of nodes [214, 215]. In principle, any measure of structural

similarity between nodes could guide a community detection heuristic toward the optimal partition. Specifically, PACO uses the Jaccard index [216], a measure of the fraction of overlap between the neighbors in common between nodes, as the guiding principle for the agglomeration of similar nodes in the same community.

In the first phase of PACO, the Jaccard metric is evaluated for every edge. More formally, for an edge $e = (u, v)$ the Jaccard index is computed as $J(e) = \frac{|\Gamma(u) \cap \Gamma(v)|}{|\Gamma(u) \cup \Gamma(v)|}$ where $\Gamma(u)$ and $\Gamma(v)$ are the neighboring nodes of u and v respectively.

The agglomerative process starts with an initial partition where every vertex represents a community on its own. This partition has n communities and no intra-cluster edges. The edges of the graph are then ranked in decreasing order by their Jaccard index and iteratively, for every edge in the sorted list, endpoint nodes are merged only if they belong to different communities. In this case, one of the two endpoints, selected by chance, is assigned to the other's endpoint community and the increment of Surprise is computed: if it is positive, the partition is updated together with the new value of Surprise (or Asymptotical Surprise), otherwise the algorithm proceeds to the next edge.

Algorithm 3 describes the details of PACO for Surprise and Asymptotical Surprise Optimization. The function PACO takes as input a graph G and returns the nodes community membership vector C. Line 1 initializes the value of Surprise to 0. Line 2 assign to each node in the graph its community. Line 3 creates a list of edges E' sorted in decreasing order by their Jaccard coefficient. Line 4 iterates on every edge $e = (u, v)$ and at line 6 checks if the endpoints they share the same community. Line 8 copies the membership vector to a temporary vector C'. Lines 9-13 choose randomly at chance if to put node u in the community of v or viceversa. Line 14 computes the new value of Surprise S' from the just updated community membership C. The function COMPUTESURPRISE returns the value of Surprise for graph G and partition C. Lines 15 to 18 checks if the new value of Surprise S' is greater than the previously stored value S and update Surprise and the membership vector, otherwise continue to the next edge. Line 19 returns the final community membership assignment. The implementation of PACO in `C++`, `Python` and `GNU Octave` is available upon request `http://forms.iit.it/view.php?id=68447`.

4.2.1 Running time analysis of PACO

I applied PACO on a full-resolution voxelwise connectivity matrix with 51,653 nodes and almost 2 million edges. PACO took 14 minutes for a single repetition on a server with Intel Xeon E5-2643@ 3.40 Ghz CPU and 256 GB ram. I estimated that 2000 repetitions of PACO would take approximately 2.5 weeks on this server. I have also tried to run PACO on a standard office PC with 16 GB memory and an Intel Core i7: it took almost 40 minutes. It is important to notice that the running time of PACO is mainly determined by the computation of Jaccard indexes in the initial step. The running time for this computation is in the order of $O(n\langle k \rangle^2)$ where n is the number of nodes and $\langle k \rangle$ is the average degree of nodes. On a desktop workstation with a 2.5 GHz CPU, PACO runs in some tenths of seconds on a single repetition for a graph of around 600 nodes with a density close to 10%, typical of brain networks. A small benchmark of PACO running time on a desktop workstation is shown in Figure 4.2, where I found optimal Surprise partitions of a

PACO(G)

1   S ← 0 ▷ *Initialize Surprise to 0*
2   C ← (1,...,|V|) ▷ *Initialize membership vector*
3   E′ ← SORT-JACCARD(E) ▷ *Sort edges in decreasing order by Jaccard index*
4   **for** each edge (u,v) in E′
5   **do**
6   **if** $C[u] \neq C[v]$ ▷ *try to move nodes only if in different communities*
7   **then**
8   C′ ← C ▷ *Create a temporary membership vector*
9   **if** UNIFRAND(0,1) < 0.5
10   **then**
11                  $C'[v] \leftarrow C[u]$
12   **else**
13                  $C'[u] \leftarrow C[v]$
14   S′ = COMPUTESURPRISE(G,C′)
15   **do if** S′ > S
16   **then**
17   C ← C′ ▷ *update membership*
18   S′ ← S ▷ *update Surprise*
19 **return** C

Algorithm 3: Pseudocode of the PACO algorithm.

LFR network with the parameters described in the text but increasing number of nodes.
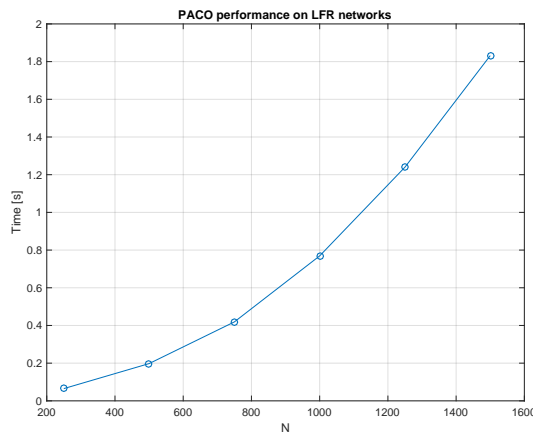
Figure 4.2: Performance of PACO on a LFR network with parameters

4.2.2   Reproducibility of PACO optimal solutions

To get a better idea of how fast the PACO method converges to a maximum, I have plotted in Figure 4.3 the optimal value of Asymptotical Surprise over the

10000 runs for one instance of the LFR network, and for the resting state data-set presented in the section 3.6.1. From these graphs, it appears that 2000 runs may be sufficient to reach stable community detection for the experimental data-set. A near-optimum value is reached earlier for the LFR network, but it should be noted that I have taken an instance of the synthetic network with no-noise added.
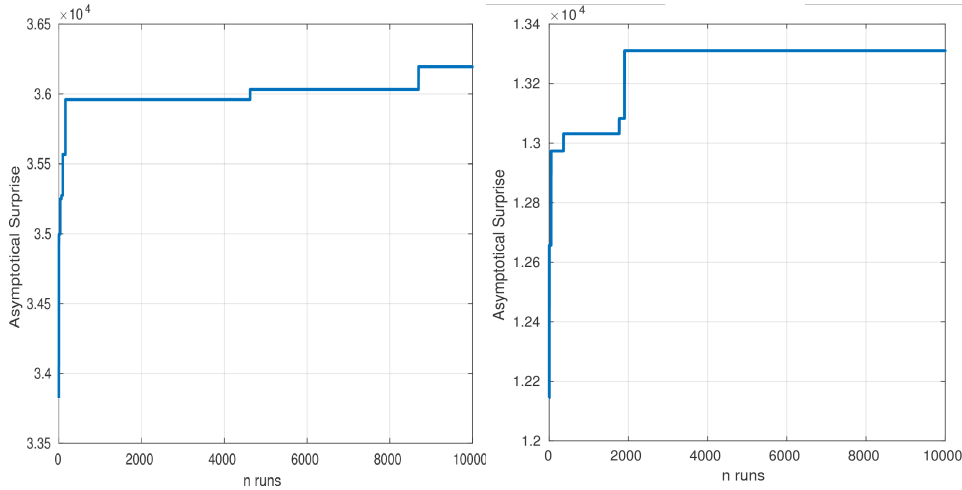
Figure 4.3: Maximum value of Asymptotical Surprise with respect to number of repetitions on a LFR networks (left panel) and on a resting state network as in [34].

4.2.3 Benchmarking PACO

Two of the most important parameters to shape the community structure of an LFR network are the topological and mixing coefficients (see section 1.3.5). The topological mixing coefficient μ_t is the average ratio of intra-cluster neighbors divided by the number of inter-cluster neighbors, as defined in [80]. The weights mixing coefficient is defined as the average ratio of node intra-cluster strength and inter-cluster strength, as defined in [95]. I explored the effects of these parameters on the ability of PACO to retrieve the planted structure in the network. I set $\mu_t = \mu_w$, as setting μ_w greater than μ_t would introduce inconsistency in the relative number and weight of the edges, with intermodule edges carrying the largest weights. I analyzed the performance of Newman's Modularity, Infomap and PACO on LFR networks when varying topological and weights mixing coefficients. Figure 4.4 shows that the performance of the three methods is comparable in terms of NMI, with a faster decay of NMI for InfoMap and Newman compared to PACO for large μ_t.
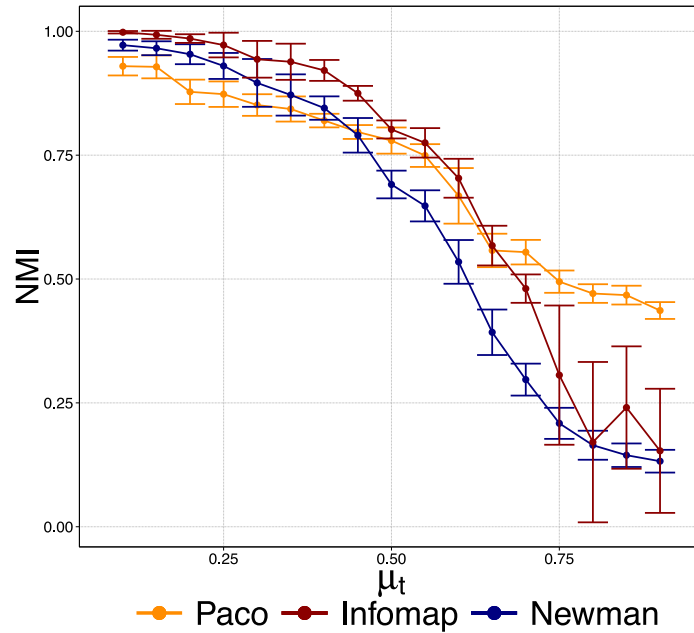
Figure 4.4: NMI of the retrieved vs planted partition of an LFR network as a function of $\mu_t = \mu_w$ for the three community detection methods.

4.3 DEGENERACY OF ASYMPTOTICAL SURPRISE

Degeneracy of nearly-optimal solutions, whereby similar values of the fitness function around its maximum correspond to substantially different partitions, has been observed for Newman's Modularity [184]. A consensus approach has been suggested in [187] as a means to mitigate the degeneracy problem, yielding a stable "average" solution over a large set of partitions. In order to ascertain whether Asymptotical Surprise suffers from a similar shortcoming I have performed degeneracy analysis by following [184], for Newman's Modularity. In created a benchmark network consisting of 24 cliques of 5 nodes, connected by a single link to form a connected ring-like structure. I sampled the configuration space of partitions during iterative steps of PACO optimization starting from different random solutions and annotating the corresponding values of Asymptotical Surprise for each partition. Afterward, I embedded the similarity matrix between all sampled partitions into a three-dimensional space maintaining similarity relations between partition with a Curvilinear Components Analysis (CCA) [184]. In the embedded manifold, two partitions are close if they are similar and the z-axis encodes the quality function. Whereas a broad plateau of solutions with similar values of maximum Modularity is observed as in Figure 2.5A, consistent with [184], Asymptotical Surprise displays a much sharper peak corresponding to the optimal solution, as shown in Figure 4.5. The existence of a globally maximum partition that is much different from the other suboptimal solutions indicates that there is no need to resort on consensus-based approaches for the identification of the best candidate solution, as picking the highest Asymptotical Surprise value already yields the correct partition.
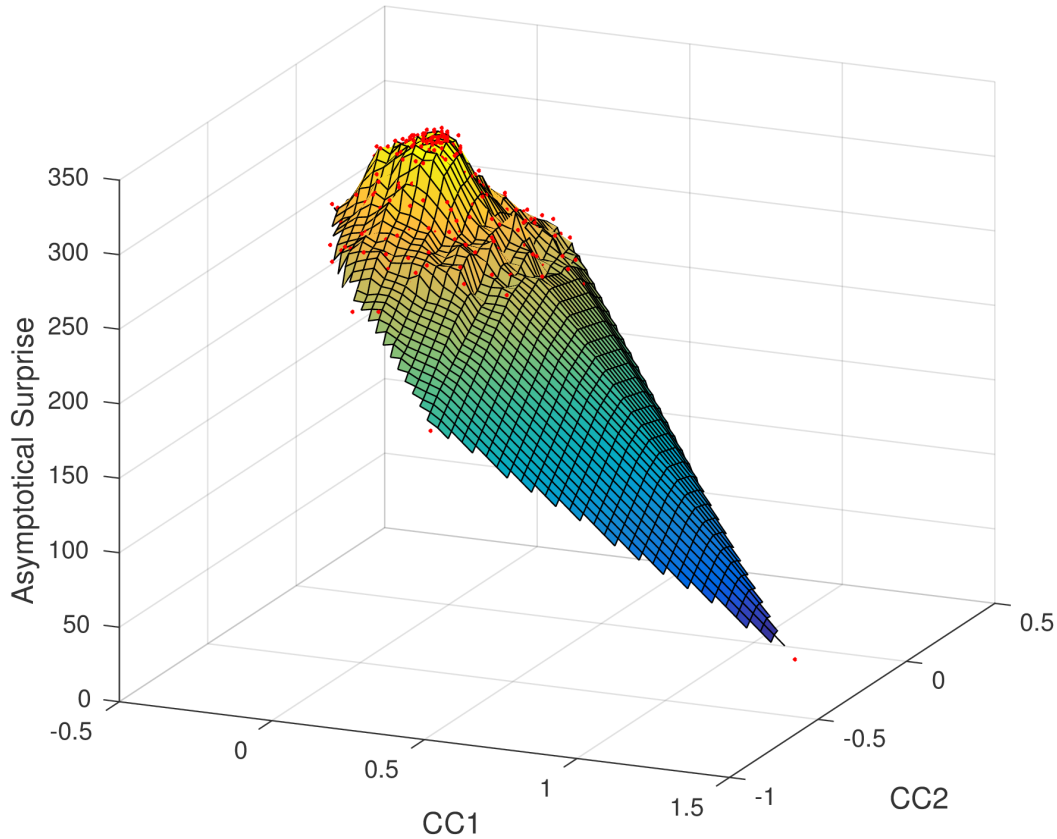
Figure 4.5: Degeneracy landscape of Asymptotical Surprise on a ring of clique model with 24 cliques of 5 nodes. The axes CC1 and CC2 are complicated functions of the original partition space computed to maintain the distance relation between points and their scale is unimportant. The height of this surface is the value of Asymptotical Surprise. Red points represent individual partitions and their metric distance is proportional to their NMI, the closer the points, the more similar the partitions.

4.4 SYNTHETIC BENCHMARK NETWORKS

One important finding in [182] is that brain networks are organized in modules with a heterogeneous size distribution. I implemented this property in our two types of benchmark networks. For the first test, I generated a ring of cliques with 300 nodes, and sizes of the cliques sampled from a power-law with exponent $\tau_c = 2$, minimum and maximum clique size respectively $\min_c = 5$, $\max_c = 50$. For each subject of the sample, I synthesized 150 time-points for each node using the neuRosim R package [217]. I set the baseline value of all the time series to 100 [218].

Finally, I correlated the original synthetic time series **X** by multiplication with the matrix **L**, obtained the correlated time series **Y** and added Rician noise [219] to **Y** independently for each area. The simulated data **Y** did not include slow drift components, simulated physiological noise, nor spatial noise. The average SNR is defined as $\text{SNR} = \bar{S}/\sigma_N$ where $\bar{S}$ is the average magnitude of the signal and σ_N is the standard deviation of the noise [220].

To be more exhaustive and extend the validity of results, I repeated the same procedure on weighted LFR networks with $N = 600$ nodes, sampling nodes degree from a power-law with exponent $\tau_d = 2$, average degree $\langle k \rangle = 12$ and maximum degree $\max_k = 50$. I set the topological and weights mixing coefficients, i.e. the average fraction of intra-cluster and inter-cluster degree and strengths, to $\mu_t = 0.1$ and $\mu_w = 0.1$, respectively. Planted community sizes ranged from 5 to 50 nodes and were sampled from a power law with exponent $\tau_c = 1$. Figure 4.6 shows two realizations of these models.
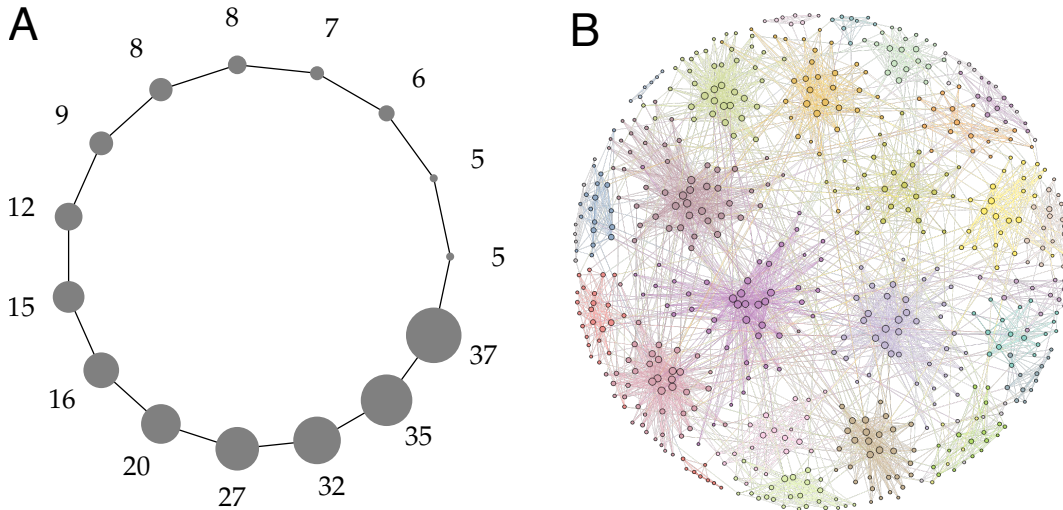
Figure 4.6: The two benchmark networks used in this study, laid out. (A) is a power-law ring of cliques, where cliques present different sizes sampled from a power-law distribution; (B) is the layout of an LFR network with parameters $N = 600$, $\langle k \rangle = 12$, $\max_k = 50$, $\mu_t = 0.1$, $\mu_w = 0.1$, $\min_c = 5$, $\max_c = 50$. The layout of (B) was generated with the graph-tool library [221].

Group-level correlation matrices were computed by Fisher-transforming and averaging individual instances of the above matrices. Sparsification was obtained by removing edges with weights below the most stringent threshold that maintained the network connectedness, a procedure known as percolation analysis [111, 116, 65]. This approach measures the size of the largest connected component of the network upon iterative removal of the weakest edges and enables data-driven determination of the optimal sparsification threshold that preserves network structure and connectedness while removing potentially spurious correlations.

The community structure of the resulting weighted sparsified matrices was detected by Asymptotical Surprise optimized with PACO and compared against two widely used methods, Infomap [82] and Newman's Modularity [171, 71], that are affected by the resolution limit, albeit to different extents. In Newman's Modularity, the size of the smallest detectable cluster is of the order of the square root of the number of edges in the entire network [72], while Infomap has a limit that depends on the overall number of inter-cluster edges [164]. Here I used the Louvain implementation available in the Brain Connectivity toolbox [110] and the Infomap implementation available in the `igraph-0.7.1` package [222].

For all methods, including PACO, I launched $10,000$ independent runs and picked the membership corresponding to the partition with the best value of the

fitness function, the maximum for Modularity and Asymptotical Surprise, the minimum for Infomap. As discussed in the previous chapter, sections 2.4, 3.4, 4.3, degeneracy of nearly-optimal solutions does not appear to severely affect Surprise or Asymptotical Surprise, and a consensus approach is not deemed necessary for these functions. This analysis supports the choice of selecting the solution with the highest value of the fitness function.

4.5 PACO BENCHMARK ON SYNTHETIC NETWORKS

I compared the quality of the partitions of the synthetic benchmark networks obtained by Asymptotical Surprise with those of Infomap [82] and Newman's Modularity [71, 171]. Figure 4.7A shows Normalized Mutual Information, Sensitivity and Specificity of the three methods applied to the ring of cliques for different sample sizes and SNRs; the no-noise condition is represented as "Inf". This model network was constructed to test the ability of the three methods to retrieve heterogeneous community structures under various noise conditions.

As expected, all methods showed better performance with increasing SNR and number of subjects, as noise and intersubject variability introduce spurious edges that hinder the ability to retrieve the planted structure. Partitions obtained with Newman's modularity showed the lowest NMI with respect to the planted partition under all conditions. Sensitivity of Newman's modularity did not exceed 0.75 even for high SNRs and a large number of subjects, a consequence of its stronger resolution limit. For this network, Infomap performed substantially better in terms of NMI against the planted partition, with a Sensitivity that was superior to that of Modularity across the spectrum of conditions.

Asymptotical Surprise showed highest NMI and Sensitivity across conditions, consistent with its resolution-limit-free behavior. It proved superior in terms of NMI and Sensitivity in the low SNR regimes, and in the presence of relatively large intersubject variability as mimicked by the generation of different instances of the ring of cliques. Specificity of Asymptotical Surprise was not inferior to the other methods under all conditions, thus ruling out increased vulnerability to False Positives, at least in this particular model network.

Comparable results were obtained for the LFR network (Figure 4.7B), a model graph that replicates the distribution of nodal degree observed in many real-world networks, including those representing functional brain connectivity. All three methods showed similar values of NMI for high SNR and a large number of subjects, with a plateau reaching maximum Sensitivity with a group sample bigger than 20 and SNR above 30. Sensitivity was only slightly worse for Modularity, but it should be noted that for the LFR network the size distribution of the planted modules was narrower than for the ring of cliques, thus making the resolution limit less evident.

In the lower SNR regime, Asymptotical Surprise presented the best performance in terms of NMI and Sensitivity, with a slower decay for decreasing SNR.

Specificity was almost equivalent across the three methods, with a quick convergence to the maximum value of 1 for high SNR and good performance (around 0.97) for low SNR. Asymptotical Surprise presented a faster decay with decreasing

SNR. However, it should be noticed that the scale of Specificity has a very narrow range (0.97-1.00), and the differences between the three methods were relatively small.

Consensus analysis applied to Newman's Modularity to assess the potential effects of the degeneracy of nearly optimal solutions did not show substantial differences in the comparison with the other methods. Notably, Infomap showed a large variability in Accuracy for lower SNRs and number of subjects. Under closer examination, however, it appeared that the increased variance for Infomap was due to occasional runs in which the algorithm only retrieved one or two large modules. This is a known problem with Infomap (and other algorithms based on random walks) that depends on the need to parametrize the teleportation step in order to make the dynamics ergodic [223].

For the sake of completeness, I also computed Accuracy and Matthew Correlation Coefficient for the same model networks. As shown in Figure 4.8A, Accuracy of Newman's Modularity is lower for small SNRs and number of subjects and in any case it does not reach 100% of true positives classification even in the no-noise condition. Infomap accuracy is high for SNR greater than 20, largely independent on the number of subjects. The large variance of Infomap for low SNRs is due to the merging of all nodes in a single large community in a few runs, as discussed in the main text. Asymptotical Surprise, behaves well in terms of Accuracy and has the least variability across all methods, plateauing at 100% for SNRs greater than 20 and more than 40 subjects. In terms of MCC, Figure 4.8A shows that Asymptotical Surprise behaves comparably or better than Infomap. Newman's modularity is the least performer, due to its resolution limit, and never reaches maximum MCC. Interestingly, Asymptotical Surprise slightly outperforms Infomap regarding MCC in the single subject case in the no-noise condition. The comparison of the three methods is similar in the case of the LFR benchmark, as shown in Figure 4.8B. Altogether, the picture that emerges from the analysis of Accuracy and MCC is entirely consistent with the results shown in this section.
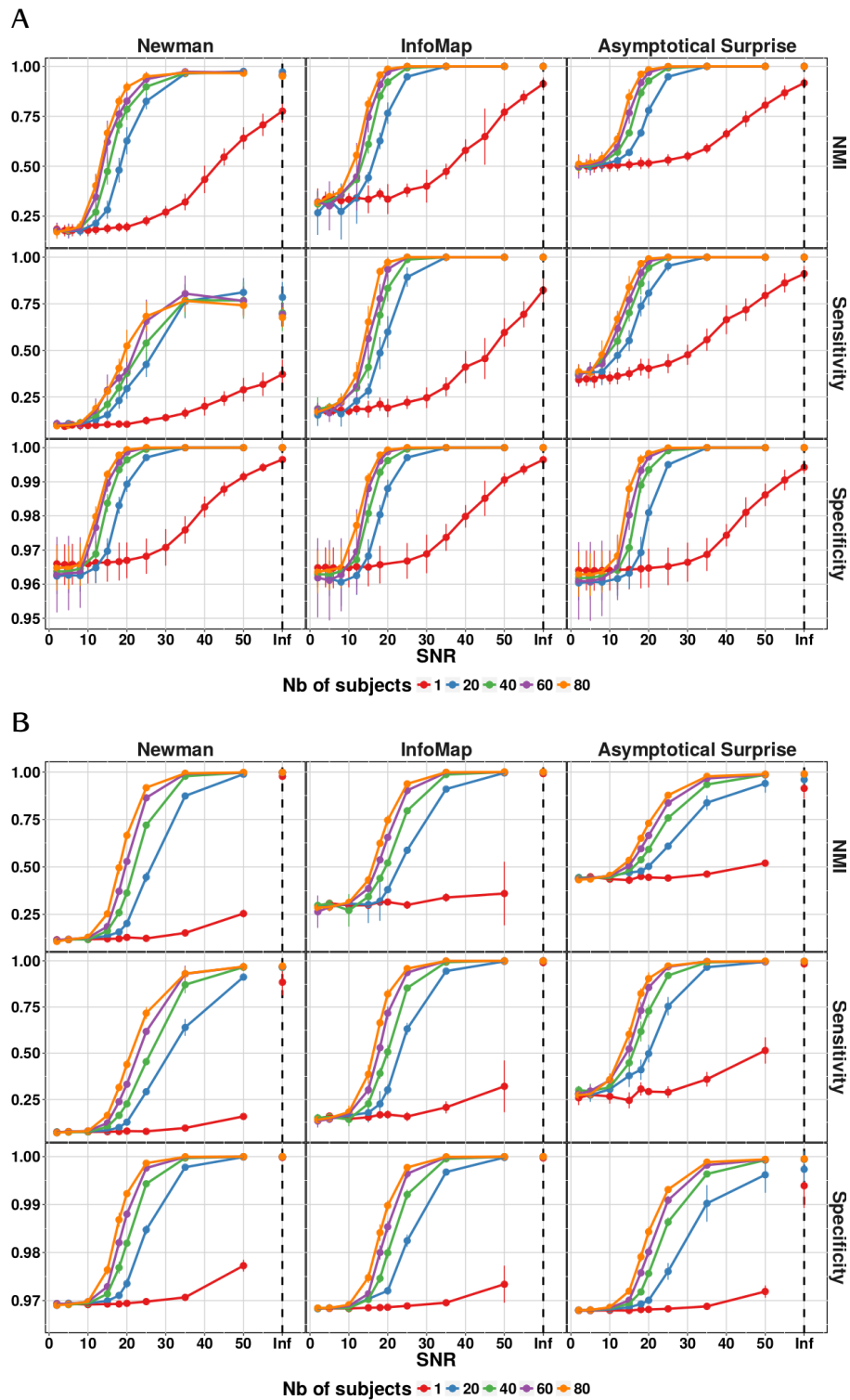
Figure 4.7: NMI, Sensitivity and Specificity of the three community detection algorithms applied to a power-law ring of clique network **(A)** and to Lancichinetti-Fortunato-Radicchi network **(B)**. SNR indicates Signal to Noise Ratio, and Inf the situation with a network structure unperturbed by noise. Number of Subjects indicates the different number of instances used to generate the group level network.

Figure 4.8: (A) Accuracy and Matthew correlation coefficient on the modified ring of clique benchmark. **(B)** Accuracy and Matthew correlation coefficient on the LFR benchmark.
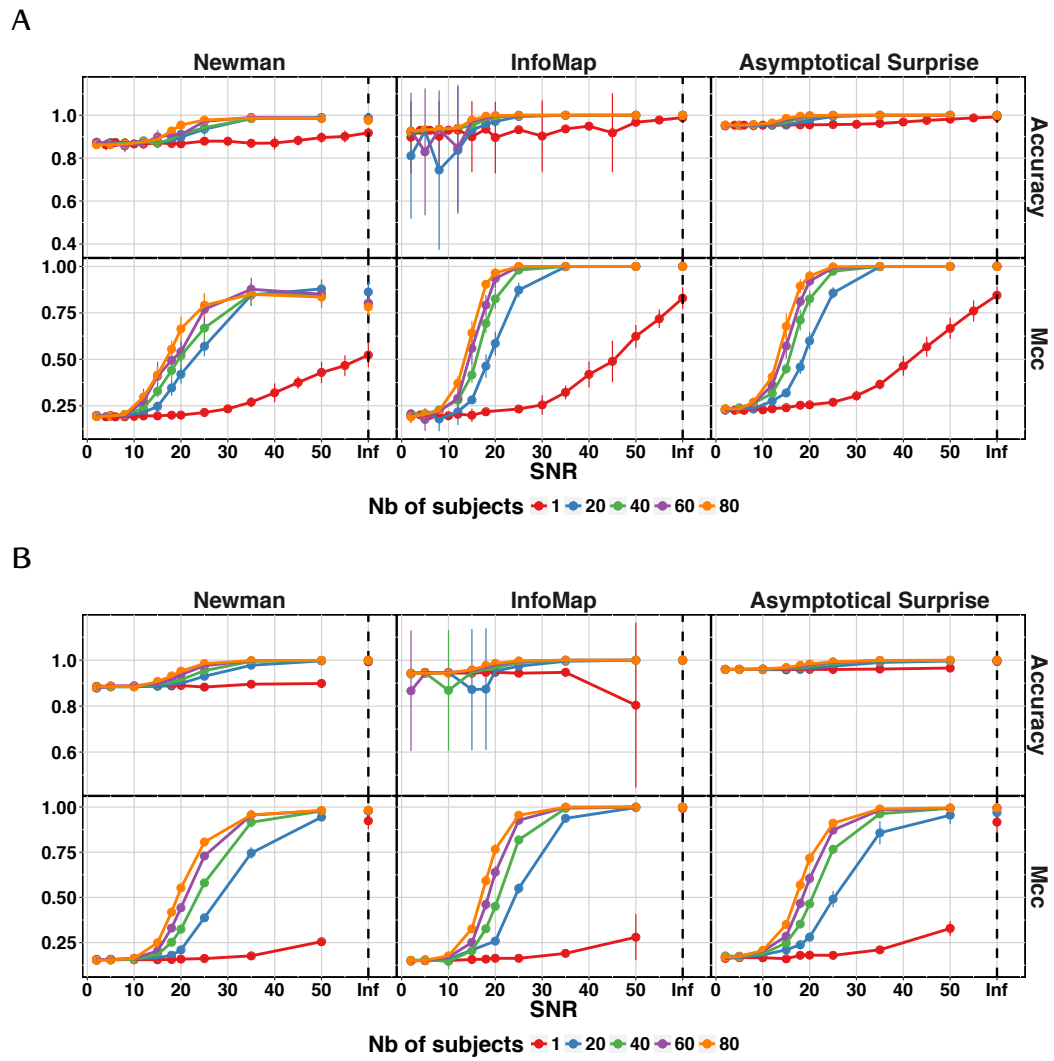
4.6 ASYMPTOTICAL SURPRISE ON RESTING STATE DATASET

Figure 4.9 shows a comparison between the modular structure of the resting state fMRI dataset obtained with Newman's Modularity, Infomap and Asymptotical Surprise. For each method, I had $10,000$ independent runs and picked the partition with the best value of the respective fitness functions ($Q = 0.4967$, $\mathcal{L} = 8.5173$, $S_a = 5925.3$, for Modularity, Infomap and Asymptotical Surprise, respectively). The three methods showed significantly different partitions (relative NMIs in Table 1), with a number of detected communities of 10, 19 and 47 for Modularity, Infomap and Asymptotical Surprise, respectively. Interestingly, Modularity detected a relatively uniform size distribution, consistent with the intrinsic scale built into the fitness function. Infomap showed a wider distribution of module sizes, with number of nodes ranging between 156 and 3, while Surprise showed the largest spread, and included communities as small as single nodes (singletons).
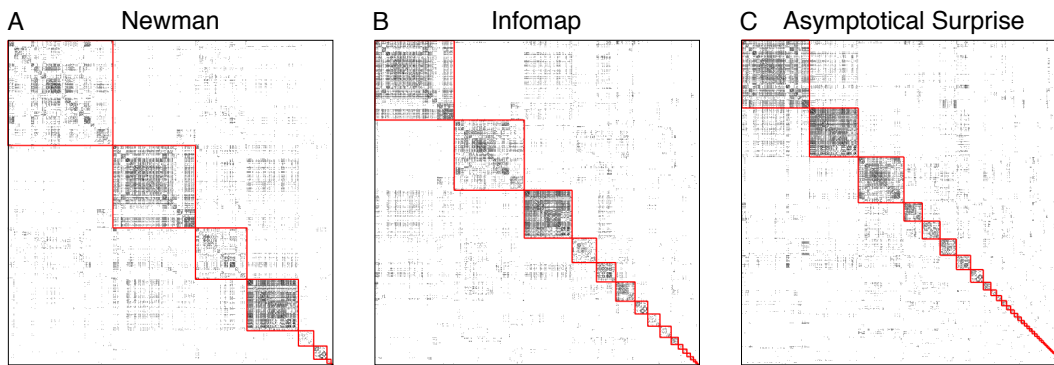
Figure 4.9: Adjacency matrix of the resting state functional connectivity network. The node indices have been reordered by module membership in each graph, and the red lines highlight the community structures obtained by A) Louvain-Newman's Modularity ($Q = 0.4967$); B) Infomap ($L = 8.5173$); C) Asymptotical Surprise ($S_a = 5925.28$).

Figure 4.10 shows the 16 largest modules detected by Asymptotical Surprise, ranked by the number of nodes comprised in each community.

The first and largest module (Figure 4.10A) includes the pre- and post-central gyri, part of the supramarginal gyrus and supplementary motor area. The second community (Figure 4.10B) consists largely of nodes belonging to the occipital lobe: the visual areas and the surrounding calcarine sulcus, the lingual and fusiform gyrus. The third module (Figure 4.10C) reflects the Default Mode Network, spanning the temporoparietal cortex, the medial prefrontal cortex and the posterior cingulate/precuneus. The nodes involved in the executive frontal functions form the fourth largest community. Interestingly, nodes in the communities D, E, G are the major players that take part in the so-called frontoparietal attentional network [202]. The auditory network, comprising temporal areas, was detected as a distinct community (Figure 4.10F). Deeper structures emerge as separate modules in Figure 4.10H, with subcortical areas including the basal ganglia, i.e. putamen, globum pallidum, caudate nucleus and the whole thalamus. The hippocampus and the parahippocampal gyrus were identified as separate communities (O and P). Additional, smaller substructures are shown in the third and fourth row of Fig-

ure 4.10, including the Supplementary Motor Area (Figure 4.10J) and the orbital (Figure 4.10M) and orbitofrontal (Figure 4.10I) modules, containing nodes from Brodmann area 47.
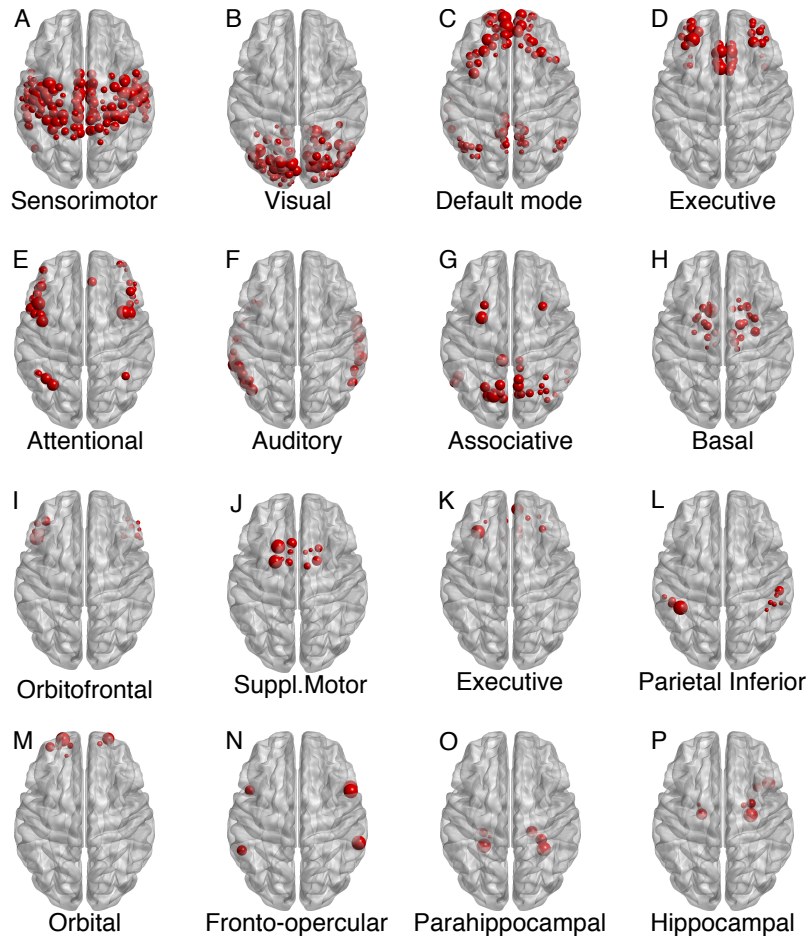
Figure 4.10: Sixteen largest modules found by Asymptotical Surprise Maximization in the resting state network overlaid on an MRI brain template. The modules are ranked by decreasing size, and named after corresponding functional networks previously identified by multivariate analysis of resting state fMRI data, or by the comprised anatomical districts.

Newman's Modularity retrieved four large, relatively uniform communities, corresponding to the Default Mode Network, the central network, occipital and frontoparietal networks. This is in keeping with previous studies using Modularity optimization by spectral decomposition [34], and consistent with the strong resolution limit that affects this method. Additionally, a few smaller modules were found by Louvain optimization of Newman's Modularity, corresponding to the basal ganglia, the hippocampal/parahippocampal formation, and two asymmetrically distributed subcortical clusters.

Infomap identified 19 communities of various sizes, also shown in Figure 4.12. The largest modules showed a close correspondence with those identified by Asymptotical Surprise, albeit with some notable differences. By way of example, the largest component includes the motor-sensory and auditory modules, identified as separate communities by Asymptotical Surprise. The Default Mode Network
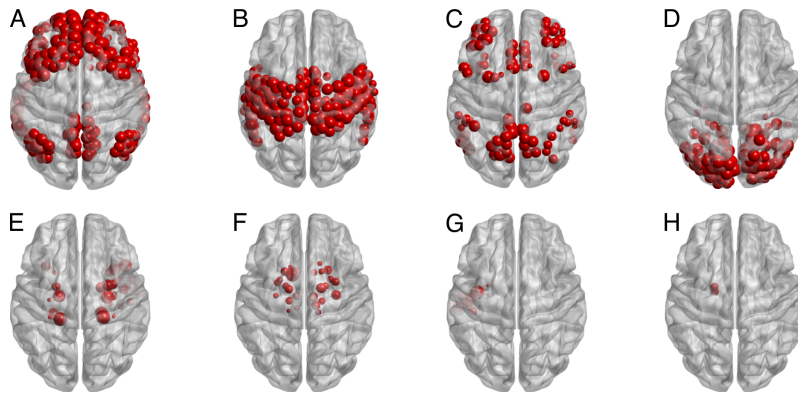
Figure 4.11: First largest 8 modules of the optimal partition as found by Newman's Modularity, with a value of the Modularity Q = 0.4967. Module 9 is not shown as it consists of a single node.
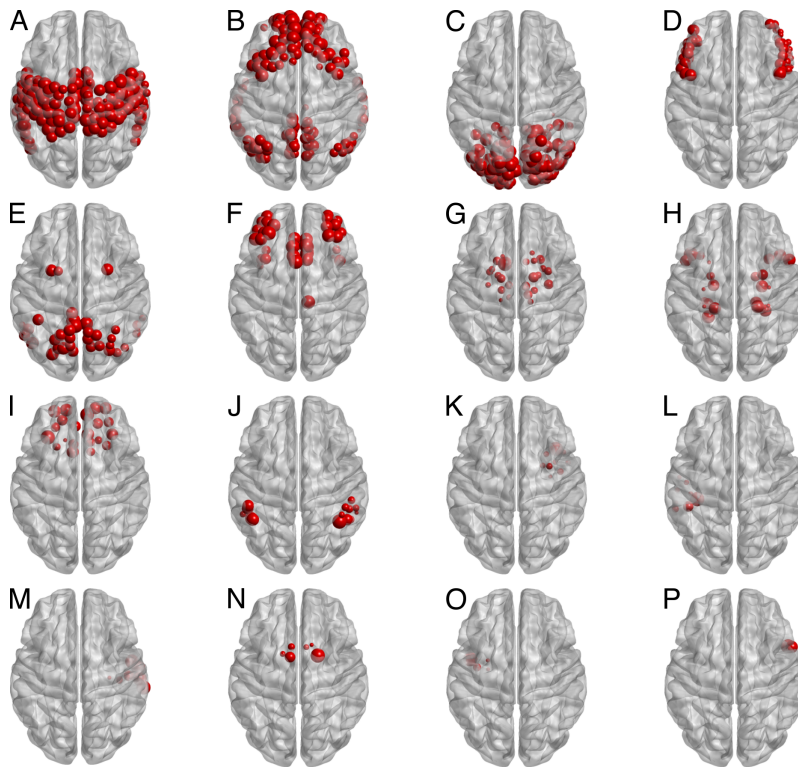
Figure 4.12: First largest 16 modules of the optimal partition as found by Infomap, with a value of the description length L = 8.5173.

retrieved by Infomap includes parts of the temporal cortices that are not normally associated with the DMN. Similarly, hippocampus and the parahippocampal modules were merged by Infomap and resolved as individual modules by Asymptotical Surprise. Other modules, including the visual, associative and executive networks (C, E and F in Figure 4.12, respectively) were qualitatively very similar to those identified by Asymptotical Surprise.

Altogether, the picture that emerges is consistent with the idea that the resolution limit is more severe in Newman's Modularity than in Infomap, and that Asymptotical Surprise presents the best resolving power among the three meth-

ods in a real-world network with finite SNR and variability as the resting state functional connectivity network used for this study.

4.6.1 Hub classification

Maps of the anatomical distribution of the participation coefficient and within-module degree show substantial differences between the three community detection methods (Figure 4.13), resulting in discrepancies in the identification of the connector hubs for the same functional connectivity network.

Figure 4.14 shows the nodes with simultaneously high values of participation coefficient and within-module degree (connector hubs, according to the Guimera and Amaral's classification). All three methods pinpoint connector hubs in the superior, superior medial and middle frontal areas, as well as in the supplementary motor area. However, substantial differences are observed for other hub regions. The partition of Asymptotical Surprise localizes connector hubs in the Temporal Middle and Frontal Middle gyri, as well as in the Rectus, Middle Cingulate Cortex, Lingual gyrus and in the Precuneus.

Community detection by InfoMap results in the identification of hubs that are partially consistent with either of the two other methods, in keeping with the idea that its resolution limit is less severe than for Newman's Modularity. Altogether, these findings indicate that node role classification is method-dependent, and may be affected by the resolution limit.
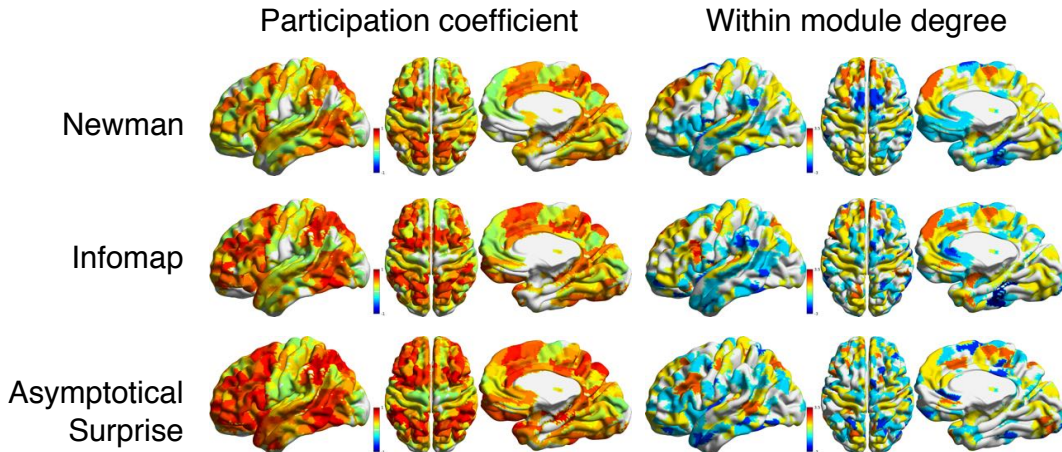
Figure 4.13: Anatomical distributions of the participation coefficient and within-module degree z-score for the resting state functional connectivity network partitioned by the three community detection methods.

4.6.2 Validation of Asymptotical Surprise in model networks

The performance of Asymptotical Surprise optimization by PACO was assessed in model graphs with a built-in community structure, and compared with two established community detection methods. I have chosen two synthetic benchmark networks, the ring of cliques and the LFR network.
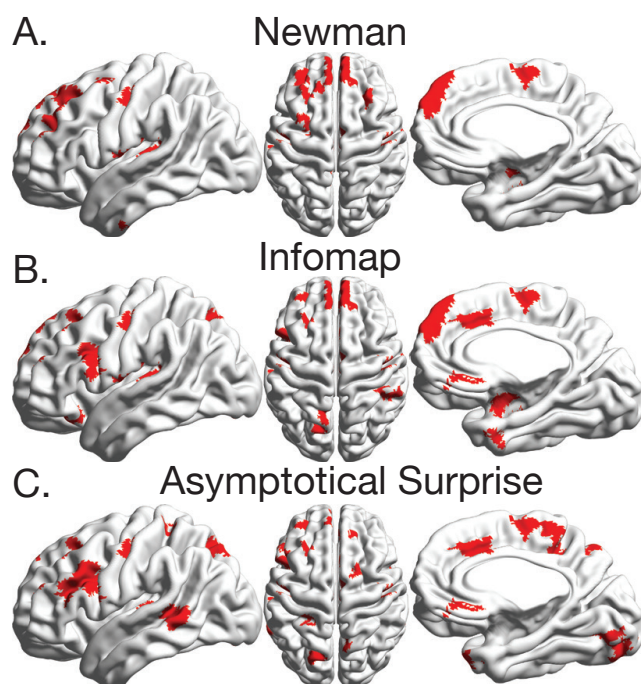
Figure 4.14: Nodes presenting simultaneously large values of the participation coefficient and within-module degree (larger than 0.6 and 1.5, respectively) for the three community detection methods. These nodes are thought to represent the connector hubs responsible for the integration of the networks's modules.

The ring of cliques presents a clear-cut modular structure by construct, with modules corresponding to complete subgraphs of variable sizes sampled from a power-law distribution.

This toy network proved useful to assess the effects of the resolution limit in the presence of a wide distribution of cluster sizes. The effects of this limit were particularly apparent for Newman's Modularity (Figure 4.7A), that showed poor Sensitivity even for noiseless rings of cliques, plateauing at a value of 0.75. This is consistent with the findings of [72], which showed that for Modularity the resolution limit is set by the square root of the total number of edges in the graph. For Infomap, this limit is less severe and is determined by the number of inter-cluster edges [164]. Accordingly, the effects of the resolution limit were not apparent in this model network, where modules are sparingly connected by single edges. Asymptotical Surprise presented the best performance, consistent with the idea that this cost function is quasi-resolution limit free [176].

However, real brain networks are characterized by heterogeneous distributions of node degree, with fat tails and power-law decays [27]. Such heterogeneity is critical, as it determines some of the remarkable features of brain connectivity networks, including resilience to random failure and rich-clubness [44, 58]. To provide a more realistic benchmark, I used the Lancichinetti-Fortunato-Radicchi algorithm [80], which made it possible to generate networks with realistic and tunable power law degree distribution and community sizes.

For LFR networks, the difference in performance in the low-noise regime was more nuanced for the three methods compared in this study, possibly a result of a

fuzzier community structure of the LFR network compared to the ring of cliques, and of the narrower distribution of cluster sizes. However, the picture appeared different when noise and intersubject variability were injected into the network structure.

Noise and other sources of variability in the data can significantly affect the structure of the resulting network representation. Noisy fMRI time-courses, for example, may introduce spurious correlations in brain functional connectivity networks. This problem may be particularly relevant for methods endowed with high resolution, like Asymptotical Surprise, that may be more vulnerable to False Positives generated by the misassignment of peripheral nodes, particularly in small clusters. Hence, the resolving power of community detection methods should be gauged against Specificity, which may be affected by noise in the distribution of edges that define the network's structure. However, to the best of our knowledge, this aspect has never been considered in the existing literature assessing the performance of community detection algorithms as applied to the study of brain connectivity.

To this end, I have devised methods to inject noise, with amplitude and spectral distribution that mimic those of experimental noise, into networks with a well defined planted structure. Moreover, I have generated different instances for each network, corresponding to different subjects in a group, to account for intersubject variability that occurs in typical neuroimaging studies.

Unsurprisingly, for all methods and networks, detection of the planted structure improved with decreasing levels of noise, and with increasing number of subjects in the study. However, Asymptotical Surprise appeared to provide a superior performance in terms of NMI and Sensitivity to the planted structure for lower SNRs in both types of networks, while its Specificity was in line with that of resolution-limited methods like Newman's and Infomap (Figures 4.7A, 4.7B). This rules out the idea that the higher sensitivity to small clusters of Asymptotical Surprise may be detrimental in noisy networks, making it more vulnerable to small, spurious modules.

4.6.3 Asymptotical Surprise optimization on resting state networks

Application of Asymptotical Surprise maximization to a group-level, resting state functional connectivity network from the brains of 27 healthy subjects revealed a heterogeneous distribution of modules, with large and small modules coexisting in the optimal partition. This is in keeping with previous findings with binary Surprise [182]. These modules closely reflect functional networks reported in many studies using Independent Component Analysis or other multivariate methods, including the sensorimotor, visual, default mode, executive, and attentional networks. Moreover, anatomically defined subcortical structures, like the hippocampus and parahippocampal formations emerged as independent modules.

While this is entirely consistent with our understanding of the neurofunctional and anatomical organization of the human brain, the accuracy of Asymptotical Surprise in identifying these networks is notable. Indeed, Surprise, like other graph-based community detection methods, divides networks into disjoint clusters of nodes on the basis of topological criteria. While a correspondence between

topological modularity and functional networks identified by, e.g. Independent Component Analysis, may be expected, it is not a given, for they are defined on different principles. Indeed, multivariate methods like ICA separate components on the basis of the statistical independence of the time-courses, and do not convey information regarding the mutual relationship between modules nor about their topological organization.

Previous studies applying resolution-limited methods like Newman's Modularity to the same dataset hereby analyzed [34] found a few, large modules encompassing large-scale networks, but failed to identify finer, neurofunctionally plausible substructures like those shown in the present study. Infomap, on the other hand, proved sensitive to heterogeneously distributed clusters, thus implying that this method does not have an intrinsic scale, like Modularity and variations thereof based on the introduction of a resolution parameter.

However, Asymptotical Surprise appears to provide superior performance in identifying small subnetworks, particularly in the presence of noise, thus suggesting that this method may represent a new standard for community detection in brain networks. It should also be noted that no symmetry constraint was imposed, and the symmetrical bilateral distribution of nodes in the retrieved modules arises entirely from Asymptotical Surprise optimization.

Hierarchical clustering methods have been extensively applied to investigate the structure of brain connectivity networks, showing smaller and smaller clusters as the modules are iteratively subdivided [59]. Maximization of Asymptotical Surprise reflects the optimal cut through the dendrogram representing connectivity at these different levels of subdivision and provides information on the optimal partition of the network. Hence, the heterogeneous distribution of cluster sizes retrieved by Asymptotical Surprise suggests that multiple scales of structure exist at the same level of the dendrogram.

Finally, abnormal functional connectivity has been observed in a number of neurological and psychiatric diseases, but the coarse resolution of methods like Newman's Modularity [26] may not have detected differences in the modular organization of networks in patients compared to healthy controls. The improved resolution and sensitivity to multiscale structure afforded by Asymptotical Surprise may provide a powerful means to assess the brain functional architecture in disease states, thus contributing a potential imaging-based marker and a key to interpret the functional effects of aberrant connectivity.

4.6.4 Hub classification

The presence of heterogeneously distributed modules in functional connectivity networks may have important consequences for our understanding of the brain functional organization. By way of example, it has been shown that highly connected nodes, or hubs, are critically important in brain connectivity networks, and may play different roles depending on their position and connectivity distribution within and between modules [27]. Hubs that primarily connect to nodes within the same community are dubbed "provincial hubs", and are thought to be responsible for the definition and stability of the modules. Conversely, hubs that connect different modules are referred to as "connector hubs" and ensure integration of

the activity of the network. The classification of hubs strongly depends on the modular structure that is considered, and inaccurate partitioning due to the resolution limit can lead to the wrong interpretation of their role in the interplay between segregation and integration of brain function [27]. The present study suggests that this may have been the case in previous studies, in which resolution limited methods characterized by an intrinsic scale have been used, and provides a solution that may enable more accurate classification of hubs and nodes.

The connector hubs identified by our three methods (Asymptotical Surprise, InfoMap and Newman) present some substantial differences, consistent with the idea that hub classification depends on community structure. These differences are particularly interesting in the light of the important role that connector hubs are thought to play in integrating information flow through the brain, and their putative role in brain disease [28, 25]. By way of example, the Precuneus and the Cingulate Cortex are highlighted by Asymptotical Surprise, but not by Newman's Modularity, as connector hubs. These are two key elements of the Default Mode Network that have been consistently identified as vulnerable regions in neurological diseases [58, 224].

Community detection by resolution limited free methods should enable more accurate classification of hub nodes, and improve our understanding of their role in brain disease.

4.7 LIMITATIONS OF ASYMPTOTICAL SURPRISE

Some caution should be taken in the interpretation of the graphs in Figures 4.7A,B. Indeed, the SNRs of the synthetic networks I have generated, reflect noise with features, like a Rician distribution, that mimic some, but not all aspects of the variability of experimental data. By way of example, the brain parcellation scheme applied to define the nodes, and the heterogeneity of voxels within these parcels may play a role that is hard to model in toy networks [225]. Hence, the simulated Sensitivity and Specificity as a function of SNR and number of subjects should not be taken as absolute values to be used in the power and sample size estimation in real experimental designs. Nevertheless, these simulations provide useful information on the dependence of these parameters on noise levels, and a rigorous means to assess the relative merits of different community detection methods.

Finally, I should note that the maximum value of Asymptotical Surprise calculated with PACO is an index of quality of the entire partition, and not of individual modules. Hence, individual modules may not all have the same strength of internal cohesiveness relative to their connection with other modules. We have found hints of this phenomenon in the comparison of nearly-optimal partitions obtained in the $10,000$ runs of PACO that I have performed to find the optimal community structure for this network. The overall community structure appeared to be robust, with most modules persistently emerging in every nearly-optimal partition, but in some cases pairs of modules split or merge in otherwise similar solutions. Most notably, this was observed for the thalamus that in some instances was merged with the basal cluster and in others, featured as a separate module. This phe-

nomenon may be less critical for methods like Newman's Modularity that have an intrinsic scale and retrieve uniformly distributed modules.

4.8 CONCLUSIONS

I have extended the use of Surprise, a resolution-limit-free fitness function for the study of the modular structure of complex networks, to weighted brain functional connectivity networks. Specifically, I have developed a novel method, dubbed PACO, for the optimization of Asymptotical Surprise, a weighted counterpart of Surprise in the limit of large networks. I have applied PACO optimization of Asymptotical Surprise in synthetic networks to evaluate the relative merits of this novel approach against Newman's Modularity and Infomap, two of the leading methods used for community detection in brain connectivity networks.

Specifically, I have implemented a process to inject noise into networks endowed with a ground-truth modular structure to assess the trade-off between improved resolution afforded by Asymptotical Surprise and potential sensitivity to spurious correlations introduced by variability in the data. Asymptotical Surprise optimization proved superior to existing methods in terms of Sensitivity and accuracy in the detection of the planted structure as measured by Normalized Mutual Information, while showing comparable Specificity.

I have also applied this approach to the partitioning of functional connectivity networks from resting state fMRI experiments. Direct comparison with other methods clearly demonstrated an improved capability to identify neurofunctionally plausible and anatomically well-defined substructures otherwise concealed by the resolution limit. Asymptotical Surprise revealed a complex modular structure of resting state connectivity, with communities of widely different sizes reflecting distributed functional networks alongside with small, anatomically or functionally defined modules.

This evidence corroborates the idea that the resolution limit may have negatively affected current models of the brain modular organization and the identification of the hubs responsible for the integration of functional modules. For this reason, the application of methods like Asymptotical Surprise provides a novel, powerful approach to study the modular structure of brain connectivity beyond this limit.

5 | BEYOND THE RESOLUTION LIMIT IN THE DISEASED BRAIN

Since the work of pioneers of neuropsychology, it has been observed that many neurological and psychiatric disorders can be described as dysconnectivity syndromes [226]. The emergence of symptoms or functional impairment in these disorders can be related to the damage or abnormal integration of spatially distributed brain regions that would normally constitute a large-scale network subserving function. Abnormal patterns of brain functional connectivity have been consistently observed in patients affected by Schizophrenia (SZ) using functional MRI and other neuroimaging methods. Graph theoretical approaches have been applied to study defective interactions and modular organization in networks of distributed brain areas, perhaps a result of dysfunctional conscious integration in SZ.

As shown in the previous chapters, current graph analysis methods suffer from a fundamental resolution limit, as they fail to detect modules that are smaller than a scale determined by the entire connectivity network.

In this chapter, the first application of Asymptotical Surprise optimization is demonstrated for the study of the modular organization of resting-state functional connectivity networks in a large cohort of SZ patients, and in matched healthy controls. These preliminary results have been obtained with the collaboration of my colleague Cécile Bordier at the Center for Neuroscience and Cognitive Systems of "Istituto Italiano di Tecnologia" in Rovereto (Italy).

Application of the methodological advances introduced in this work showed a substantial fragmentation and reorganization involving primary sensory, auditory and visual areas in SZ patients. Conversely, frontal and prefrontal areas related with higher cognitive functions appeared to be less affected, with changes involving mostly language-processing regions. These findings, not yet published, support the hypothesis that cognitive deficits in SZ may be driven by impairments in basic perceptual processes that localize to primary sensory brain regions.

5.1 FUNCTIONAL CONNECTIVITY IN SCHIZOPHRENIA

Schizophrenia has been associated with aberrant functional connectivity as measured by neuroimaging methods in a number of studies [227, 228, 229, 230, 231, 65, 66, 232]. This growing evidence is in keeping with the disconnectivity hypothesis of Schizophrenia that posits that the core dysfunction of this disease may correspond to a weakening of the functional interactions between specialized brain areas [233, 234, 235], resulting in defective integration of activity in distributed networks and cognitive disintegration [236]. Indeed, psychotic symptoms akin to those of schizophrenia, including hallucinations and delusions, are also observed in certain neurological disorders that involve disruption of corticocortical

and cortico- subcortical connections [237]. Understanding the nature of connectivity alterations in SZ patients and their effects on brain functional integration may provide important insights into the etiology of this devastating disease, as well as potential diagnostic or prognostic markers.

To this end, graph theoretical approaches have been proposed as a powerful framework to assess topological features of functional connectivity networks [238, 27, 239, 140]. Several alterations in graph-related metrics of resting state connectivity have been identified in schizophrenia patients, including reduction in global network efficiency [230, 27, 240], small worldness [230, 241, 242] and rich-club organization of high-connectivity nodes [243].

Several studies have assessed the modular structure of resting state functional connectivity networks derived from functional MRI in Schizophrenia patients compared to healthy controls [230, 65, 244]. Disrupted modular organization and reduction in Modularity, a measure of segregation of functional modules within the network, was found in Childhood Onset Schizophrenia [65]. Reduced Modularity was associated with a proportional increase in inter-cluster edges and decrease in intra-cluster edges [66]. Lerman-Sinkoff et al. [244] reported similar community structures in adult schizophrenia patients and healthy subjects under stringent control of potential sources of imaging artifacts, with small but significant alterations of node community membership in specific networks. Yu et al. [245] found reduced overall connectivity strength and a larger number of communities in the patients' group (6 in SZ subjects vs 5 in healthy controls).

These pioneering investigations provide important indications that the modular organization of functional connectivity networks may be altered in patients affected by schizophrenia. However, graph theory as applied to the study of brain networks is still in its infancy, and several methodological and conceptual issues that are still open may have affected early studies.

In the previous chapters, I have shown that the resolution limit severely hampered the ability to resolve the modular organization of human brain connectivity networks and to capture their complex community structure. This pervasive limit is likely to have biased previous studies in clinical populations and may have prevented detection of differences in the organization of functional connectivity in patients and controls at a finer scale. Indeed, even though previous studies in SZ populations systematically report a substantial reduction in functional connectivity strength and modularity compared to healthy controls, differences in the number, size, and boundaries of functional modules appear to be inconsistent across studies and dependent on the specific clustering approach that was adopted.

The deleterious effects of the resolution limit propagate to the evaluation of relevant topological parameters that depend on the network's community structure. These include within-module-degree and participation coefficient, parameters that enable the identification of highly connected nodes, or hubs, responsible for the integration and efficient exchange of information between modules [27].

5.2 ASYMPTOTICAL SURPRISE OPTIMIZATION ON SCHIZOPHRENIC PATIENTS

Here, capitalizing on these important methodological advances discussed in the previous chapters, we applied Asymptotical Surprise to resolve and compare the modular structures of resting state functional connectivity networks in two cohorts of 78 schizophrenia subjects and 91 controls.

In contrast with previous studies, we have found profound changes in the resting state brain connectivity structure of schizophrenia patients, with a substantial functional reorganization reflecting both fragmentation and merging of functional modules. Additionally, we investigated alterations in node-wise participation coefficients and the resulting rearrangement of integrative brain regions in patients.

5.2.1 Material and methods

MRI data were downloaded for 78 people with schizophrenia strict (SCZ) (64 males, 14 females) and 91 healthy controls (CON) (65 males, 26 females) from the open COBRE database `http://fcon_1000.projects.nitrc.org/indi/retro/cobre.html`. Ranging ages going from 18 to 65 years in both groups. All the subjects in the COBRE were screened and excluded if they had history of neurological disorder, history of mental retardation, history of severe head trauma with more than 5 minute loss of consciousness, history of substance abuse or dependence within the last 12 months. Ethical statements are contained in the original publication of this dataset. Images were acquired with a Siemens MIND TRIO 3T scanner equipped for echo-planar imaging (EPI). Echo-planar imaging was used for resting state fMRI data collection with (Repetition Time) TR=2s, (Echo Time) TE=29ms, matrix size: 64x64, slices=33, voxel size=$3 \times 3 \times 4$ mm^3 (for more details see [246]). A total of 150 volumes of functional images were obtained for all the subjects except one (this subject was excluded from the present study). The data were pre-processed using SPM8 (Wellcome Trust Centre for Neuroimaging, London, UK). After discarding the four initial volumes, the remaining volumes were slice-timed, head-motion realigned and normalized to the standard MNI EPI template space (voxel-size re-sampled to $3 \times 3 \times 3$ mm^3). Then, for each participant, 638 regional mean time series were computed by averaging the voxel time series within each of the parcellized areas of [34] template. Finally, after regressing the movement parameters from each areas signal, we estimated the connectivity matrix by computing pairwise inter-regional correlation for each participation. In order to have a connectivity matrix by population, we Fisher-transformed individual correlation matrix and average them by group to get an adjacency matrix. In order to determine the optimal modular partitions for the experimental groups beyond the resolution limit, we applied optimization of Asymptotical Surprise by PACO. Prior to community detection, the group level adjacency matrices were sparsified using a percolation analysis approach to remove weaker edges and reduce the effects of noise, thus maximizing information about the network's modular structure [97, 116, 111], as described in section 1.6.

5.2.2 Disruption of modular structure in SZ patients

Figure 5.1 shows the group-level adjacency matrices, with the node indexes rearranged by module membership, for the control and schizophrenia groups. Disjoint clusters of nodes, or modules, are delineated by red lines. We found 44 communities in the control group, with module sizes ranging between 141 and 1 nodes. The optimal partition of the patients' group comprised 39 communities and showed a less heterogeneous size distribution. The statistical significance of the difference in community structure was assessed using a recently proposed permutation approach [66], resulting in a p-value of $p = 0.009$ (1000 permutations, FDR corrected). The smaller number of modules in the schizophrenia group may appear somewhat counterintuitive, in the light of overall weaker functional connectivity strength in this group. However, while the largest modules appear to break up into smaller modules in patients, the tail of the distribution of community sizes is fatter in the schizophrenia group, thus indicating aggregation and reorganization of smaller modules.
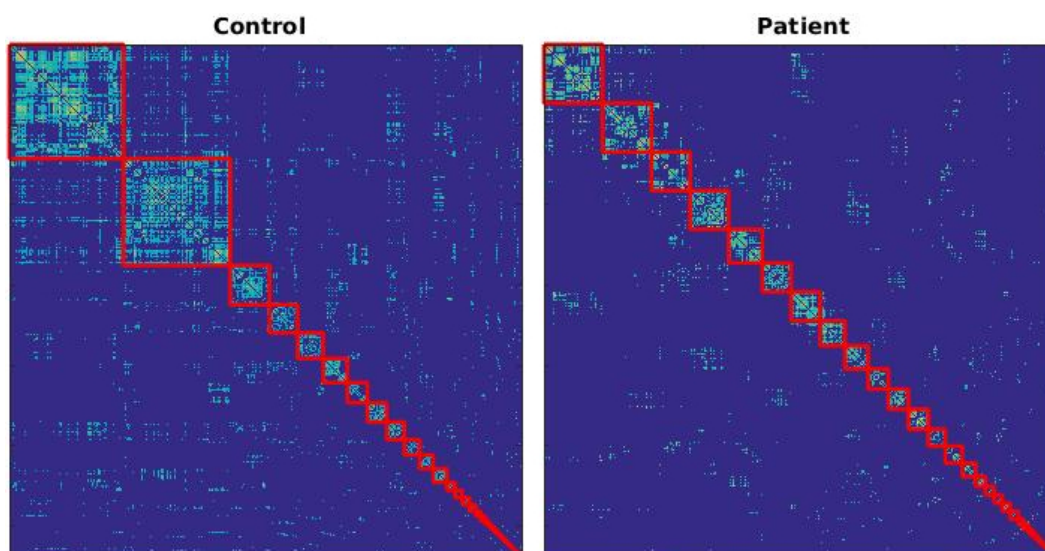
Figure 5.1: Optimal Asymptotical Surprise partitions for the two populations.

Figure 5.2 shows the distribution of functional modules in the two groups overlaid on an anatomical template. Note that the colors denoting the communities were chosen independently in the two groups to maximize contrast between adjacent modules. Differences in the modular structures of functional connectivity in the two groups are apparent and involve complex reorganization of nodal membership across modules. The main differences in modular organization between the two groups involve the sensorimotor, visual and auditory cortices. The large central module in the healthy controls' group (dark blue in Figure 5.2A) comprises somatosensory cortices and temporal auditory cortices, consistent with previous findings in healthy volunteers [247]. In schizophrenia patients, this module breaks up into four different clusters of nodes. Similarly, the controls' large occipital module (light brown in Figure 5.2) is split in the patients' group, with primary visual cortex standing as an independent community together with part of the inferior temporal lobe. The more dorsal part of the occipital community includes part of

A. Controls
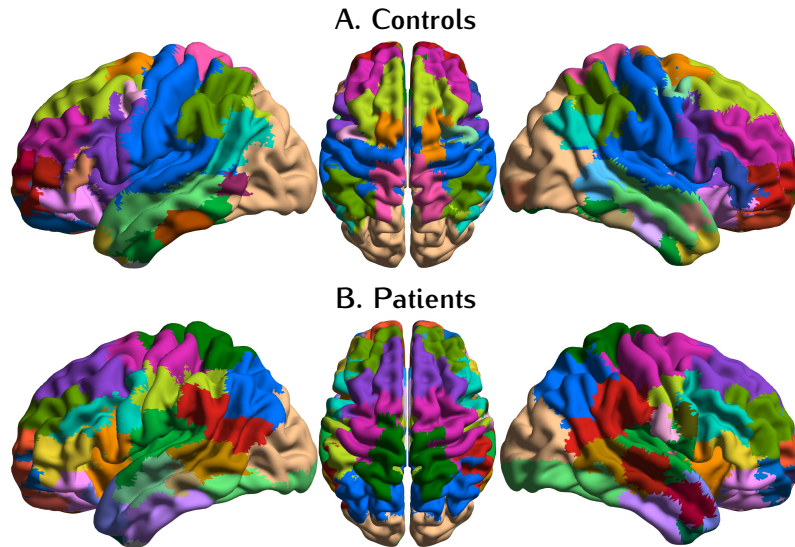
B. Patients

Figure 5.2: Maps representation of the communities for each group: A. Healthy control, B. Schizophrenia patient. Please note that the colors of each community are random even across groups.

the superior parietal lobule in HCs, but not in SZ patients, where the boundary of this community lies in the vicinity of the parieto-occipital fissure.

Substantial fragmentation and reorganization is also observed in the modular structure of the temporal cortex. In controls, we find well-delineated modules comprising the middle temporal gyrus and the inferior temporal gyrus, while the superior temporal gyrus is part of larger community that includes somatosensory cortices. In patients, the superior temporal gyrus is separated from the larger somatosensory community, and is split into two modules, anterior and posterior, respectively, that comprise also part of the supramarginal cortex. The middle temporal gyrus is split rostrocaudally into four different communities that include parts of the superior and inferior gyri. The inferior gyrus is split in ttwo modules, anterior and posterior, respectively. The anterior module includes part of the middle temporal gyrus, while the posterior one extends to visual cortices.

Interestingly, the angular gyrus and the supramarginal gyrus appear as separate modules in healthy controls, but, in patients these areas are merged into a single community including the temporoparietal junction. Finally, between-group differences in frontal lobe organization pertain particularly the language regions, with the Broca area forming an independent community in patients. The modular structure of other frontal and prefrontal areas is consistent in the two groups.

Figure 5.3 shows statistically significant voxelwise differences in participation coefficient, a measure of diversity in intermodular connections of individual nodes. Nodes characterized by high participation coefficients have many links pointing to different modules, and are thought to play an integrative role. Conversely, nodes with most links pointing to other nodes within the same community are dubbed provincial hubs, and contribute to defining functional segregation of their communities. Significantly larger participation coefficients are observed observed in sensorimotor, visual and auditory areas of SZ patients. Conversely, lower participation coefficients in patients are observed in frontal and parietal regions, with

A. SCZ > CON

B. CON > SCZ
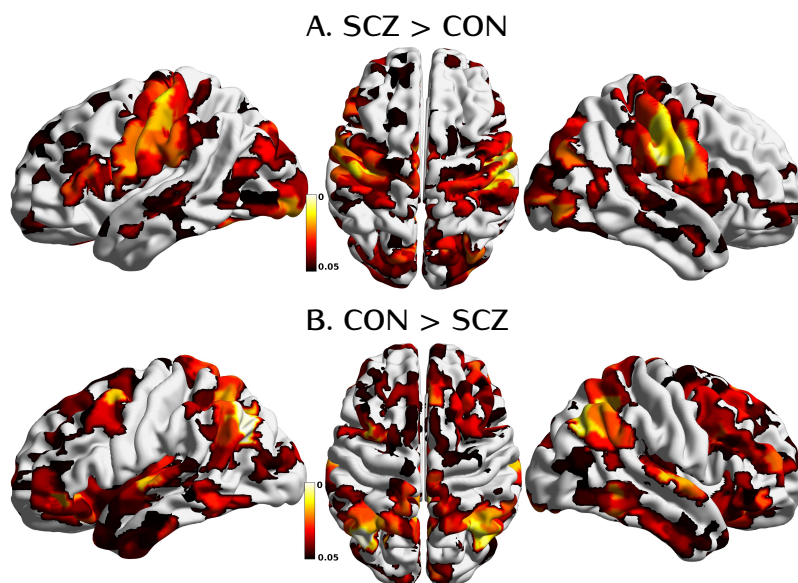
Figure 5.3: Participation coefficient map of the difference between the two populations. a. Nodes with higher participation coefficient in SCZ than for CON; b. Nodes with lower participation coefficient in SCZ than for CON.

the most prominent decrease in the temporal primary auditory cortex in the vicinity of the Heschl gyrus. Substantial differences are also observed in functional connectivity of areas related with language generation and processing.

The anterior part of the Broca area (Brodmann area 45), which receives afferent projections from the PFC, shows a decrease in participation coefficient, while the posterior part (Brodmann area 44), which has more structural connections with sensory cortices and inferior parietal cortices, shows an increase in PC.

Network nodes characterized by high degree and high participation coefficient represent the integrative hubs of connectivity networks, and are dubbed "connector hubs". Differences in centrality and connectivity structure between the two populations may result in a different distribution of connector hubs in the brain. Figure 5.4 shows the regions with simultaneously high values of degree and participation coefficient in the SZ and healthy control groups. Notable differences are observed in the parietal regions, where the superior parietal lobule represents a connector hub in control subjects, but not in patients. Importantly, the Broca area appears as prominent connector hub only in the SZ subjects.

To get an idea of the main changes due to the difference of network, we looked at the participation coefficient of the nodes. This information points out the changes of node role in the different population, i.e. when nodes get important because of its links towards the other communities. As shown on Figure 5.4, the nodes of the motor cortex of the patient group increase their participation coefficient in the inter-community connection when conversely, the parietal as well as some frontal and temporal nodes reduce the importance of their external links compared to the second groups.

In summary, the modular structure of functional connectivity is substantially reorganized in SZ patients, with prominent differences involving primary sensory and sensorimotor areas, and language related areas.

A. Controls

B. Patients
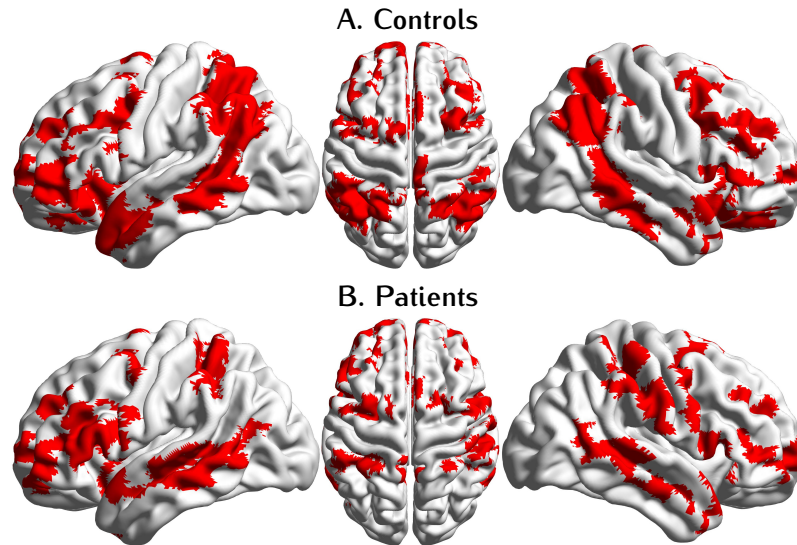
Figure 5.4: Maps of connector hubs with a participation coefficient higher than 0.6 and a degree higher than 1 for both populations.

5.3 CONCLUSIONS

In conclusion, we have applied a novel graph theoretical approach, dubbed Asymptotical Surprise, to study the structure of brain functional connectivity networks in a large cohort of Schizophrenia patients. Global and node-wise connectivity parameters show an overall reduction in connectivity in patients compared to healthy controls, in line with previous studies. The improved resolution afforded by our method reveals substantial reorganization of the modular structure of functional connectivity in patients, with a fragmentation of visual, auditory and sensorimotor cortices. This evidence is in keeping with bottom-up theories of schizophrenia that posit that cognitive dysfunction and conscious disintegration observed in SZ may arise from deficits occurring already at early stages of sensory processing. The reorganization of auditory and language modules, and the merging with multimodal association cortices is interesting in the light of the auditory hallucinations often experienced by SZ patients. Significant changes were observed in the participation coefficient of sensory, visual, and in primary auditory cortices, including the Heschl gyrus, a region critically implicated in auditory hallucinations. This evidence indicates that these regions play a different role in the integration of the network of functional connectivity in the patient's brain. Previous studies using similar, but resolution-limited methods, may have failed to detect the abnormal organization of functional connectivity at the scale reported here due to intrinsic methodological limitations. The present approach may provide a novel and powerful tool to study alterations in the brain functional organization in other neuropsychiatric conditions that are thought to be associated with aberrant connectivity.

6 | CONCLUSIONS AND FUTURE DIRECTIONS

Functional brain connectivity, measured by fMRI, greatly increased our knowledge of the architecture and function of the brain. In particular, investigation of the modular organization of brain functional connectivity networks is specifically explored in this thesis by means of a new graph theoretical approach.

In the past, these functional networks have been investigated with methods hampered by a limit that impeded the detection of modules smaller than a scale determined, in the best case, by the amount of interconnections and, in the worst case, by the overall size of the graph under consideration.

In this work, I pioneered the use of Surprise and Asymptotical Surprise, two approaches for community detection in brain networks that are demonstrably free from these severe limitations.

The methodological advantage of the approach proposed in this thesis is two-fold. Firstly, the resolution limit-free method for community detection offered a richer view on the modular structure of brain networks. Indeed, direct comparison with other methods clearly demonstrated an improved capability to identify neurofunctionally plausible and anatomically well-defined substructures otherwise concealed by the resolution limit. Moreover, the more heterogeneously distributed partitions obtained by Surprise maximization resulted in non-trivial differences of the roles of single nodes, compared to current models.

As a second point, the validation of a method to maximize Asymptotical Surprise and its application on synthetic networks endowed with a ground-truth community structure provided a comparative term for benchmark of other community detection methods. In particular, I described a procedure to artificially simulate the effects of noise on time-courses of BOLD signals, to see its detrimental effects on the performance of community detection. Altogether the picture that emerged is that Asymptotical Surprise has proved superior to existing methods in terms of Sensitivity and accuracy in detection of the planted structure as measured by Normalized Mutual Information, while showing comparable Specificity.

As the last test of the validity of this newly developed approach for community detection beyond the resolution limit, Asymptotical Surprise optimization revealed a functional fragmentation and reorganization of the primary sensory areas in schizophrenic patients. Although preliminary, this study is in keeping with bottom-up theories of schizophrenia, hypothesizing that alterations in early sensory processing may be at the heart of the cognitive and behavioural dysfunctions that characterize this tremendous disease.

As a future direction of this work, I plan to implement a multilevel method for the optimization of Asymptotical Surprise that may be of help in dealing with huge graphs as those coming from voxelwise approaches to functional connectivity. Graphs in the order of fifty thousand nodes and several million edges are the next natural step in this "big data" challenge that is pushing the frontiers

of efficient algorithms design. While PACO has been already tested on graphs with millions of edges, its performance is still poor, requiring an entire week for a single run. On the other hand, a deeper theoretical analysis of the behaviour of Surprise/Asymptotical Surprise on networks non-trivial architecture could shed light on more efficient implementations.

As the last chapter of this work confirmed, I see much space for the utilization of my new methods in the realm of the analysis of neuroimaging data for the investigation of neuropsychiatric alterations. As observed in schizophrenia patients, the greater resolution afforded by the methods presented in this work may provide novel markers for diagnosis and prognosis of neuropsychiatric disorders.

BIBLIOGRAPHY

[1] Amunts, K. *et al.* The human brain project: Creating a european research infrastructure to decode the human brain. *Neuron* **92**, 574–581 (2016).

[2] Sporns, O., Tononi, G. & Kötter, R. The human connectome: A structural description of the human brain. *PLoS Comput. Biol.* **1**, 0245–0251 (2005).

[3] Deco, G., Jirsa, V. K., Robinson, P. A., Breakspear, M. & Friston, K. The dynamic brain: From spiking neurons to neural masses and cortical fields. *PLoS Comput. Biol.* **4** (2008).

[4] Ogawa, S., Lee, T. M., Kay, A. R. & Tank, D. W. Brain magnetic resonance imaging with contrast dependent on blood oxygenation. *Proc. Natl. Acad. Sci. U. S. A.* **87**, 9868–9872 (1990).

[5] Logothetis, N. K., Pauls, J., Augath, M., Trinath, T. & Oeltermann, A. Neurophysiological investigation of the basis of the fMRI signal. *Nature* **412**, 150–157 (2001).

[6] Cordes, D. *et al.* Mapping functionally related regions of brain with functional connectivity MR imaging. *Am. J. Neuroradiol.* **21**, 1636–1644 (2000).

[7] Cordes, D. *et al.* Frequencies contributing to functional connectivity in the cerebral cortex in "resting-state" data. *Am. J. Neuroradiol.* **22**, 1326–1333 (2001).

[8] Shmuel, A. Sustained Negative BOLD, Blood Flow and Oxygen Consumption Response and Its Coupling to the Positive Response in the Human Brain. *Neuron* **36**, 1195–1210 (2002).

[9] Biswal, B., Zerrin Yetkin, F., Haughton, V. M. & Hyde, J. S. Functional connectivity in the motor cortex of resting human brain using echo-planar mri. *Magn. Reson. Med.* **34**, 537–541 (1995).

[10] Fox, P. T. & Lancaster, J. L. Opinion: Mapping context and content: the BrainMap model. *Nat. Rev. Neurosci.* **3**, 319–321 (2002).

[11] Gorgolewski, K. J. *et al.* NeuroVault.org: a web-based repository for collecting and sharing unthresholded statistical maps of the human brain. *Front. Neuroinform.* **9**, 8 (2015).

[12] Brown, J. A., Rudie, J. D., Bandrowski, A., Van Horn, J. D. & Bookheimer, S. Y. The UCLA multimodal connectivity database: a web-based platform for brain connectivity matrix sharing and analysis. *Front. Neuroinform.* **6**, 28 (2012).

[13] Honey, C. J., Kotter, R., Breakspear, M. & Sporns, O. Network structure of cerebral cortex shapes functional connectivity on multiple time scales. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 10240–10245 (2007).

[14] Hutchison, R. M. *et al.* Dynamic functional connectivity: promise, issues, and interpretations. *Neuroimage* **80**, 360–378 (2013).

[15] Raichle, M. E. *et al.* A default mode of brain function. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 676–682 (2001).

[16] Gusnard, D. A. & Raichle, M. E. Searching for a baseline: Functional imaging and the resting human brain. *Nat. Rev. Neurosci* **2**, 685–694 (2001).

[17] Fransson, P. How default is the default mode of brain function? Further evidence from intrinsic BOLD signal fluctuations. *Neuropsychologia* **44**, 2836–2845 (2006).

[18] Jonckers, E., Shah, D., Hamaide, J., Verhoye, M. & Van der Linden, A. The power of using functional fmri on small rodents to study brain pharmacology and disease. *Front. pharmacol.* **6** (2015).

[19] Fox, M. D., Corbetta, M., Snyder, A. Z., Vincent, J. L. & Raichle, M. E. Spontaneous neuronal activity distinguishes human dorsal and ventral attention systems. *Proc. Natl. Acad. Sci.* **103**, 10046–10051 (2006).

[20] Horovitz, S. G. *et al.* Decoupling of the brain's default mode network during deep sleep. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 11376–11381 (2009).

[21] Greicius, M. D. *et al.* Persistent default-mode network connectivity during light sedation. *Hum. Brain Mapp.* **29**, 839–847 (2008).

[22] Martuzzi, R., Ramani, R., Qiu, M., Rajeevan, N. & Constable, R. T. Functional connectivity and alterations in baseline brain state in humans. *Neuroimage* **49**, 823–834 (2010).

[23] Rudie, J. *et al.* Altered functional and structural brain network organization in autism. *NeuroImage Clin.* **2**, 79–94 (2013).

[24] van den Heuvel, M. & Fornito, A. Brain networks in schizophrenia. *Neuropsychology Review* **24**, 32–48 (2014).

[25] Stam, C. J. Modern network science of neurological disorders. *Nat. Rev. Neurosci.* **15**, 683–695 (2014).

[26] Fornito, A., Zalesky, A. & Breakspear, M. The connectomics of brain disorders. *Nat. Rev. Neurosci.* **16**, 159–172 (2015).

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

[28] Crossley, N. A. *et al.* The hubs of the human connectome are generally implicated in the anatomy of brain disorders. *Brain* **137**, 2382–2395 (2014).

[29] Stanley, M. L. *et al.* Defining nodes in complex brain networks. *Front. Comput. Neurosci.* **7**, 169 (2013).

[30] van den Heuvel, M. P. & Hulshoff Pol, H. E. Exploring the brain network: a review on resting-state fMRI functional connectivity. *Eur. Neuropsychopharmacol.* **20**, 519–534 (2010).

[31] Pereda, E., Quiroga, R. Q. & Bhattacharya, J. Nonlinear multivariate analysis of neurophysiological signals. *Prog. Neurobiol.* **77**, 1–37 (2005). `0510077`.

[32] Evans, A. C. Networks of anatomical covariance. *Neuroimage* **80**, 489–504 (2013).

[33] He, Y., Chen, Z. J. & Evans, A. C. Small-world anatomical networks in the human brain revealed by cortical thickness from mri. *Cereb. cortex* **17**, 2407–2419 (2007).

[34] Crossley, N. A. *et al.* Cognitive relevance of the community structure of the human brain functional coactivation network. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 11583–11588 (2013).

[35] Schwarz, A. J., Gozzi, A., Reese, T. & Bifone, A. Functional connectivity in the pharmacologically activated brain: Resolving networks of correlated responses to d-amphetamine. *Magnet. Reson. Med.* **57**, 704–713 (2007).

[36] Schwarz, A. J., Gozzi, A. & Bifone, A. Community structure and modularity in networks of correlated brain activity. *Magnetic resonance imaging* **26**, 914–920 (2008).

[37] Watts, D. J. & Strogatz, S. H. Collective dynamics of 'small-world'networks. *Nature* **393**, 440–442 (1998).

[38] Sporns, O., Tononi, G. & Edelman, G. Theoretical neuroanatomy and the connectivity of the cerebral cortex. *Behav. brain res.* **135**, 69–74 (2002).

[39] Sporns, O. & Zwi, J. D. The small world of the cerebral cortex. *Neuroinformatics* **2**, 145–162 (2004).

[40] van den Heuvel, M. P., Stam, C. J., Boersma, M. & Pol, H. H. Small-world and scale-free organization of voxel-based resting-state functional connectivity in the human brain. *Neuroimage* **43**, 528–539 (2008).

[41] van den Heuvel, M. P., Stam, C. J., Kahn, R. S. & Pol, H. E. H. Efficiency of functional brain networks and intellectual performance. *J. Neurosci.* **29**, 7619–7624 (2009).

[42] Collin, G., Sporns, O., Mandl, R. C. & van den Heuvel, M. P. Structural and functional aspects relating to cost and benefit of rich club organization in the human cerebral cortex. *Cereb. cortex* **24**, 2258–2267 (2014).

[43] Senden, M., Deco, G., de Reus, M. A., Goebel, R. & van den Heuvel, M. P. Rich club organization supports a diverse set of functional network configurations. *NeuroImage* **96**, 174 – 182 (2014).

[44] van den Heuvel, M. P. & Sporns, O. Rich-Club Organization of the Human Connectome. *J. Neurosci.* **31**, 15775–15786 (2011).

[45] Harriger, L., Van Den Heuvel, M. P. & Sporns, O. Rich club organization of macaque cerebral cortex and its role in network communication. *PloS one* **7**, e46497 (2012).

[46] de Reus, M. a. & van den Heuvel, M. P. Rich club organization and inter-module communication in the cat connectome. *J. Neurosci.* **33**, 12929–12939 (2013).

[47] Sporns, O., Honey, C. J. & Kötter, R. Identification and classification of hubs in brain networks. *PloS ONE* **2**, e1049 (2007).

[48] Power, J. D., Schlaggar, B. L., Lessov-Schlaggar, C. N. & Petersen, S. E. Evidence for hubs in human functional brain networks. *Neuron* **79**, 798–813 (2013). `NIHMS150003`.

[49] Kandel, E. R., Schwartz, J. H., Jessell, T. M. & Mack, S. (eds.). *Principles of neural science,* chap. 15, 346–347 (McGraw-Hill Medical, New York, Chicago, San Francisco, 2013).

[50] Zamora-López, G., Zhou, C. & Kurths, J. Cortical hubs form a module for multisensory integration on top of the hierarchy of cortical networks. *Front. Neuroinform.* **4**, 1 (2010).

[51] Honey, C. J. & Sporns, O. Dynamical consequences of lesions in cortical networks. *Hum. Brain Mapp.* **29**, 802–809 (2008).

[52] Ball, G. *et al.* Rich-club organization of the newborn human brain. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 7456–7461 (2014).

[53] Collin, G., Kahn, R. S., de Reus, M. A., Cahn, W. & van den Heuvel, M. P. Impaired rich club connectivity in unaffected siblings of schizophrenia patients. *Schizophr. Bull.* **40**, 438–448 (2014).

[54] Fortunato, S. Community detection in graphs. *Phys. Rep.* **486**, 75–174 (2010).

[55] Karrer, B. & Newman, M. E. J. Stochastic blockmodels and community structure in networks. *Phys. Rev. E* **83**, 1–10 (2011). `1008.3926`.

[56] Simon, H. A. The architecture of complexity. In *Facets of systems science*, 457–476 (Springer, 1991).

[57] Simon, H. A. Near decomposability and complexity: How a mind resides in a brain. In *Santa Fe Institute Studies in the science of complexity*, vol. 22, 25–25 (Addison-Wesley publishing Co., 1995).

[58] van den Heuvel, M. P. & Sporns, O. Network hubs in the human brain. *Trends Cogn. Sci.* **17**, 683–696 (2013).

[59] Meunier, D., Lambiotte, R. & Bullmore, E. T. Modular and hierarchically modular organization of brain networks. *Front. Neurosci.* **4**, 200 (2010).

[60] Mumford, D. On the computational architecture of the neocortex - II The role of cortico-cortical loops. *Biol. Cybern.* **66**, 241–251 (1992). `arXiv:1408.1149`.

[61] Tononi, G., Sporns, O. & Edelman, G. M. A measure for brain complexity: relating functional segregation and integration in the nervous system. *Proc. Natl. Acad. Sci. U. S. A.* **91**, 5033–5037 (1994). `NIHMS150003`.

[62] Tononi, G., Edelman, G. M. & Sporns, O. Complexity and coherency: Integrating information in the brain. *Trends Cogn. Sci.* **2**, 474–484 (1998).

[63] Friston, K. J. Transients, metastability, and neuronal dynamics. *Neuroimage* **5**, 164–171 (1997).

[64] Tononi, G., Sporns, O. & Edelman, G. M. Measures of degeneracy and redundancy in biological networks. *Proc.Natl.Acad.Sci.,USA* **96**, 3257–3262 (1999).

[65] Alexander-Bloch, A. F. *et al.* Disrupted modularity and local connectivity of brain functional networks in childhood-onset schizophrenia. *Front. Syst. Neurosci.* **4**, 147 (2010).

[66] Alexander-Bloch, A. *et al.* The discovery of population differences in network community structure: New methods and applications to brain functional networks in schizophrenia. *Neuroimage* **59**, 3889–3900 (2012).

[67] Meunier, D., Lambiotte, R., Fornito, A., Ersche, K. D. & Bullmore, E. T. Hierarchical modularity in human brain functional networks. *Front. Neuroinform.* **3**, 37 (2010).

[68] Sporns, O. & Betzel, R. F. Modular Brain Networks. *Annu. Rev. Psychol.* **67**, 613–640 (2016). `15334406`.

[69] Lancichinetti, A. & Fortunato, S. Community detection algorithms: a comparative analysis. *Phys. Rev. E* **80**, 056117 (2009).

[70] Fortunato, S. & Hric, D. Community detection in networks: A user guide. *Phys. Rep.* **659**, 1–44 (2016).

[71] Newman, M. E. J. Modularity and community structure in networks. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 8577–8582 (2006).

[72] Fortunato, S. & Barthélemy, M. Resolution limit in community detection. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 36–41 (2007).

[73] Newman, M. *Networks: An Introduction* (OUP Oxford, 2010).

[74] Estrada, E. *The Structure of Complex Networks: Theory and Applications* (Oxford University Press, Inc., New York, NY, USA, 2011).

[75] Cormen, T. H., Stein, C., Rivest, R. L. & Leiserson, C. E. *Introduction to Algorithms*, chap. 21 (McGraw-Hill Higher Education, 2001), 2nd edn.

[76] Wasserman, S. & Faust, K. *Social network analysis: Methods and applications*, vol. 8 (Cambridge university press, 1994).

[77] Alba, R. D. A graph-theoretic definition of a sociometric clique. *J. Math. Sociol.* **3**, 113–126 (1973).

[78] Radicchi, F., Castellano, C., Cecconi, F., Loreto, V. & Parisi, D. Defining and identifying communities in networks. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 2658–2663 (2004).

[79] Condon, A. & Karp, R. M. on the Planted Partition Model. *Electr. Eng.* 116–140 (2000).

[80] Lancichinetti, A., Fortunato, S. & Radicchi, F. Benchmark graphs for testing community detection algorithms. *Phys. Rev. E* **78**, 046110 (2008).

[81] Pons, P. & Latapy, M. Computing communities in large networks using random walks. *J. Graph Algorithms Appl.* **10**, 191–218 (2006).

[82] Rosvall, M. & Bergstrom, C. T. Maps of random walks on complex networks reveal community structure. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 1118–1123 (2008).

[83] Fox, M. D. *et al.* The human brain is intrinsically organized into dynamic, anticorrelated functional networks. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 9673–9678 (2005). NIHMS150003.

[84] De Luca, M., Beckmann, C. F., De Stefano, N., Matthews, P. M. & Smith, S. M. fMRI resting state networks define distinct modes of long-distance interactions in the human brain. *Neuroimage* **29**, 1359–1367 (2006).

[85] Salvador, R., Suckling, J., Schwarzbauer, C. & Bullmore, E. Undirected graphs of frequency-dependent functional connectivity in whole brain networks. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **360**, 937–946 (2005).

[86] Scannell, J. W., Blakemore, C. & Young, M. P. Analysis of connectivity in the cat cerebral cortex. *J. Neurosci.* **15**, 1463–1483 (1995).

[87] Felleman, D. J. & Van Essen, D. C. Distributed hierarchical processing in the primate cerebral cortex. *Cereb. Cortex* **1**, 1–47 (1991).

[88] Deco, G., Tononi, G., Boly, M. & Kringelbach, M. L. Rethinking segregation and integration: contributions of whole-brain modelling. *Nat. Rev. Neurosci.* **16**, 430–439 (2015).

[89] Guimera, R., Nunes Amaral, L. A. & R. Guimera, L. A. Functional cartography of complex metabolic networks. *Nature* **433**, 895–900 (2005).

[90] Erdös, P. & Rényi, a. On random graphs. *Publ. Math.* **6**, 290–297 (1959). 1205.2923.

[91] Guimerà, R., Sales-Pardo, M. & Amaral, L. Modularity from fluctuations in random graphs and complex networks. *Phys. Rev. E* **70**, 025101 (2004).

[92] Barabasi, A.-L. *et al.* Emergence of scaling in random networks. *Science* **286**, 509–512 (1999). 9910332.

[93] Holland, P. W., Laskey, K. B. & Leinhardt, S. Stochastic blockmodels: First steps. *Social networks* **5**, 109–137 (1983).

[94] Peixoto, T. P. Model Selection and Hypothesis Testing for Large-Scale Network Models with Overlapping Groups. *Phys. Rev. X* **5**, 011033 (2015). 1409.3059.

[95] Lancichinetti, A. & Fortunato, S. Benchmarks for testing community detection algorithms on directed and weighted graphs with overlapping communities. *Phys. Rev. E* **80**, 1–8 (2009). 0904.3940.

[96] Barrat, a., Barthélemy, M., Pastor-Satorras, R. & Vespignani, a. The architecture of complex weighted networks. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 3747–3752 (2004).

[97] Nicolini, C., Bordier, C. & Bifone, A. Community detection in weighted brain connectivity networks beyond the resolution limit. *Neuroimage* **146**, 28–39 (2017). 1609.04316.

[98] Higham, N. J. Computing a nearest symmetric positive semidefinite matrix. *Linear Algebra and its Applications* **103**, 103–118 (1988).

[99] Hagmann, P. *et al.* Mapping the structural core of human cerebral cortex. *PLoS Biol.* **6**, 1479–1493 (2008).

[100] Zachary, W. W. An Information Flow Model for Conflict and Fission in Small Groups. *J. Anthropol. Res.* **33**, 452–473 (1977).

[101] Adamic, L. A. & Glance, N. The political blogosphere and the 2004 u.s. election: Divided they blog. In *Proceedings of the 3rd International Workshop on Link Discovery*, LinkKDD '05, 36–43 (ACM, New York, NY, USA, 2005).

[102] Danon, L., Díaz-Guilera, A., Duch, J. & Arenas, A. Comparing community structure identification. *J. Stat. Mech. Theory Exp.* **09008**, 219–228 (2005).

[103] Vinh, N. X., Epps, J. & Bailey, J. Information theoretic measures for clusterings comparison: Variants, properties, normalization and correction for chance. *J. Mach. Learn. Res.* **11**, 2837–2854 (2010).

[104] Cover, T. M. & Thomas, J. A. *Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing)* (Wiley-Interscience, 2006).

[105] M., M. Comparing clusterings—an information based distance. *J. Multivar. Anal.* **98**, 873–895 (2007).

[106] Zhang, P. A revisit to evaluating accuracy of community detection using the normalized mutual information. *arXiv Prepr. arXiv1501.03844* 1–6 (2015). 1501.0384.

[107] Karrer, B., Levina, E. & Newman, M. E. J. Robustness of community structure in networks. *Phys. Rev. E* **77**, 046119 (2008).

[108] Thomas, A. C. & Blitzstein, J. K. Valued Ties Tell Fewer Lies: Why Not To Dichotomize Network Edges With Thresholds. *http://arxiv.org/abs/1101.0788* 1–36 (2011). `1101.0788`.

[109] Zahoránszky-Kőhalmi, G., Bologa, C. G. & Oprea, T. I. Impact of similarity threshold on the topology of molecular similarity networks and clustering outcomes. *J. Cheminform.* **8**, 16 (2016).

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

[111] Gallos, L. K., Makse, H. a. & Sigman, M. A small world of weak ties provides optimal global integration of self-similar modules in functional brain networks. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 2825–2830 (2012).

[112] Lee, H., Chung, M. K., Kang, H., Kim, B.-N. & Lee, D. S. Discriminative persistent homology of brain networks. In *Biomedical Imaging: From Nano to Macro, 2011 IEEE International Symposium on*, 841–844 (IEEE, 2011).

[113] Scheinost, D. *et al.* The intrinsic connectivity distribution: A novel contrast measure reflecting voxel level functional connectivity. *Neuroimage* **62**, 1510–1519 (2012). `NIHMS150003`.

[114] Callaway, D. S., Newman, M. E. J., Strogatz, S. H. & Watts, D. J. Network robustness and fragility: percolation on random graphs. *Phys. Rev. Lett.* **85**, 5468–5471 (2000).

[115] Goerdt, A. "The giant component threshold for random regular graphs with edge faults". *Theoretical Computer Science* **259**, 307 – 321 (2001).

[116] Bardella, G., Bifone, A., Gabrielli, A., Gozzi, A. & Squartini, T. Hierarchical organization of functional connectivity in the mouse brain: a complex network approach. *Sci. Rep.* **6**, 32060 (2016). `1601.07574`.

[117] Towlson, E. K., Vertes, P. E., Ahnert, S. E., Schafer, W. R. & Bullmore, E. T. The Rich Club of the C. elegans Neuronal Connectome. *J. Neurosci.* **33**, 6380–6387 (2013).

[118] Park, H.-J. & Friston, K. Structural and functional brain networks: from connections to cognition. *Science* **342**, 1238411 (2013).

[119] Sporns, O., Chialvo, D. R., Kaiser, M. & Hilgetag, C. C. Organization, development and function of complex brain networks. *Trends in cognitive sciences* **8**, 418–425 (2004).

[120] Bullmore, E. & Sporns, O. The economy of brain network organization. *Nat. Rev. Neurosci.* **13**, 336–349 (2012).

[121] Bressler, S. L. & Menon, V. Large-scale brain networks in cognition: emerging methods and principles. *Trends Cogn. Sci.* **14**, 277–290 (2010).

[122] Clune, J., Mouret, J.-B. & Lipson, H. The evolutionary origins of modularity. *Proc. Biol. Sci.* **280**, 20122863 (2013). `1207.2743v1`.

[123] Kashtan, N. & Alon, U. Spontaneous evolution of modularity and network motifs. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 13773–13778 (2005).

[124] Kashtan, N., Noor, E. & Alon, U. Varying environments can speed up evolution. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 13711–13716 (2007).

[125] Yamaguti, Y. & Tsuda, I. Mathematical modeling for evolution of heterogeneous modules in the brain. *Neural Networks* **62**, 3–10 (2015).

[126] Kitano, H. Biological robustness. *Nat. Rev. Genet.* **5**, 826–837 (2004).

[127] Richard F. Betzel & Bassett, D. S. Multi-scale brain networks. *arXiv Prepr. arXiv1308.5499* **7**, 446–451 (2016). `arXiv:1608.08828v1`.

[128] Kirschner, M. & Gerhart, J. Evolvability. *Proc. Natl. Acad. Sci.* **95**, 8420–8427 (1998).

[129] French, R. M. Catastrophic forgetting in connectionist networks. *Trends Cogn. Sci.* **3**, 128–135 (1999).

[130] Ellefsen, K. O., Mouret, J. B. & Clune, J. Neural Modularity Helps Organisms Evolve to Learn New Skills without Forgetting Old Skills. *PLoS Comput. Biol.* **11** (2015).

[131] Stam, C. J. & van Straaten, E. C. W. The organization of physiological brain networks. *Clin. Neurophysiol.* **123**, 1067–1087 (2012).

[132] Pan, R. K. & Sinha, S. Modular networks emerge from multiconstraint optimization. *Phys. Rev. E* **76** (2007). `0703033`.

[133] Doucet, G. *et al.* Brain activity at rest: a multiscale hierarchical functional organization. *J. Neurophysiol.* **105**, 2753–2763 (2011).

[134] Betzel, R. F. *et al.* The modular organization of human anatomical brain networks: Accounting for the cost of wiring. *Netw Neuroscience* (2017).

[135] Kaiser, M. & Hilgetag, C. C. Nonoptimal component placement, but short processing paths, due to long-distance projections in neural systems. *PLoS Comput Biol* **2**, e95 (2006).

[136] Bongard, J. C. Spontaneous Evolution of Structural Modularity in Robot Neural Network Controllers. *Gecco-2011 Proc. 13th Annu. Genet. Evol. Comput. Conf.* 251–258 (2011).

[137] Biswal, B. B. Resting state fMRI: A personal history. *Neuroimage* **62**, 938–944 (2012).

[138] Beckmann, C. F., DeLuca, M., Devlin, J. T. & Smith, S. M. Investigations into resting-state connectivity using independent component analysis. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **360**, 1001–1013 (2005).

[139] Damoiseaux, J. S. *et al.* Consistent resting-state networks across healthy subjects. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 13848–13853 (2006).

[140] Stam, C. J. & Reijneveld, J. C. Graph theoretical analysis of complex networks in the brain. *Nonlinear Biomed Phys* **1**, 3 (2007).

[141] Tomasi, D. & Volkow, N. D. Resting functional connectivity of language networks: characterization and reproducibility. *Mol. psychiat.* **17**, 841–854 (2012).

[142] Alnaes, D. *et al.* Attentional load modulates large-scale functional brain connectivity beyond the core attention networks. *Neuroimage* **109**, 260–272 (2015).

[143] Calhoun, V. D. *et al.* Independent Component Analysis for Brain fMRI Does Indeed Select for Maximal Independence. *PLoS One* **8**, 1–8 (2013).

[144] Lee, M. H., Smyser, C. & Shimony, J. S. Resting-State fMRI: A Review of Methods and Clinical Applications. *Am. J. Neuroradiol.* **34**, 1866–1872 (2013).

[145] Thomas, C. G., Harshman, R. A. & Menon, R. S. Noise reduction in bold-based fmri using component analysis. *Neuroimage* **17**, 1521–1537 (2002).

[146] Tohka, J. *et al.* Automatic independent component labeling for artifact removal in fMRI. *Neuroimage* **39**, 1227–1245 (2008). `NIHMS150003`.

[147] Rosazza, C., Minati, L., Ghielmetti, F., Mandelli, M. L. & Bruzzone, M. G. Functional connectivity during resting-state functional MR imaging: study of the correspondence between independent component analysis and region-of-interest-based methods. *AJNR Am J Neuroradiol* **33**, 180–187 (2012).

[148] Dehaene, S., Kerszberg, M. & Changeux, J.-P. A neuronal model of a global workspace in effortful cognitive tasks. *Proc. Natl. Acad. Sci. USA* **95**, 14529–14534 (1998).

[149] Varela, F., Lachaux, J.-P., Rodriguez, E. & Martinerie, J. The brainweb: phase synchronization and large-scale integration. *Nat. Rev. Neuro.* **2**, 229–239 (2001).

[150] Greicius, M. D., Krasnow, B., Reiss, A. L. & Menon, V. Functional connectivity in the resting brain: a network analysis of the default mode hypothesis. *Proc Natl Acad Sci USA* **100**, 253–258 (2003).

[151] Power, J. D. *et al.* Functional Network Organization of the Human Brain. *Neuron* **72**, 665–678 (2011).

[152] Golland, Y., Golland, P., Bentin, S. & Malach, R. Data-driven clustering reveals a fundamental subdivision of the human cortex into two global systems. *Neuropsychologia* **46**, 540–553 (2008).

[153] Buckner, R. L., Andrews-Hanna, J. R. & Schacter, D. L. The brain's default network: Anatomy, function, and relevance to disease. *Ann. N. Y. Acad. Sci.* **1124**, 1–38 (2008).

[154] Washington, S. D. *et al.* Dysmaturation of the Default Mode Network in Autism. *Hum. Brain Mapp.* **35**, 1284–1296 (2014).

[155] Garrity, A. G. *et al.* Aberrant "default mode" functional connectivity in schizophrenia. *Am J Psychiat* **164**, 450–457 (2007).

[156] Lee, M. H. *et al.* Clustering of resting state networks. *PLoS One* **7** (2012).

[157] Zeki, S. & Bartels, a. The autonomy of the visual systems and the modularity of conscious vision. *Philos. Trans. B* **353**, 1911–1914 (1998).

[158] Baddeley, A. Working memory: looking back and looking forward. *Nat. Rev. Neurosci.* **4**, 829–839 (2003).

[159] Meunier, D., Achard, S., Morcom, A. & Bullmore, E. Age-related changes in modular organization of human brain functional networks. *Neuroimage* **44**, 715–723 (2009).

[160] Song, J. *et al.* Age-related reorganizational changes in modularity and functional connectivity of human brain networks. *Brain connectivity* **4**, 662–676 (2014).

[161] Traag, V., Van Dooren, P. & Nesterov, Y. Narrow scope for resolution-limit-free community detection. *Phys. Rev. E* **84**, 016114 (2011).

[162] Squartini, T., Mastrandrea, R. & Garlaschelli, D. Unbiased sampling of network ensembles. *New J. Phys.* **17**, 023052 (2015).

[163] Lancichinetti, A. & Fortunato, S. Limits of modularity maximization in community detection. *Phys. Rev. E* **84**, 1–9 (2011).

[164] Kawamoto, T. & Rosvall, M. Estimating the resolution limit of the map equation in community detection. *Phys. Rev. E* **91**, 012809 (2015).

[165] Reichardt, J. & Bornholdt, S. Statistical mechanics of community detection. *Phys. Rev. E* **74**, 016110 (2006).

[166] Arenas, A., Fernández, A. & Gómez, S. Analysis of the structure of complex networks at different resolution levels. *New J. Phys.* **10**, 53039 (2008).

[167] Ronhovde, P. & Nussinov, Z. Multiresolution community detection for megascale networks by information-based replica correlations. *Phys. Rev. E* **80**, 1–18 (2009).

[168] Ronhovde, P. & Nussinov, Z. Local resolution-limit-free Potts model for community detection. *Phys. Rev. E* **81**, 1–15 (2010).

[169] Thomas Yeo, B. T. *et al.* The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *J. Neurophysiol.* **106**, 1125–1165 (2011).

[170] Hilgetag, C. C., Burns, G. A., O'Neill, M. A., Scannell, J. W. & Young, M. P. Anatomical connectivity defines the organization of clusters of cortical areas in the macaque monkey and the cat. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **355**, 91–110 (2000).

[171] Blondel, V. D., Guillaume, J.-l., Lambiotte, R. & Lefebvre, E. Fast unfolding of communities in large networks. *J. Stat. Mech. Theor. Exp.* **2008**, P10008 (2008).

[172] Raghavan, U. N., Albert, R. & Kumara, S. Near linear time algorithm to detect community structures in large-scale networks. *Phys. Rev. E* **76**, 036106 (2007).

[173] Mézard, M., Parisi, G. & Virasoro, M.-A. *Spin glass theory and beyond.* (World Scientific Publishing Co., Inc., Pergamon Press, 1990).

[174] Fosdick, B. K., Larremore, D. B., Nishimura, J. & Ugander, J. Configuring Random Graph Models with Fixed Degree Sequences. *arXiv* 1–42 (2016). 1608.00607.

[175] Kehagias, A. & Pitsoulis, L. Bad communities with high modularity. *Eur. Phys. J. B* **86**, 330 (2013).

[176] Traag, V. A., Aldecoa, R. & Delvenne, J.-C. Detecting communities using asymptotical surprise. *Phys. Rev. E* **92**, 022816 (2015).

[177] Li, X., Guo, L. & Zhao, Y. E. Tag-based social interest discovery. In *Proc. 17th int. conf. on WWW*, 675–684 (ACM, 2008).

[178] Saraç, Ö. S., Pancaldi, V., Bähler, J. & Beyer, A. Topology of functional networks predicts physical binding of proteins. *Bioinformatics* **28**, 2137–2145 (2012).

[179] Java, A., Song, X., Finin, T. & Tseng, B. Why we twitter: understanding microblogging usage and communities. In *Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis*, 56–65 (ACM, 2007).

[180] Leskovec, J., Adamic, L. A. & Huberman, B. A. The dynamics of viral marketing. *ACM T. Web* **1**, 5 (2007).

[181] Berry, J. W., Hendrickson, B., LaViolette, R. a. & Phillips, C. a. Tolerating the community detection resolution limit with edge weighting. *Phys. Rev. E* **83**, 056119 (2011).

[182] Nicolini, C. & Bifone, A. Modular structure of brain functional networks: breaking the resolution limit by surprise. *Sci.Rep.* **6**, 19250 (2016).

[183] Fair, D. A. *et al.* Functional Brain Networks Develop from a "Local to Distributed" Organization. *PLOS Comput. Biol.* **5**, 1–14 (2009).

[184] Good, B. H., de Montjoye, Y.-A. & Clauset, A. Performance of modularity maximization in practical contexts. *Phys. Rev. E* **81**, 046106 (2010).

[185] Radicchi, F., Lancichinetti, A. & Ramasco, J. J. Combinatorial approach to modularity. *Phys. Rev. E* **82**, 026102 (2010).

[186] Stanley, R. P. *Enumerative Combinatorics. Vol. 1, vol. 49 of Cambridge Studies in Advanced Mathematics* (Cambridge University Press, Cambridge, 1997).

[187] Lancichinetti, A. & Fortunato, S. Consensus clustering in complex networks. *Sci. Rep.* **2**, 336 (2012). 1203.6093.

[188] Browet, A. *Algorithms for community and role detection in networks*. Phd thesis, Universitè catholique de Louvain (2014).

[189] Feller, W. *An introduction to probability theory and its applications: volume I*, vol. 3 (John Wiley & Sons London-New York-Sydney-Toronto, 1968).

[190] Aldecoa, R. & Marín, I. Deciphering network community structure by surprise. *PLoS ONE* **6**, e24195 (2011).

[191] Arnau, V., Mars, S. & Marín, I. Iterative Cluster Analysis of Protein Interaction Data. *Bioinformatics* **21**, 364–378 (2005).

[192] Fleck, T., Kappes, A. & Wagner, D. Graph clustering with surprise: Complexity and exact solutions. In Geffert, V., Preneel, B., Rovan, B., Štuller, J. & Tjoa, A. (eds.) *SOFSEM 2014: Theory and Practice of Computer Science*, vol. 8327 of *Lecture Notes in Computer Science*, 223–234 (Springer International Publishing, 2014).

[193] Aldecoa, R. & Marín, I. Surprise maximization reveals the community structure of complex networks. *Sci. Rep.* **3**, 1060 (2013).

[194] Jiang, Y., Jia, C. & Yu, J. An efficient community detection algorithm using greedy surprise maximization. *J. Phys. A* **47**, 165101 (2014).

[195] Laird, A. R., Lancaster, J. L. & Fox, P. T. BrainMap: the social evolution of a human brain mapping database. *Neuroinformatics* **3**, 65–78 (2005).

[196] Xia, M., Wang, J. & He, Y. BrainNet Viewer: a network visualization tool for human brain connectomics. *PLoS ONE* **8**, e68910 (2013).

[197] He, Y. *et al.* Uncovering intrinsic modular organization of spontaneous brain activity in humans. *PLoS ONE* **4**, e5226 (2009).

[198] Smith, S. M. *et al.* Correspondence of the brain's functional architecture during activation and rest. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 13040–13045 (2009).

[199] Laird, A. R. *et al.* Comparison of the disparity between Talairach and MNI coordinates in functional neuroimaging data: validation of the Lancaster transform. *Neuroimage* **51**, 677–683 (2010).

[200] Lancaster, J. L. *et al.* Bias between MNI and Talairach coordinates analyzed using the ICBM-152 brain template. *Hum. Brain Mapp.* **28**, 1194–1205 (2007).

[201] Tzourio-Mazoyer, N. *et al.* Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage* **15**, 273–289 (2002).

[202] Markett, S. *et al.* Assessing the function of the fronto-parietal attention network: insights from resting-state fMRI and the attentional network test. *Hum. Brain Mapp.* **35**, 1700–1709 (2014).

[203] Roy, A. K. *et al.* Functional connectivity of the human amygdala using resting state fMRI. *Neuroimage* **45**, 614–626 (2009).

[204] Chen, A. C. & Etkin, A. Hippocampal network connectivity and activation differentiates post-traumatic stress disorder from generalized anxiety disorder. *Neuropsychopharmacol.* **38**, 1889–1898 (2013).

[205] Rubinov, M. & Sporns, O. Weight-conserving characterization of complex functional brain networks. *Neuroimage* **56**, 2068–2079 (2011).

[206] Mišić, B., Goñi, J., Betzel, R. F., Sporns, O. & McIntosh, A. R. A Network Convergence Zone in the Hippocampus. *PLoS Comput. Biol.* **10**, e1003982 (2014).

[207] Moser, M.-B. & Moser, E. I. Functional differentiation in the hippocampus. *Hippocampus* **8**, 608–619 (1998).

[208] Fanselow, M. S. & Dong, H.-W. Are the dorsal and ventral hippocampus functionally distinct structures? *Neuron* **65**, 7–19 (2010).

[209] Swanson, L. W. & Cowan, W. M. An autoradiographic study of the organization of the efferent connections of the hippocampal formation in the rat. *J. Comp. Neurol.* **172**, 49–84 (1977).

[210] Cavanna, A. E. & Trimble, M. R. The precuneus: a review of its functional anatomy and behavioural correlates. *Brain* **129**, 564–583 (2006).

[211] Palla, G., Derényi, I., Farkas, I. & Vicsek, T. Uncovering the overlapping community structure of complex networks in nature and society. *Nature* **435**, 814–818 (2005).

[212] Ahn, Y.-Y., Bagrow, J. P. & Lehmann, S. Link communities reveal multiscale complexity in networks. *Nature* **466**, 761–764 (2010).

[213] Kullback, S. & Leibler, R. A. On Information and Sufficiency. *Ann. Math. Stat.* **22**, 79–86 (1951).

[214] Bianconi, G., Darst, R. K., Iacovacci, J. & Fortunato, S. Triadic closure as a basic generating mechanism of communities in complex networks. *Phys. Rev. E* **90** (2014).

[215] Eustace, J., Wang, X. Y. & Cui, Y. Z. Community detection using local neighborhood in complex networks. *Phys. A.* **436**, 665–677 (2015).

[216] Jaccard, P. Étude comparative de la distribution florale dans une portion des alpes et des jura. *Bull. Soc. vaud. Sci. nat* **37**, 547–579 (1901).

[217] Welvaert, M., Durnez, J., Moerkerke, B., Verdoolaege, G. & Rosseel, Y. neuRosim: An R package for generating fmri data. *J Stat Soft* **44**, 1–18 (2011).

[218] Welvaert, M. & Rosseel, Y. On the definition of signal-to-noise ratio and contrast-to-noise ratio for fmri data. *PLoS ONE* **8**, e77089 (2013).

[219] Gudbjartsson, H. & Patz, S. The Rician distribution of noisy MRI data. *Magn. Reson. Med.* **34**, 910–914 (1995).

[220] Krüger, G. & Glover, G. H. Physiological noise in oxygenation-sensitive magnetic resonance imaging. *Magnet Reson Med* **46**, 631–637 (2001).

[221] Peixoto, T. P. The graph-tool python library. *figshare* (2014).

[222] Csardi, G. & Nepusz, T. The igraph software package for complex network research. *InterJournal* **Complex Systems**, 1695 (2006).

[223] Lambiotte, R. & Rosvall, M. Ranking and clustering of nodes in networks with smart teleportation. *Phys. Rev. E* **85**, 056107 (2012).

[224] Buckner, R. L. *et al.* Cortical hubs revealed by intrinsic functional connectivity: mapping, assessment of stability, and relation to Alzheimer's disease. *J Neurosci.* **29**, 1860–1873 (2009).

[225] Fornito, A., Zalesky, A. & Bullmore, E. T. Network scaling effects in graph analytic studies of human resting-state FMRI data. *Front. Syst. Neurosci.* **4**, 22 (2010).

[226] Catani, M. *et al.* The rises and falls of disconnection syndromes. *Brain* **128**, 2224–2239 (2005).

[227] Friston, K. J. & Frith, C. D. Schizophrenia: a disconnection syndrome. *Clin Neurosci* **3**, 89–97 (1995).

[228] Bullmore, E. T. *et al.* Does dysplasia cause anatomical dysconnectivity in schizophrenia? *Schizophr. Res.* **30**, 127–135 (1998).

[229] Liang, M. *et al.* Widespread functional disconnectivity in schizophrenia with resting-state functional magnetic resonance imaging. *Neuroreport* **17**, 209–213 (2006).

[230] Liu, Y. *et al.* Disrupted small-world networks in schizophrenia. *Brain* **131**, 945–961 (2008).

[231] Calhoun, V. D., Eichele, T. & Pearlson, G. Functional brain networks in schizophrenia: a review. *Front. Hum. Neurosci.* **3**, 17 (2009).

[232] Alexander-Bloch, A. F. *et al.* The anatomical distance of functional connections predicts brain network topology in health and schizophrenia. *Cereb. Cortex* **23**, 127–138 (2013).

[233] Ellison-Wright, I. & Bullmore, E. Meta-analysis of diffusion tensor imaging studies in schizophrenia. *Schizophr. Res.* **108**, 3–10 (2009).

[234] Fornito, A., Yücel, M. & Pantelis, C. Reconciling neuroimaging and neuropathological findings in schizophrenia and bipolar disorder. *Curr. Opin. Psychiatry* **22**, 312–319 (2009).

[235] Kubicki, M. *et al.* DTI and MTR abnormalities in schizophrenia: Analysis of white matter integrity. *Neuroimage* **15**, 1109–1118 (2005).

[236] Tononi, G. & Edelman, G. M. Schizophrenia and the mechanisms of conscious integration. *Brain Res. Rev.* **31**, 391–400 (2000).

[237] Hyde, T. M., Ziegler, J. C. & Weinberger, D. R. Psychiatric disturbances in metachromatic leukodystrophy. Insights into the neurobiology of psychosis. *Arch. Neurol.* **49**, 401–406 (1992).

[238] Bassett, D. S. & Bullmore, E. Small-world brain networks. *Neuroscientist* **12**, 512–523 (2006). `arXiv:1608.05665v1`.

[239] Kaiser, M. A tutorial in connectome analysis: Topological and spatial features of brain networks. *Neuroimage* **57**, 892–907 (2011). `1105.4705`.

[240] Bassett, D. S. *et al.* Hierarchical organization of human cortical networks in health and schizophrenia. *J. Neurosci.* **28**, 9239–9248 (2008).

[241] Anderson, A. & Cohen, M. S. Decreased small-world functional network connectivity and clustering across resting state networks in schizophrenia: an fMRI classification tutorial. *Front. Hum. Neurosci.* **7**, 520 (2013).

[242] Yu, Q. *et al.* Altered topological properties of functional network connectivity in schizophrenia during resting state: A small-world brain network study. *PLoS One* **6** (2011).

[243] van den Heuvel, M. P. *et al.* Abnormal rich club organization and functional brain dynamics in schizophrenia. *JAMA Psychiatry* **70**, 783–792 (2013).

[244] Lerman-Sinkoff, D. B. & Barch, D. M. Network community structure alterations in adult schizophrenia: Identification and localization of alterations. *NeuroImage Clin.* **10**, 96–106 (2016).

[245] Yu, Q. *et al.* Modular Organization of Functional Network Connectivity in Healthy Controls and Patients with Schizophrenia during the Resting State. *Front. Syst. Neurosci.* **5**, 1–16 (2012).

[246] Çetin, M. S. *et al.* Thalamus and posterior temporal lobe show greater inter-network connectivity at rest and across sensory paradigms in schizophrenia. *NeuroImage* **97**, 117 – 126 (2014).

[247] Javitt, D. C. & Sweet, R. A. Auditory dysfunction in schizophrenia: integrating clinical and basic features. *Nat. Rev. Neurosci.* **16**, 535–550 (2015). `15334406`.

A | TABLES AND ABBREVIATIONS

Region	Abbreviation
ACG	Anterior Cingulate Gyrus
CAL	Calcarine
CAU	Caudate
CUN	Cuneus
DCG	Median cingulate and paracingulate gyri
FFG	Fusiform Gyrs
HES	Heschl Gyrus
HIP	Hippocampus
IFGoperc	Inferior Frontal Gyrus, opercular part
IFGtriang	Inferior Frontal Gyrus, triangular part
INS	Insula
IOG	Inferior Occipital Gyrus
IPL	Inferior Parietal Gyrus
ITG	Inferior Temporal Gyrus
LING	Lingual Gyrus
MFG	Middle Frontal Gyrus
MOG	Middle Occipital Gyrus
MTG	Middle Temporal Gyrus
ORBinf	Inferior Frontal Gyrus, orbital part
ORBmid	Middle Frontal Gyrus, orbital part
ORBsup	Superior Frontal Gyrus, orbital part
ORBsupmed	Superior Frontal Gyrus, medial orbital
PAL	Globus Pallidum
PCG	Posterior cingulate Gyrus
PCL	Paracentra Lobule
PCUN	Precuneus
PHG	Parahippocampal Gyrus
PoCG	Postcentral Gyrus
PreCG	Precentral Gyrus
PUT	Putamen
REC	Rectus
ROL	Rolandic Operculum
SFG	Superior Frontal Gyrus
SFGmed	Superior Frontal Gyrus, medial part
SMA	Supplementary Motor Area
SMG	Supramarginal Gyrus
SOG	Superior Occipital Gyrus
SPG	Superior Parietal Gyrus
STG	Superior Temporal Gyrus

THA	Thalamus
TPOmid	Temporal Pole, Middle temporal Gyrus
TPOsup	Temporal Pole, Superior temporal Gyrus

Table A.1: Abbreviation of brain areas.

Table A.2: Modularity Coactivation

Community	Density	$|E_c|$	$|V_c|$
0	0.122457	2486	202
1	0.311761	4167	164
2	0.321541	4090	160
3	0.342825	2131	112

Table A.3: Modularity Resting state

Community	Density	$|E_c|$	$|V_c|$
0	0.154097	3545	215
1	0.109951	1674	175
2	0.489336	5323	148
3	0.671313	3323	100

Table A.4: Surprise Coactivation

Community	Density	$|E_c|$	$|V_c|$
0	0.491261	3626	122
1	0.482883	2948	111
2	0.553282	1298	69
3	0.388811	834	66
4	0.573077	447	40
5	0.458128	186	29
6	0.442029	122	24
7	0.394737	75	20
8	0.452632	86	20
9	0.397661	68	19
10	0.415205	71	19
11	0.379085	58	18
12	0.541667	65	16

Table A.5: Surprise Resting State

Community	Density	$|E_c|$	$|V_c|$
0	0.653183	4586	119
1	0.755375	3162	92
2	0.543101	2224	91
3	0.495277	367	39
4	0.560317	353	36
5	0.616756	346	34
6	0.513228	194	28
7	0.525362	145	24
8	0.601307	92	18
9	0.454545	30	12
10	0.583333	21	9
11	0.678571	19	8
12	0.607143	17	8