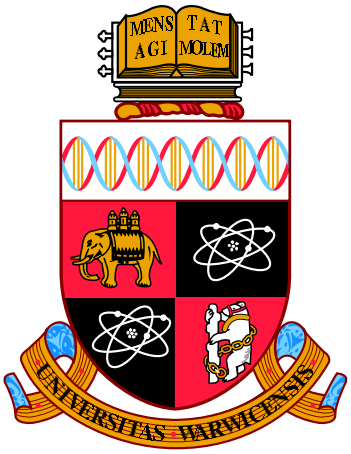


Generalised Stochastic Blockmodels and their Applications in the Analysis of Brain Networks

by

Dragana Mile Pavlović

Thesis

Submitted to the University of Warwick

for the degree of

Doctor of Philosophy

Department of Statistics

September 2015

THE UNIVERSITY OF WARWICK

Contents

List of Tables

List of Figures

Acknowledgments

I would like to express my special appreciation and gratitude to my supervisors, Prof. Thomas E. Nichols and Prof. Edward T. Bullmore, whose immense neuroimaging expertise, patience and continuous support shaped me as a researcher. On these lines, I would also like to thank my past committee members, Dr. Adam Johansen and Prof. Jane Hutton, for their continuous support and detailed feedback at various stages of this thesis project, which helped me evolve as a writer. Finally, I would like to thank Dr. Petra E. Vértes whose dedication and enthusiasm for network analysis made the worm project especially enjoyable and I am very honoured to have had the opportunity to work with her.

A very special thanks goes out to Dr. Bryan Guillaume, without whose motivation and encouragement I would have never completed this work. At the moments of greatest despair, his never-ending support and abundance of clever engineering tips have helped tremendously at various stages of my practical implementations. Also, special thanks go out to all the past and present members of the Neuroimaging Reading Group. Without their cheerful attitude, friendship and occasional homemade cake during the 'brain tea', my time at Warwick would have been a very lonely experience. I am very grateful that I got the opportunity to be a part of this research cohort and to present and discuss my work in such a lovely atmosphere where the collegiate spirit and teamwork are truly exemplary in the department.

I must also acknowledge Dr. Christine Farmer, whose persistence, understanding and kindness helped tremendously in my transition to Warwick. It was

under her tutelage that I was inspired to consider research and to be courageous to pursue my dreams. Also, I would like to thank Dr. Charlotte Price for helping me come out of my shell and do some teaching in Business school which was a very rewarding experience.

To my family (aunts, uncles, brothers and sisters and especially to Nedeljko Pavlović), thank you for the abundance of unconditional love and support during the various health scares and not so pretty life episodes. I am so sorry for putting you through all this and I promise not do it again. Also, I would like to thank Patrick Garry for his never-ending friendship. His kindness, love and support helped me in many difficult moments in my life. You will stay treasured in my heart, until we meet again some day. Rest in peace my friend.

In conclusion, I recognise that this research would not have been possible without the financial assistance of the MRC Industrial CASE award with the Glaxo-SmithKline's Clinical Unit Cambridge (UK).

To my inspirational mathematics teacher Jasna Filipović (IXth Belgrade Gymnasium) and to Ljubica Crnjanska

Declarations

I hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification in these, or any other Universities. This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration, except where specifically indicated in the text.

- The work in Chapter 3 has been published in Plos One (Pavlović et al., 2014).
- A large part work in Chapter 4 is based on a manuscript which is currently under a revision for a resubmission.
- The work in Chapter 5 will shortly be submitted for publication.

Dragana Pavlović 2015

Abstract

Recently, there has been a great interest in methods that can decompose brain networks into clusters with similar connection patterns. However, most of the currently used clustering methods in neuroimaging are based on the stringent assumption that the cluster structure is modular, that is, the nodes are densely connected within clusters, but sparsely connected between clusters. Furthermore, multi-subject network data is typically fit by several subject-by-subject analyses, which are limited by the fact that there is no obvious way to combine the results for group comparisons, or on a group-averaged network analysis, which does not reflect the variability between subjects. In the first part of this thesis, we consider the analysis of a single binary-valued brain network using the Stochastic Blockmodel (Daudin et al., 2008) and compare it to the widely used clustering methods, Louvain and Spectral algorithms. For this, we use the *Caenorhabditis elegans* (*C. elegans*) worm nervous system as a model organism whose wealth of prior biological knowledge can be used to validate the functional relevance of network decompositions. We show that the 'cores-in-modules' decomposition of the worm brain network estimated by the Stochastic Blockmodel is more compatible with prior biological knowledge about the *C. elegans* than the purely modular decompositions found by the Louvain and Spectral algorithms. In the second part of this thesis, we propose three multi-subject extensions of Daudin et al.'s Stochastic Blockmodel that can estimate a common cluster structure across subjects. Two of these (non-trivial) models use subject-specific covariates to model variation in connection rates in the data. The first and trivial model assumes no variability between subjects, the second model accounts for a global variability in connections between subjects, and the third model accounts for local variability in connections between subjects that can differ across individual within/between-cluster connectivity elements. In the third part of this thesis, we propose a mixed-effect multi-subject model which can account for the repeated-measures aspects of multi-subject network data by including a random intercept. For the second and third part of the thesis, we use intensive Monte Carlo simulations to investigate the accuracy of the estimated parameters as well as the validity of inference procedures. Furthermore, we illustrate the proposed models on a resting state fMRI dataset with two groups of subjects: healthy volunteers and individuals diagnosed with schizophrenia.

CHAPTER 1

Introduction

The past thirty years have seen increasing interest in the systems-level understanding of structures as diverse as the global economy (Dicken et al., 2002), ecosystems (Ulanowicz, 2004), living cells (Ravasz et al., 2002), power grids (Albert et al., 2004) and more. To obtain deeper insights into the operational mechanisms governing these systems, the scientific focus has gradually moved away from the analysis of their isolated components to the ways in which these components interact to perform the functions that characterise the system as a whole. In this manner, a wide range of systems can all be studied as networks, defined through their elements (nodes) and the connections (edges) that link them. A system's functional properties can then be studied in terms of the connection structure that is associated with its network.

In the context of the brain, the same approach can be used to study how simple neuronal elements (e.g., neurons) are organised into circuits to process information. This allows us to gain greater insights than the study of a single and isolated element would normally provide. For example, individual neurons can engage in complex physiological responses that are triggered by interactions between a larger number of neurons locked in networks and, thus, the knowledge of such networks may provide a better understanding of brain activity (Sporns, 2010). Indeed, characterising brain activity in terms of networks has opened up the neuroimaging community to new research avenues and opportunities to make enquiries about how a coordinated activity between various neuronal elements gives rise to a specific behaviour. In neuroscience, this has marked a shift of scientific focus from a classical brain mapping (or 'functional segregation') approach to a more modern connectomics (or 'functional integration') approach.

In general, classical neuroscience describes the human brain as an organ whose many parts specialise in various functions by actively engaging in local processes. In the conventional sense, such a view of the brain's infrastructure is known as 'functional segregation'. The concept of functional segregation was an important milestone in classical neuroscience because it has provided a framework in which it was possible to classify parts of the brain by only considering if they acted as independent mechanisms for completing a certain cognitive function. However, although this approach has proved to be a successful theoretical framework for describing regional functions of the human brain, it has been shown to be inadequate for explaining higher cognitive processes like visual recognition, social cognition, emotions or others (see, e.g., Van Den Heuvel and Sporns, 2013, for a review). Indeed, such cognitive processes are not necessarily linked to a single anatomical brain region, but typically to several brain regions, demonstrating that the brain, for these functions, actively engages in processes at multiple spatial locations (Tononi et al., 1998). Thus, in order to study such large scale aspects of the brain's functional organisation, modern neuroscience has adopted a view of 'functional integration', in which the brain is defined as a complex network of neuronal elements (nodes) whose connections (edges) facilitate different functional processes. This alternative view of the brain helps to represent two important functional aspects of the brain. First, long distance connections between different brain regions facilitate complex functions. Second, some of these brain regions and their circuits may be involved in several functions and, thus, act as important focal points in the brain. These two aspects clearly demonstrate the advantage of functional integration compared to functional segregation and explain the growing interest of representing the brain as a network.

Nevertheless, brain networks are expected to have a complex topology which can adapt to a variety of cognitive functions and, thus, their analysis is generally non-trivial. In recent years, however, different empirical studies (Bullmore and Sporns, 2009; Sporns, 2010; Bullmore and Bassett, 2011; Chen et al., 2006; Towlson et al., 2013), conducted on both animals and humans, have provided growing evidence that the organisational principles governing brain networks conform to a common architectural standard. This suggests that brain networks can be decomposed into groups of nodes which share similar connectivity profiles with regards to both their internal (within-group) and external (between-group) connections. Thus, instead of trying to analyse a brain network directly, we can try to decompose it into groups of nodes with similar properties (i.e. blocks), which, in turn, yield a more interpretable view of the network. This compressed view of a network is generally referred to as a cluster or block structure.

However, the classification of nodes into blocks is generally unknown and,

thus, we require methods or models which can estimate this latent structure. So far, most network studies have used deterministic methods which seek for a decomposition that maximises a goodness of fit measure called modularity (Newman, 2006). In contrast, in this thesis, we consider stochastic network models which use likelihood-based methods to estimate the network's block structure. One example of such network models is the Stochastic Blockmodel of Snijders and Nowicki (Snijders and Nowicki, 1997). This model uses a framework of finite mixture models to explicitly account for different sources of variation in the distribution of edges which give rise to the network's cluster structure. In particular, Snijders and Nowicki (1997) investigated different optimisation strategies, including the Expectation Maximisation (EM) and Markov Chain Monte Carlo (MCMC) algorithms, which can be used to estimate the latent node assignments and model's parameters. Their results showcased some computational challenges encountered in the optimisation of the model likelihood and pointed out that these optimisations techniques are appropriate only for small networks (e.g., < 100 nodes). More recently, Daudin, Picard and Robin (Daudin et al., 2008) revisited the Stochastic Blockmodel of Snijders and Nowicki, for which they considered a 'variational approximation' as an optimisation strategy. These authors also suggested the use of the Integrated Classification Likelihood (ICL) criterion (Biernacki et al., 1998) as a way to estimate the optimal number of blocks. Their work also touched upon a plethora of interesting questions related to general issues in network clustering. For example, they showed (a) that their model can be used to detect a wider range of different types of clustering structures, including modular, disassortative and star-patterned structures, (b) how the network's empirical degree distribution can be approximated with a mixture of Poisson densities and (c) how the model parameters can be used to approximate the empirical clustering coefficient proposed by Newman et al. (2002). Throughout this thesis, we will explicitly use the SBM to note the Stochastic Blockmodel of Snijders and Nowicki (Snijders and Nowicki, 1997) with implementation proposed by Daudin, Picard and Robin (Daudin et al., 2008).

In this work, we make a number of contributions to the network analysis of brain data.

1. Motivated by a lack of comparative studies in the literature, we use the binary and undirected brain network of the *C. elegans* worm[1] to benchmark the SBM against popular and widely used deterministic clustering algorithms, such as the Fast Louvain and Spectral algorithms. In a detailed assessment, we show that the 'cores-in-modules' decomposition estimated by the SBM is more biologically plausible than the purely modular decompositions defined

[1] *C.elegans* is a simple model organism with a large body of results related to its anatomy and behaviour.

deterministically (Pavlović et al., 2014).

2. Regardless of the clustering methods or models, most of the existing multi-subject work in human neuroimaging studies is based on a subject-by-subject or a group-averaged network analysis. The main limitation of per-subject models is that there is no obvious way to combine the cluster results for group comparisons, and of group-averaged models that they do not reflect the variability between subjects. In Chapter 4, we propose three extensions of the classical SBM. To our knowledge, two of these (non-trivial) models are the only multi-subject clustering approaches which can capture between subject variability and, at the same time, estimate a common network decomposition (cluster structure).

3. Our proposed methods use a regression framework and, thus, allow inference on group differences and covariate effects. We are unaware of any comparable statistical analysis in the literature; in contrast, the usual approach is to submit network summary metrics to univariate statistical tests.

4. We use the Firth Maximum Likelihood Estimate (MLE) for good performance with small numbers of subjects, and use punishing (low number of nodes n, highly imbalanced block sizes) simulations to evaluate our methods. The methods in Chapter 4 found good performance with 10 or more subjects, thus, demonstrating the utility for the typical analyses with small samples.

5. We propose a mixed effects multi-subject model which uses a random intercept for each subject's block. The advantage of this model is that it can account for within subject correlation. This correlation may be present if the covariates don't completely explain the between-subject variation in connection rates, or if multiple networks per subject are used. Again, we are unaware of any comparable analysis capability in published methods.

This dissertation is organised as follows. In Chapter 2, we review some background information related to the analysis of functional connectivity. In particular, we describe how functional connectivity can be measured using functional Magnetic Resonance Imaging (fMRI) and how its pre-processed data can be used to construct binary and undirected networks. In the remaining part of Chapter 2, we review the existing clustering models and methods that are relevant for this thesis. In Chapter 3, we present the results of the SBM fit to the *C. elegans* data and compare it against the fits of the Fast Louvain and Spectral algorithms. In Chapter 4, we describe three multi-subjects SBMs and we discuss their optimisation strategies and their inference procedures. We use intensive Monte Carlo simulations to investigate the accuracy of the estimated parameters, as well as the validity of inference procedures based on

the Ordinary and Firth MLEs. Furthermore, we illustrate the multi-subject SBM with heterogeneous effects on a resting state fMRI dataset with two groups of subjects: healthy volunteers (Controls) and individuals diagnosed with schizophrenia (Patients). In Chapter 5, we describe the multi-subject SBM with heterogeneous fixed and per-subject random effects. We use Monte Carlo simulations to investigate the accuracy of the estimated parameters and illustrate the model on a resting state fMRI data sets considered in Chapter 4. In the final chapter (Chapter 6), we give conclusions and directions of future work.

CHAPTER 2

Background

One of the major sources of brain's complexity can be attributed to the intricate patterns of connectivity which facilitate complex physiological and cognitive processes between various neuronal elements. For brains at higher evolutionary stages, such patterns of connectivity can be segregated into three different scales of connectivity, a microscale level (links are formed at the levels of neurons), a mesoscale level (links are formed at the levels of neuronal populations), and a macroscale level (links are formed at the levels of brain regions). At each scale level, it is also possible to make further classifications because the connectivity measures can be derived from anatomical observations or from observations related to physiological processes that somewhat indirectly measure an ongoing functional activity. In that regard, connectivity can be distinguished along three major classes, referred to as *structural*, *functional* and *effective* connectivity. These three types of connectivity are described next.

Structural connectivity generally denotes a network of anatomical connections between neuronal elements. In the case of the human brain, structural connections are mediated by the part of brain known as white matter. White matter is mostly made of myelinated axons[1] that make connections between neuron cells and, thus, allow communications between different regions of the brain. A bundle of white matter fibres typically contains between 10^3 and 10^5 compressed axons which tend to follow a common trajectory. Structural human brain networks are generally estimated using Diffusion Tensor Imaging (DTI) data and they are typically weighted and undirected because the directionality of axonal projections cannot, at present,

[1] An axon is a projection of a nerve cell, encased in a sheet of lipid called myelin, which improves the transmission of electric impulses.

be determined (Hagmann et al., 2007; Zalesky and Fornito, 2009; Iturria-Medina et al., 2007).

Functional connectivity, according to Friston et al. (1993), is defined as the temporal correlations in functional measurements between anatomically separated brain regions. In this context, functional measurements generally entail time series data that can be collected from a variety of neuroimaging techniques, including Electroencephalography (EEG), Magnetoencephalography (MEG) and functional Magnetic Resonance Imaging (fMRI), within either a framework of task or resting state studies. Functional connectivity is represented by correlations between time series and, as such, does not bear information about directionality or causality. It may be represented by a binary or weighted network, but it is always undirected.

Effective connectivity, according to Friston et al. (1994), is defined as a network of causes which prompts one neuronal system to induce an effect in another. Effective connectivity is generally estimated from time series data (e.g., EEG, MEG, fMRI), either from task or resting state experiments. In particular, this approach necessitates a generative statistical model (Costa et al., 2015) which estimates the direction of edges by taking into account the time courses at multiple nodes.

Although effective connectivity holds a promise of new insights into the organisational principles of brain networks, it is important to acknowledge that the majority of brain networks studies are carried out on functional connectivity data sets, and hence we will mainly focus on this type of connectivity in this thesis. In the continuation of this chapter, we give a general overview of the data generating pipeline which begins with the acquisitions and preprocessing of fMRI images and finishes with the construction of binary and undirected functional brain networks. We then review some popular methods for network clustering, focusing especially on the class of statistical models known as the Stochastic Blockmodels (SBMs).

2.1 Measuring Functional Connectivity with fMRI

In this section, we review functional Magnetic Resonance Imaging (fMRI), currently the most popular non-invasive neuroimaging modality for estimating functional connectivity in the human brain. We begin by briefly detailing the fundamentals of fMRI and then discuss some of the common preprocessing steps used on its data. Finally, we describe how fMRI data is used to estimate brain networks.

2.1.1 Magnetic Resonance Imaging

Magnetic Resonance Imaging (MRI) is based on the magnetic properties of certain atomic species that have a magnetic moment. Specifically, hydrogen atoms, prevalent in the body through water (H_2O), have a magnetic moment and respond to

the presence of a magnetic field. MRI uses a very strong but static magnetic field (the bulk of the machine) and weaker but dynamic magnetic fields (created by thin coils surrounding the subject). Careful manipulation of the dynamic magnetic fields allows the creation of images which reveal the properties of tissue. MRI can create images sensitive to different features of brain tissue, and in particular can produce detailed images of brain anatomy. Such images are the mainstay of clinical MRI use, and are referred to as anatomical or structural images, to distinguish them from 'functional' images. In Figure 2.1, we contrast respectively the static and dynamic nature of structural and functional MRI images.

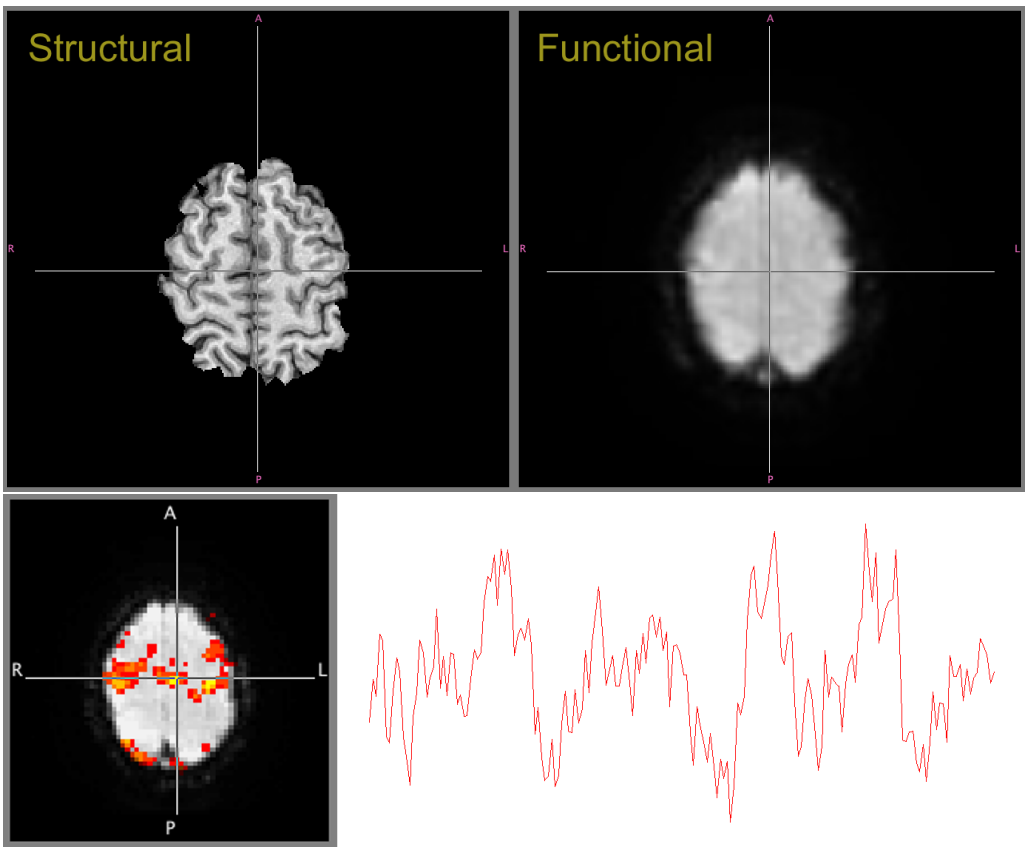


Figure 2.1: **Comparison between a structural and a functional MRI image**. Top left, a structural MRI image. Top right, a functional MRI image at a particular time. Bottom left, thresholded statistical image with the functional image as background. Bottom right, functional MRI signal at the selected voxel located in the supplementary motor area. The structural image is static while the functional image is dynamic.

2.1.2 Functional Magnetic Resonance Imaging (fMRI)

Functional Magnetic Resonance imaging or fMRI is an indirect measure of brain activity. It is based on the local changes in concentrations of oxygenated and deoxygenated haemoglobin in the brain. In particular, when neuron cells are active, they fire electrical impulses which lead to an increase in their demand for oxygen, after which there is an increase in blood flow. A signal is generated by the different magnetic properties associated with oxygenated/deoxygenated haemoglobin: oxygenated blood has a negative magnetic susceptibility (diamagnetic) and deoxygenated blood has a positive magnetic susceptibility (paramagnetic). As the bulk of the brain tissue is diamagnetic, the presence of paramagnetic deoxygenated blood creates an inhomogeneity that changes the MRI intensity. To indicate this physiological dependence, the fMRI signal is generated as the Blood-Oxygen-Level Dependent (BOLD) signal (Ogawa et al., 1990).

2.1.3 Preprocessing

Raw fMRI data take the form of a time series of three-dimensional images or, equivalently, a times series at each image voxel[2]. Before useful information can be extracted from the raw data, several preprocessing steps are needed to eliminate sources of immense nuisance variation that would otherwise swamp the subtle BOLD signal. In the rest of this section, we briefly review standard pre-processing steps, including slice timing correction, spatial realignment, spatial normalisation and spatial smoothing. We provide only a cursory review of each area and provide references for further details.

Slice Timing Correction

The acquisition of each 3D fMRI image is generally carried out as a sequence of individual 2D slice-scans rather than one single 3D snapshot of the brain at a given time point. Thus, for each time point, it usually takes up to two seconds ($t = 2$) to completely sample the brain. As a results of this, some misalignments between the individual acquisition times of individual scans are present. This effect is generally alleviated with an algorithm that shifts each scan relative to the acquisition time of the initial scan. See Henson et al. (1999) for further details.

Spatial Realignment

When lying in the MRI scanner, it is very difficult for a subject to stay perfectly still and head motion is likely to occur. Yet any head motion contributes to misalign-

[2] A voxel is the smallest graphical unit upon which a three-dimensional image is built. By analogy with a two-dimensional image, a voxel is a three-dimensional generalisation of a pixel.

ments of the images over time. This type of motion is known as the bulk-motion and it can be induced by different factors. For example, a subject can move more as a result of fatigue or as a result of simple actions like swallowing or blinking. Bulk-motion is typically corrected with a rigid body realignment procedure (Ashburner and Friston, 2007), which assumes that the head motion can be described using translation and rotation about three axes. The motion relative to a reference scan (e.g., the first scan) is estimated and a spatial transformation applied to correct the motion.

Spatial Normalisation

In a multi-subject study, each subject's brain must be transformed to a common space or to an atlas space. This is called 'spatial normalisation' and its goal is to remove natural anatomical differences between individual brains (e.g., size and shape). Spatial normalisation proceeds by finding a transformation (linear or non-linear) that minimises some measure of difference between the individual images and the image of reference (Friston et al., 1995). The transformation is generally estimated on the structural images (which have fine anatomical detail) and then applied to the functional images (which have poor anatomical detail).

Spatial Smoothing

In fMRI preprocessing, spatial smoothing consists of convolving each image with a three-dimensional Gaussian kernel. In this process, high frequency information is removed from the images while reducing variance. This trade-off is generally acceptable as the spatial scale of expected BOLD effects is generally larger than the individual voxels. Spatial smoothing after spatial normalisation also discounts any residual misregistrations. Finally, some researchers use spatial smoothing in order to make the data more compliant with the assumptions of the statistical analysis.

Additional Steps

Once the standard preprocessing steps are completed, some confounding effects due to physiological noise (including cardiac, respiratory or head motion activity) may still exist in the data. One of the most effective ways of dealing with this is to regress such information out of the time series (Varoquaux and Craddock, 2013). For head motion, this can be done by regressing out the parameters obtained from the rigid body realignment (i.e. the six time series of the translations and rotations). The effects of cardiac and respiratory activity are, however, generally harder to control for and, hence, these components are not regressed out directly, but instead they

are accounted for by regressing out the mean signal measured in white matter[3] and in cerebrospinal fluid[4] (Varoquaux and Craddock, 2013).

In addition, some researchers also apply a temporal high-pass filtering. The high-pass filtering keeps the signal with a frequency greater than a certain cutoff frequency which usually corresponds to the lowest frequency of interest (0.1Hz). This accounts for the notable low-frequency 'drift' effects that occur in fMRI data. In addition to the high-pass filtering, some authors (e.g., Lynall et al., 2010; Bullmore et al., 2003; Bassett et al., 2006) use a discrete wavelet transform of the time series data, which decomposes the data over a set of frequency scales. One or more of these scales is then used as the signal of interest while the remaining scales are assumed to capture mainly confounding signals or noise variance.

2.1.4 Defining Connectivity with fMRI

Once the fMRI data is acquired and preprocessed, the estimation and construction of a functional network is carried out next. This mainly entails two tasks: the formulation of the nodes and the estimation of the functional connections (or edges) between the nodes.

For the first task, we could consider each voxel of the preprocessed fMRI data as a node. However, as the number of voxels typically varies in the order of 200,000, we would have a very large network and, consequently, a very limited number of methodological choices for its analysis. To keep the data in a more manageable format, it is customary to aggregate the voxels into regions of interests (ROIs) which define the nodes of the brain network. This can be generally achieved by the means of brain atlases or data driven approaches.

Brain atlases are typically based on brain anatomy, either from a single subject like the Automatic Anatomic Labelling (AAL) (Tzourio-Mazoyer et al., 2002) and the Talairach Tournoux (TT) atlases, or from a group of subjects like the Harvard-Oxford (HO) (Smith et al., 2004) atlas. A limitation of these atlases is that they are based on the brain structure seen in anatomical MRI images while the connectivity of interest is based on brain function revealed by the BOLD signal. To bypass these inconsistencies, brain atlases can alternatively be based on functional data alone. Thus, using a parcellation method on the BOLD data, it is possible to estimate contiguous regions of maximal homogeneity (Craddock et al., 2012; Thirion et al., 2006; Yeo et al., 2011) which define ROIs. Examples of two such atlases are the Craddock 200 (CC200) and the Craddock 400 (CC400) (Craddock et al., 2012).

Another important issue to consider when selecting atlases is their number

[3]White matter is a part of the brain which consists mostly of axonal projections from grey matter cells.

[4]Cerebrospinal fluid is a clear body fluid found in the centre of the brain.

of ROIs. The HO atlas has fewer than 100 regions and, therefore, it is reasonable to expect that some of its ROIs may not be representative of a single function but of several. As shown by Smith et al. (2011) such overly-large ROIs may lead to incorrect network estimates. In contrast, this issue is circumvented in the CC200 and CC400 atlases whose larger number of ROIs tend to provide a more accurate representation of brain functions than the ROIs from an anatomical atlas (see Figure 2.2 for images of the HO and CC200 atlas).

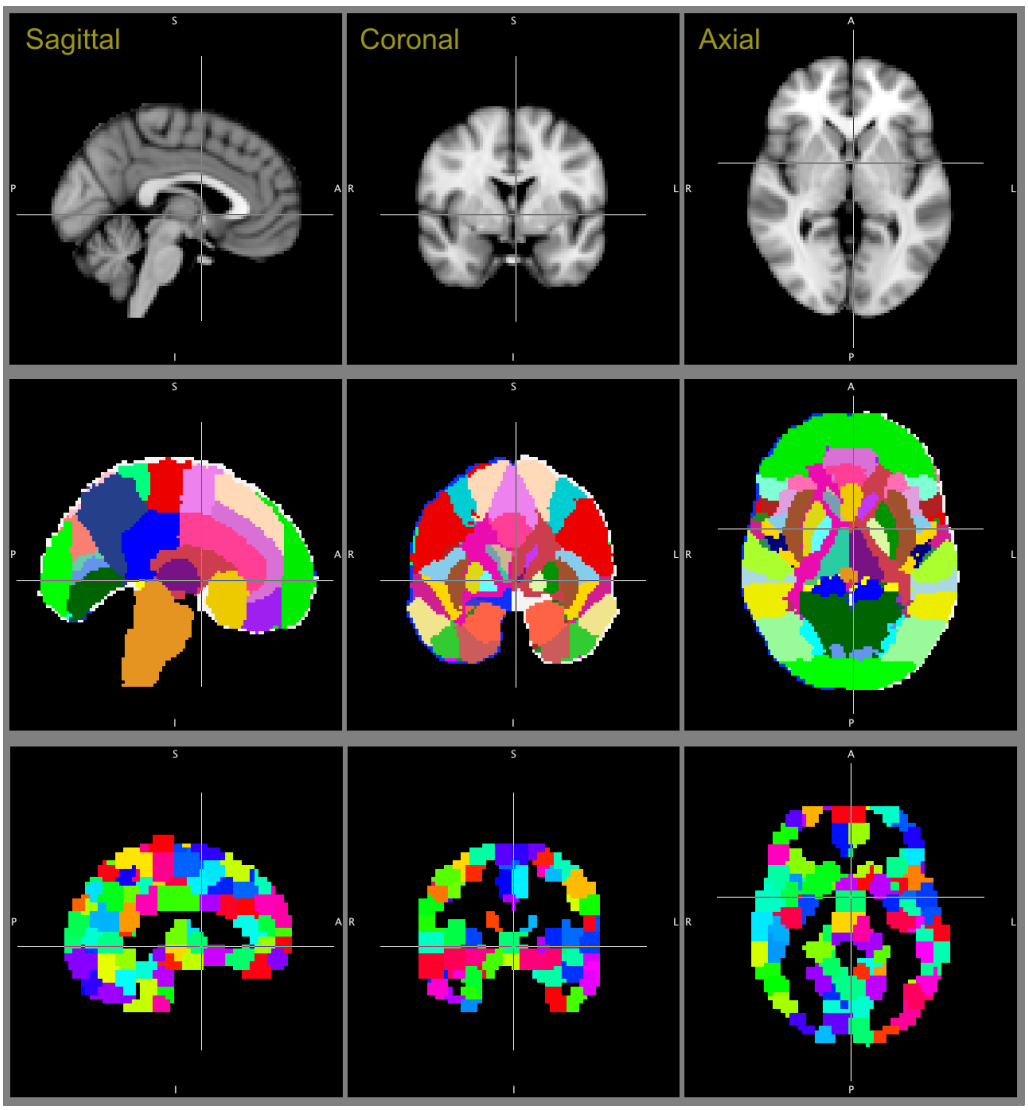


Figure 2.2: **Brain parcellation using different atlases**. First row, the anatomical image showing the Sagittal (y, z), Coronal (x, z) and Axial (x, y) view of the standard MNI brain template. Second row, the whole brain parcellation using the Harvard-Oxford atlas with 69 cortical and subcortical regions. Third row, the whole brain parcellation using a data driven approach (Craddock et al., 2012) with 400 cortical and subcortical ROIs.

Once the brain is parcellated into anatomically -or functionally- defined ROIs these regions are then considered as the nodes of the brain network. Each node is represented by a mean time series, that is, the mean over the voxels in each ROI for each time point. If there are n ROIs, we compute the Pearson correlations r_{ij} for each ROI pair, $1 \leq i \neq j \leq n$, and transformed them via Fisher's formula (Fisher, 1915)

$$z_{ij} = \frac{1}{2} \log \left[\frac{1 + r_{ij}}{1 - r_{ij}} \right]. \tag{2.1}$$

Under the null hypothesis that the population correlation is zero, Z_{ij} approximatively follows a Normal distribution with zero mean and variance $1/(N-3)$, where N is the number of time points in the time series. However, if the time series were previously filtered, the variance estimate needs to be adjusted for the loss of degrees of freedom induced by the filtering. A relatively recent result of Davey et al. (2013) suggests to use an adjusted variance estimate that is equal to $1/(2t(f_h - f_l)N)$, where t is the sampling time for the acquisition (e.g., $t = 2$ seconds) and $f_h - f_l$ is the difference between the high cut-off frequency (f_h) and the low cut-off frequency (f_l) of the filter used.

As noted in Section 2.1.3, it is also possible to use a discrete wavelet transformation on the regional mean time series, and hence represent the data in terms of different frequency scales. According to Achard and Bullmore (2007), the motivation for the use of wavelets is twofold. First, some prior practical evidence suggests that resting state fMRI time series often have non-stationary features, such as a slowly decaying positive autocorrelation, which may affect the estimation of simple cross-correlations between time series. Contrary to this, wavelet-based estimators are generally able to adapt to potential non-stationary features in the time series and, hence, seems to be more appropriate when estimating the dependance between time series (Bullmore et al., 2003). Second, wavelet-based connectivity analysis allows researchers to focus on the frequencies less than 0.1Hz, which are known to be subtended by the most salient functional connectivities.

Similarly to the example of Pearson correlations, with wavelet analysis, we can compute correlations between each ROI pair and we can use Fisher's transform formula (Eq. (2.1)) but with the caveat that N represents the total number of wavelet coefficients in the frequency band rather than the original number of time series points. Nevertheless, in both cases, z_{ij} can be used to test for the presence or absence of an edge between each node. Since positive dependence is the predominant and expected type of dependence, a one-sided test is typically used. The raw z_{ij} data can directly be used to construct a weighted network, or they can be binarised, yielding a binary network. For the latter, a $n \times n$ binary connection matrix is formed

based on significant (1) or non-significant (0) tests at each edge.

When constructing binary connection matrices, one of the most important decisions to make is how to set the critical significance threshold. In the literature, it is common to set this threshold in a data-dependent fashion and select a value that produces a given proportion (e.g., 10%) of significant edges (e.g., Mørup et al., 2010; Zalesky et al., 2012; Van Wijk et al., 2010; Ginestet et al., 2011). Alternatively, a multiple comparisons based threshold can usually be obtained by using either a Bonferroni (e.g., Cao et al., 2014) or an False Discovery Rate method (e.g., Achard et al., 2006).

2.2 Formulation and Properties of a Brain Network

As discussed in Section 2.1.4, a network is defined in terms of its nodes and edges. In a mathematical sense, a network is denoted as $\mathcal{G} = \mathcal{G}(V, E)$, where V is a set of n nodes and E is a set of m edges. The connectivity structure can be coded as an $n \times n$ adjacency (or connectivity) matrix $\boldsymbol{x} = ((x_{ij}))_{1 \leq i,j \leq n}$, whose element x_{ij} quantifies a relationship between the nodes V_i and V_j, $i, j = 1, \ldots, n$. In this thesis, we consider networks with no self connected nodes, that is, $x_{ii} = 0$ for all $i = j$. Such networks can be directed or undirected, and weighted or binary. For example, a network based on inter ROI correlations can be said to be weighted and undirected, but after thresholding, it is binary and undirected.

For the simplicity of discussion, hereafter, we will only consider binary and undirected networks, represented by a $n \times n$ connectivity matrix $\boldsymbol{x}$ such that $x_{ij} = 1$ if there is an edge between nodes V_i and V_j and $x_{ij} = 0$ otherwise. The network is considered to be undirected, that is, $x_{ij} = x_{ji}$.

In the literature, there are many ways in which a given network can be summarised, and these summary metrics are generically referred to as network statistics. In this section, we review some of the most crucial network statistics used in the analysis of brain networks, including *degree*, *degree sequence*, *clustering coefficient* and *cluster structure* (for a full review, see Rubinov and Sporns, 2010). The simplest way in which we can summarise each node V_i is to consider the number of edges which terminate at that particular node. This summary is known as the *degree* of a node $\rho(\cdot)$ and it is a rough indicator of the way a particular node is embedded in a network. Thus, the degree of node V_i is defined as

$$\rho(V_i) = \sum_{j=1}^{n} x_{ij}. \tag{2.2}$$

The *degree sequence* of a network is the set of all observed degrees in this network. It is a global description of a network and it is generally not sensitive to mild

perturbations of edge structure in a network. For instance, it is always possible to construct two slightly different networks which have the same degree sequence. This window of variation often serves as a basis for randomisation strategies whose end goal is to find the distribution of some other network summary statistic (Maslov et al., 2004; Maslov and Sneppen, 2002).

Another interesting property of brain networks is the heterogeneity of edges, which suggests that the edges tend to be more concentrated in some areas of the network than in others. This can be quantified with the *clustering coefficient* C. Newman et al. (2002) defines C as the prevalence of fully connected triplets of nodes among the set of triplets that have at least two connections,

$$C = \frac{3\#\triangle}{\#\vee} = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n} x_{ij}x_{jk}x_{ik}}{2\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k<j}^{n} x_{ij}x_{ik}}. \tag{2.3}$$

A much more difficult task is to find a complete segmentation of network's nodes into clusters. This is known as a *cluster structure*. Determining an optimal cluster structure in a given network is a procedure which generally entails a good working definition of the cluster structure, the ability to score individual fits, the ability to determine the optimal number of clusters Q and, finally, it entails having a fast optimisation algorithm which can work for networks with a relatively large number of nodes.

In a broad sense, clustering approaches can be classified into two groups: deterministic approaches based on heuristic objective functions and model-based approaches that relate the observable data to unobservable parameters of interest with a statistical model. In sections 2.3 and 2.4, we review the principle representatives of each group.

2.3 Deterministic Network Clustering Algorithms

The problem of network clustering has been actively investigated in a variety of academic fields, including computer science, statistical physics (Newman, 2003; Newman and Girvan, 2004) and data mining (Wei et al., 2006; Beverly and Sollins, 2008; Moghaddam et al., 2010). This research has generated a large body of clustering algorithms that can be broadly categorised as 'deterministic' because no randomness is involved in the formation of the observed networks.

In his extensive exposition of the topic, Fortunato (2010) classifies the clustering algorithms according to how they define the cluster structure in a network. He notes that the most popular approach is to assume that a network is a set of 'fairly independent clusters, each of which plays a similar role like, e.g., the tissues or the organs in the human body'. According to this view, the clusters are

selected to match the criteria of maximising internal (or within-cluster) connection count and minimising external (or between-cluster) connection count. Each such network decomposition is evaluated with a quality function, whose most frequently used version is modularity (Newman and Girvan, 2004).

The idea behind this concept of modularity is that the clusters are found by a comparison between the observed number of connections in the clusters and the number of connections one would expect to find if the nodes were connected at random. This randomised version of the original network, which acts as a reference point in the modularity function, is called a null model and is typically tailored to preserve some attractive features of the original network like the total number of edges or the degree distribution (Luczak, 1989; Molloy and Reed, 1995; Pattison and Robins, 2007). The modularity score takes the form of the difference between the edge count (0 or 1) and its expected value under the null model considered, summed over all intra-cluster edges and normalised:

$$f_{mod} = \frac{1}{2m} \sum_{i=1}^{n} \sum_{j=1}^{n} (x_{ij} - \mathbb{E}_0[X_{ij}]) \delta(c_i, c_j), \tag{2.4}$$

where $\mathbb{E}_0[X_{ij}]$ is the expected value of an edge under the null model considered, c_i and c_j denote the clusters of nodes V_i and V_j, respectively, and $\delta(c_i, c_j) = 1$ if V_i and V_j are located in the same cluster, 0 otherwise.

The standard null model of modularity conserves a network's observed degree sequence by generating half-edges[5], so that each node in the null model has as many half-edges as its corresponding degree in the original network. Thus, the probability to randomly select a half-edge for node V_i is expressed as the ratio of its degree $\rho(V_i)$ and the total sum of all degrees in the network. Noting that the total sum of all degrees in a network corresponds to twice the number of edges ($2m$), this probability is then $\rho(V_i)/2m$. Furthermore, the nodes in the standard null model of modularity are independent and, hence, the probability that two nodes V_i and V_j make a full edge is just the product of their individual probabilities $\frac{\rho(V_i)\rho(V_j)}{4m^2}$. As given in Fortunato (2010) and Newman (2010), the expected value of an edge in this model ($\mathbb{E}_0[X_{ij}]$) is $\frac{\rho(V_i)\rho(V_j)}{2m}$. With this, we can formulate the modularity function as

$$f_{mod} = \frac{1}{2m} \sum_{i=1}^{n} \sum_{j=1}^{n} \left(x_{ij} - \frac{\rho(V_i)\rho(V_j)}{2m} \right) \delta(c_i, c_j). \tag{2.5}$$

In the remainder of this section, we review two modularity based algorithms, the Spectral and Fast Louvain algorithms. We choose these two algorithms because

[5] A half-edge is an edge with only one end. Thus, a full edge is composed of two such half-edges.

of their widespread use in the literature and because of their ability to determine an optimal number of clusters. The latter is an especially appealing feature of these two algorithms compared to many alternative algorithms that require a prior guess of the number of clusters.

2.3.1 Spectral Algorithm

Newman (2006) proposed an optimisation procedure of Eq. (2.5) which is based on a spectral decomposition approach. Considering first just the case of 2 clusters, let $\boldsymbol{s}$ to be an indicator vector for nodes assignment, with $s_i = 1$ if the node V_i is located in the first cluster and $s_i = -1$ if the node is located in the second cluster. The modularity function can then be written in a matrix form

$$\begin{aligned} f_{mod} &= \frac{1}{4m}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(x_{ij} - \frac{\rho(V_i)\rho(V_j)}{2m}\right)(s_i s_j + 1) \quad (\text{setting, } x_{ij} - \frac{\rho(V_i)\rho(V_j)}{2m} = D_{ij}), \\ &= \frac{1}{4m}\sum_{i=1}^{n}\sum_{j=1}^{n} D_{ij} s_i s_j, \\ &= \frac{1}{4m}\boldsymbol{s}^\top \boldsymbol{D}\boldsymbol{s}. \end{aligned} \tag{2.6}$$

Moreover, the vector $\boldsymbol{s}$ can be written as a linear combination of the normalised eigenvectors $\boldsymbol{u}_i$ associated with the matrix $\boldsymbol{D}$ such that $\boldsymbol{s} = \sum_{i=1}^{n} a_i \boldsymbol{u}_i$ and $a_i = \boldsymbol{u}_i^\top \boldsymbol{s}$. Using this along with the fact that β_i is the eigenvalue of $\boldsymbol{D}$ corresponding to the eigenvector $\boldsymbol{u}_i$, we get

$$\begin{aligned} f_{mod} &= \frac{1}{4m}\left(\sum_{i=1}^{n} a_i \boldsymbol{u}_i^\top\right)\boldsymbol{D}\left(\sum_{i=1}^{n} a_j \boldsymbol{u}_j\right) \\ &= \frac{1}{4m}\sum_{i=1}^{n} a_i \sum_{j=1}^{n} a_j \boldsymbol{u}_i^\top \boldsymbol{D}\boldsymbol{u}_j, \quad \text{as} \quad \boldsymbol{u}_i^\top \boldsymbol{D} = \beta_i \boldsymbol{u}_i^\top, \\ &= \frac{1}{4m}\sum_{i=1}^{n} a_i \sum_{j=1}^{n} a_j \beta_i \boldsymbol{u}_i^\top \boldsymbol{u}_j, \quad \text{this is non-zero when} \quad i = j, \\ &= \frac{1}{4m}\sum_{i=1}^{n} a_i^2 \beta_i \\ &= \frac{1}{4m}\sum_{i=1}^{n} \left(\boldsymbol{u}_i^\top \boldsymbol{s}\right)^2 \beta_i. \end{aligned} \tag{2.7}$$

Thus, the idea is to look for the largest positive eigenvalue of $\boldsymbol{D}$, and then group the nodes according to the elements of $\boldsymbol{u}_i$.

The extension of the algorithm to more than two clusters considers the additional contribution to the modularity after dividing a cluster g with n_g nodes into

two clusters

$$\begin{aligned}\Delta f_{mod} &= \frac{1}{2m}\left[\frac{1}{2}\sum_{i,j\in g} D_{ij}(s_i s_j + 1) - \sum_{i,j\in g} D_{ij}\right] \\ &= \frac{1}{2m}\left[\frac{1}{2}\sum_{i,j\in g} D_{ij}s_i s_j - \frac{1}{2}\sum_{i,j\in g} D_{ij}\right] \\ &= \frac{1}{4m}\left[\sum_{i,j\in g} D_{ij}s_i s_j - \sum_{i,j\in g} D_{ij}\right] \\ &= \frac{1}{4m}\sum_{i,j\in g}\left[D_{ij} - \delta_{ij}\sum_{k\in g} D_{ik}\right]s_i s_j \\ &= \frac{1}{4m}\boldsymbol{s}^\top \boldsymbol{D}^{(g)}\boldsymbol{s}, \end{aligned} \tag{2.8}$$

where for the last step, we note that, using the Kronecker delta notation δ_{ij}, we can write

$$\sum_{i,j\in g} D_{ij} = \sum_{i\in g}\delta_{ii}\sum_{k\in g} D_{ik}s_i^2 = \sum_{i,j\in g}\delta_{ij}\sum_{k\in g} D_{ik}s_i s_j,$$

and $\boldsymbol{D}^{(g)}$ is an $n_g \times n_g$ matrix whose elements are given by $D_{ij}^{(g)} = D_{ij} - \delta_{ij}\sum_{k\in g} D_{ik}$. The algorithm stops when there are no more positive eigenvalues.

2.3.2 Fast Louvain Algorithm

The Fast Louvain algorithm (Blondel et al., 2008) optimises a more general version of modularity which is also appropriate for a weighted undirected network. To distinguish it from an ordinary binary undirected network data, we use the notation $\boldsymbol{w} = ((w_{ij}))_{1\le i,j\le n}$ to indicate the corresponding weighted adjacency matrix. Note that a weighted network may also have self connected nodes, in which case the convention of this algorithm is to treat such edges with twice their weights. For example, if a node V_i is connected to itself and this connection has a weight 1, then by convention $w_{ii} = 2 \times 1 = 2$. Similarly to the binary network case, the degree of node V_i is $\rho^w(V_i) = \sum_{j=1}^n w_{ij}$ and W is the total sum of weights of all edges. In such a network, the weighted modularity is given as

$$f_{mod}^w = \frac{1}{2W}\sum_{i=1}^n\sum_{j=1}^n (w_{ij} - \frac{\rho^w(V_i)\rho^w(V_j)}{2W})\delta(c_i, c_j). \tag{2.9}$$

The Fast Louvain algorithm optimises the weighted modularity in two stages that are repeated iteratively. The algorithm is initialised with the number of clusters equal to the number of nodes, that is, each node is assigned to its own cluster. In the first stage, for each node V_i, the algorithm considers each of its neighbours[6] and

[6]Two nodes are said to be neighbours or adjacent to each other if they have a common edge.

computes the gain of modularity that would have been obtained if the node V_i was placed in the same cluster as its neighbour V_j. The node V_i is assigned to the cluster for which this gain is the largest or, in the case of no positive gain, the node stays in its initial cluster. This process is applied sequentially, cycling through every node until no individual move can improve the modularity at which point the first stage stops.

In the second stage, the algorithm builds a new weighted network whose nodes are identified as the clusters found in the first stage. The weight between every node pair of this simplified network is obtained by summing the connection weights between their corresponding clusters and the weight of every node is obtained by counting twice the sum of the weights of its corresponding intra-cluster connections. This gives a simplified weighted network that is used as the initialisation for the next pass of the first stage. These two stages are repeated until the maximal weighted modularity is attained.

2.4 Stochastic Blockmodels

In the domain of social sciences, network analysis has long been a subject of active research. Some of the first work in network analysis is linked to the work of Moreno (1953), who used a graphical representation of the social connections among individuals, which he called a sociogram. This idea was further mathematically formulated by his successors and, in particular, by Luce (1950), who used an adjacency matrix to represent social networks.

The next wave of research came from the work of Lorrain and White (1971), who introduced the concept of *structural equivalence*, by which two nodes are said to be structurally equivalent if they relate to other nodes in the network in the same way. Moreover, Lorrain and White (1971) referred to the collection of nodes sharing the same equivalence structure as an equivalence class or a block. This idea profoundly influenced many researchers, who tried to find optimal ways in which nodes can be clustered into blocks, and, in the sociometric literature, this problem became known as blockmodelling.

However, the evolution of blockmodelling into stochastic blockmodeling is inseparable from the probabilistic network ideas considered by Erdős and Rényi (Erdős, 1947; Erdős and Rényi, 1959). In particular, Erdős and Rényi studied special types of random networks whose connections were independently and identically generated according to some global network probability (i.e. the Erdős-Rényi (ER) model). This implied that the observed network data can be modelled as a product of independent Bernoulli trials which shared a common probability rate. The ER model cannot be directly applied to social networks as the assumption that every node in

the network has approximately the same number of neighbours is not characteristic of observed social networks.

To overcome the issues of the simple ER model, Holland and Leinhardt proposed the p_1 model (Holland and Leinhardt, 1981), which introduced some dependences in the overall distribution of edges by the reparametrisation of the ER model in terms of dyads (i.e. two way node relationships). However, this model was shown to be a poor fit to an already known block structure induced by a classification of individuals according to their gender (Wang and Wong, 1987). To compensate for this, Holland et al. (1983) proposed a stochastic blockmodel which establishes the idea of *stochastic equivalence*, that is to say, that the probability distribution of a network is invariant under permutations of nodes within a block. Thus, the overall distribution of edges can be expressed in terms of block level parameters which quantify the probabilities of connections between the blocks and within the blocks.

However, the classification of nodes into blocks is generally unknown and, thus, we require models which can estimate this latent structure. An appealing way to achieve this is to use Stochastic Blockmodels which are based on mixture models and use likelihood methods for the estimation of the model parameters. In this section, we first review the classical Stochastic Blockmodel (SBM) initially proposed by Snijders and Nowicki (1997). Then, we describe the variational optimisation and the model selection criterion, called Integrated Classification Likelihood (ICL; Biernacki et al., 1998), considered by Daudin et al. (2008) for the SBM. Finally, we review two extensions of the SBM: (1) the extension proposed by Mariadassou et al. (2010) who posed a generalised linear model with an edge-based covariate to model the cluster structure of a single network and (2) the extension proposed by (Zanghi et al., 2010) who instead considered node-based covariates to model the cluster structure of a single network data.

Before going further, let us fix some notation. Hereafter, we employ the usual statistical convention that capital Roman letters denote random variables and lower case letters denote their observed realisations. To distinguish between scalar and non-scalar values, we employ light and bold face fonts, respectively. As an example, a non-scalar random variable is denoted as $\boldsymbol{X}$ and its realisation is denoted as $\boldsymbol{x}$. Furthermore, if $\boldsymbol{X}$ is a discrete random variable, we use three equivalent notations for its probability mass function or its discrete density $f_{\boldsymbol{X}}(\boldsymbol{x}) \equiv f(\boldsymbol{x}) \equiv \mathrm{P}(\boldsymbol{X} = \boldsymbol{x})$. When a density depends on fixed parameters, we write them after a semicolon, as in $f(\boldsymbol{x}; \boldsymbol{\alpha})$; a density conditional on a random variable is written using a vertical bar, e.g., $f(\boldsymbol{x}|\boldsymbol{z})$.

2.4.1 The Stochastic Blockmodel (SBM)

In this section, we consider an undirected and binary network with n nodes whose observed $n \times n$ connectivity matrix $\boldsymbol{x}$ is symmetric (see Section 2.2). The classical Stochastic Blockmodel (SBM) uses a framework of finite mixture models to explicitly account for the variation in the distribution of edges, which give rise to the network cluster structure (for a general review, see Everitt and Hand, 1981). In order to describe the model, let us first assume that the number of blocks Q is fixed, and that the individual blocks are labeled with the indices $q, l = 1, \ldots, Q$. Note, however, that we will later discuss how to estimate Q.

The classification of nodes into Q blocks is represented by a $n \times Q$ matrix of binary latent random variables $\boldsymbol{Z}$. In this matrix, the individual row vectors $\boldsymbol{Z}_i = (Z_{i1}, \ldots, Z_{iQ})$ denote the block assignment of each node in a network. Thus, if the node V_i is assigned to the block q, Z_{iq} is 1 and 0 otherwise. An individual node can only be a member of one block and, thus, $\sum_{q=1}^{Q} z_{iq} = 1$. Marginally, the latent random variables $\boldsymbol{Z}$ are assumed to be independent and to follow a Categorical distribution,

$$\boldsymbol{Z}_i \sim \text{Categorical}(Q, \boldsymbol{\alpha}), \tag{2.10}$$

with individual probabilities of success $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_Q)$ such that $\sum_{q=1}^{Q} \alpha_q = 1$. The probability mass function of $\boldsymbol{Z}$ is then

$$f(\boldsymbol{z}; \boldsymbol{\alpha}) = \prod_{q=1}^{Q} \prod_{i=1}^{n} \alpha_q^{z_{iq}}. \tag{2.11}$$

The SBM specifies that, given the block assignments of the nodes, the elements of $\boldsymbol{X}$ are conditionally independent Bernoulli random variables whose rates are given by their corresponding elements in an association matrix $\boldsymbol{\pi} = ((\pi_{ql}))_{1 \leq q,l \leq Q}$. In other words, if a node V_i belongs to block q and a node V_j belongs to block l, then

$$X_{ij} | Z_{iq} = 1, Z_{jl} = 1 \sim \text{Bernoulli}(\pi_{ql}). \tag{2.12}$$

The conditional distribution of the connectivity matrix given the block assignments is thus

$$f(\boldsymbol{x} | \boldsymbol{z}; \boldsymbol{\pi}) = \prod_{i=1}^{n} \prod_{j<i}^{n} \prod_{q=1}^{Q} \prod_{l=1}^{Q} \left(\pi_{ql}^{x_{ij}} (1 - \pi_{ql})^{1 - x_{ij}} \right)^{z_{iq} z_{jl}}. \tag{2.13}$$

When dealing with mixture models, it is common to make a distinction between the complete dataset $(\boldsymbol{x}, \boldsymbol{z})$ and the incomplete data set $(\boldsymbol{x})$. The likelihood of the SBM is stated as the incomplete data problem. The complete data likelihood is

$$\log f(\boldsymbol{x}, \boldsymbol{z}; \boldsymbol{\pi}, \boldsymbol{\alpha}) = \log f(\boldsymbol{x} | \boldsymbol{z}; \boldsymbol{\pi}) + \log f(\boldsymbol{z}; \boldsymbol{\alpha})$$

$$= \frac{1}{2}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\sum_{q=1}^{Q}\sum_{l=1}^{Q} z_{iq}z_{jl}\log\left[\pi_{ql}^{x_{ij}}(1-\pi_{ql})^{1-x_{ij}}\right] + \sum_{i=1}^{n}\sum_{q=1}^{Q} z_{iq}\log\alpha_q, \tag{2.14}$$

and its marginal likelihood or equivalently the incomplete data likelihood $f(\boldsymbol{x};\boldsymbol{\pi},\boldsymbol{\alpha})$ has the same parameters as the complete data likelihood. Ideally, we would like to base the estimation on $f(\boldsymbol{x};\boldsymbol{\pi},\boldsymbol{\alpha})$ but its implicit computation is generally challenging due to the fact that for each Q, we have Q^n possible node assignments. For this reason, Daudin et al. (2008) suggested the use of a variational approximation, which is reviewed next.

Estimation with Variational Approximation

In order to describe the variational approach, we first explicitly derive a lower bound for the incomplete data, represented by the marginal density $f(\boldsymbol{x};\boldsymbol{\pi},\boldsymbol{\alpha})$. Denoting a sample space of $\boldsymbol{Z}$ as $\mathcal{S}$, we can derive the following inequality

$$\begin{aligned}
\log f(\boldsymbol{x};\boldsymbol{\pi},\boldsymbol{\alpha}) &= \log\left[\sum_{z\in\mathcal{S}} f(\boldsymbol{x},\boldsymbol{z};\boldsymbol{\pi},\boldsymbol{\alpha})\right] \\
&= \log\left[\sum_{z\in\mathcal{S}} f(\boldsymbol{x},\boldsymbol{z};\boldsymbol{\pi},\boldsymbol{\alpha})\frac{f^*(\boldsymbol{z};\boldsymbol{\tau})}{f^*(\boldsymbol{z};\boldsymbol{\tau})}\right] \\
&= \log\left[\mathbb{E}\left(\frac{f(\boldsymbol{x},\boldsymbol{Z};\boldsymbol{\pi},\boldsymbol{\alpha})}{f^*(\boldsymbol{Z};\boldsymbol{\tau})}\right)\right] \\
&\geq \mathbb{E}\left(\log\left[\frac{f(\boldsymbol{x},\boldsymbol{Z};\boldsymbol{\pi},\boldsymbol{\alpha})}{f^*(\boldsymbol{Z};\boldsymbol{\tau})}\right]\right) \quad \text{(by Jensen's inequality)} \\
&= \mathbb{E}\Big(\log[f(\boldsymbol{x},\boldsymbol{Z};\boldsymbol{\pi},\boldsymbol{\alpha})]\Big) - \mathbb{E}\Big(\log[f^*(\boldsymbol{Z};\boldsymbol{\tau})]\Big).
\end{aligned} \tag{2.15}$$

In particular, the lower bound in Eq. (2.15) is closely related to the Kullback-Leibler (KL) divergence of $f^*(\boldsymbol{z};\boldsymbol{\tau})$ to $f(\boldsymbol{z}|\boldsymbol{x};\boldsymbol{\pi},\boldsymbol{\alpha})$, where $\boldsymbol{\tau}$ is a set of variational parameters, as

$$\begin{aligned}
\mathrm{KL}\Big[f^*(\boldsymbol{z};\boldsymbol{\tau})\Big\|f(\boldsymbol{z}|\boldsymbol{x};\boldsymbol{\pi},\boldsymbol{\alpha})\Big] &= \log f(\boldsymbol{x};\boldsymbol{\pi},\boldsymbol{\alpha}) \\
&\quad - \Big[\mathbb{E}\Big(\log[f(\boldsymbol{x},\boldsymbol{Z};\boldsymbol{\pi},\boldsymbol{\alpha})]\Big) - \mathbb{E}\Big(\log[f^*(\boldsymbol{Z};\boldsymbol{\tau})]\Big)\Big].
\end{aligned} \tag{2.16}$$

Rearranging the last equation, we write the lower bound Eq. (2.15) as

$$\log f(\boldsymbol{x};\boldsymbol{\pi},\boldsymbol{\alpha}) - \mathrm{KL}\Big[f^*(\boldsymbol{z};\boldsymbol{\tau})\Big\|f(\boldsymbol{z}|\boldsymbol{x};\boldsymbol{\pi},\boldsymbol{\alpha})\Big]. \tag{2.17}$$

This shows that the lower bound in Eq. (2.15) is precisely attained when the KL divergence is zero, that is, when $f^*(\boldsymbol{z};\boldsymbol{\tau})$ coincides with $f(\boldsymbol{z}|\boldsymbol{x};\boldsymbol{\pi},\boldsymbol{\alpha})$. As we cannot explicitly calculate $\log f(\boldsymbol{x};\boldsymbol{\pi},\boldsymbol{\alpha})$ due to the intractability of $f(\boldsymbol{z}|\boldsymbol{x};\boldsymbol{\pi},\boldsymbol{\alpha})$, we find a family of densities with variational parameter $\boldsymbol{\tau}$ $f^*(\boldsymbol{z};\boldsymbol{\tau})$ to approximate $\log f(\boldsymbol{x};\boldsymbol{\pi},\boldsymbol{\alpha})$ by its lower bound,

$$\begin{aligned}\log f(\boldsymbol{x};\boldsymbol{\pi},\boldsymbol{\alpha}) - \mathrm{KL}\left[f^*(\boldsymbol{z};\boldsymbol{\tau})\middle\|f(\boldsymbol{z}|\boldsymbol{x};\boldsymbol{\pi},\boldsymbol{\alpha})\right]\\ \geq \mathbb{E}\Big(\log[f(\boldsymbol{x},\boldsymbol{Z};\boldsymbol{\pi},\boldsymbol{\alpha})]\Big) - \mathbb{E}\Big(\log[f^*(\boldsymbol{Z};\boldsymbol{\tau})]\Big).\end{aligned} \tag{2.18}$$

Although the lower bound depends on expectations, it is still easier to work with it directly than with $\log f(\boldsymbol{x};\boldsymbol{\pi},\boldsymbol{\alpha})$. Indeed, since the random variables in $\boldsymbol{Z}$ are independent, the expectations of their products, as we will show later, can be factorised. This suggests that we can maximise $\log f(\boldsymbol{x};\boldsymbol{\pi},\boldsymbol{\alpha})$ (or more precisely the likelihood) without its calculation and, consequently, estimate $\boldsymbol{\tau}$ and other parameters of interest. Note, that the accuracy of the variational approximation is of course dependant of our choice of $f^*(\boldsymbol{z};\boldsymbol{\tau})$ and the true log likelihood. Since the random variables $\boldsymbol{Z}$ are independent, a class of densities $f^*(\boldsymbol{z};\boldsymbol{\tau})$ is generally referred to as mean-field or fully-factored densities. To indicate the role of $\boldsymbol{\tau}$, the lower bound is formally denoted as

$$\mathcal{J}(f^*(\boldsymbol{z};\boldsymbol{\tau});\boldsymbol{\pi},\boldsymbol{\alpha}) = \mathbb{E}\Big(\log[f(\boldsymbol{x},\boldsymbol{Z};\boldsymbol{\pi},\boldsymbol{\alpha})]\Big) - \mathbb{E}\Big(\log[f^*(\boldsymbol{Z};\boldsymbol{\tau})]\Big). \tag{2.19}$$

In particular, the natural option for $f^*(\boldsymbol{z};\boldsymbol{\tau})$ is a Categorical distribution with block-specific probabilities, independently for each node:

$$f^*(\boldsymbol{z};\boldsymbol{\tau}) = \prod_{i=1}^{n}\prod_{q=1}^{Q}\tau_{iq}^{z_{iq}}, \tag{2.20}$$

where $\sum_{q=1}^{Q}\tau_{iq} = 1$. Note that this gives the probability of node V_i belonging to block q as $\mathbb{E}(Z_{iq}) = \tau_{iq}$ and that $\mathbb{E}(Z_{iq}Z_{jl}) = \tau_{iq}\tau_{jl}$. Thus, the lower bound $\mathcal{J}(f^*(\boldsymbol{z};\boldsymbol{\tau});\boldsymbol{\pi},\boldsymbol{\alpha})$ can be written as

$$\begin{aligned}\mathcal{J}(f^*(\boldsymbol{z};\boldsymbol{\tau});\boldsymbol{\pi},\boldsymbol{\alpha}) = \frac{1}{2}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\sum_{q=1}^{Q}\sum_{l=1}^{Q}\tau_{iq}\tau_{jl}\log\left[\pi_{ql}^{x_{ij}}(1-\pi_{ql})^{1-x_{ij}}\right]\\ + \sum_{i=1}^{n}\sum_{q=1}^{Q}\tau_{iq}\log\alpha_q - \sum_{i=1}^{n}\sum_{q=1}^{Q}\tau_{iq}\log\tau_{iq},\end{aligned} \tag{2.21}$$

and the full details of this derivation are given in Appendix A.1.

The point estimates for the model parameters $\boldsymbol{\tau}$, $\boldsymbol{\pi}$ and $\boldsymbol{\alpha}$ are obtained by

maximising $\mathcal{J}(f^*(\boldsymbol{z};\boldsymbol{\tau});\boldsymbol{\pi},\boldsymbol{\alpha})$ with respect to each of the parameters, and we get

$$\hat{\tau}_{iq} \propto \hat{\alpha}_q \prod_{i \neq j}^{n} \prod_{l=1}^{Q} \left[\hat{\pi}_{ql}^{x_{ij}} (1 - \hat{\pi}_{ql})^{1-x_{ij}} \right]^{\hat{\tau}_{jl}}, \tag{2.22}$$

$$\hat{\pi}_{ql} = \frac{\sum_{i \neq j}^{n} \hat{\tau}_{iq} \hat{\tau}_{jl} x_{ij}}{\sum_{i \neq j}^{n} \hat{\tau}_{iq} \hat{\tau}_{jl}}, \tag{2.23}$$

$$\hat{\alpha}_q = \frac{1}{n} \sum_{i=1}^{n} \hat{\tau}_{iq}. \tag{2.24}$$

Finally, for each node, the largest variational parameter estimate in $\hat{\boldsymbol{\tau}}_i = (\hat{\tau}_{i1}, \ldots, \hat{\tau}_{1Q})$ determines the classification vector estimate $\boldsymbol{z}_i = (z_{i1}, \ldots, z_{iQ})$, such that

$$\hat{z}_{iq} \;=\; \begin{cases} 1 & \text{if } q \;=\; \text{argmax}_{q'}\, \hat{\tau}_{iq'} \\ 0 & \text{otherwise.} \end{cases} \tag{2.25}$$

Estimation of Optimal Number of Blocks

In a general context, the number of components in a mixture model can be estimated using the Bayesian Information criterion (BIC; Leroux and Puterman, 1992). However, there is some controversy about its accuracy when it is used in the context of clustering. On one side, the studies of Fraley and Raftery (1998), Dasgupta and Raftery (1998), Steele and Raftery (2009) and Gormley and Murphy (2011) are inclined to recommend the BIC for clustering as it gives good results in practice, while, on the other side, the studies of Biernacki and Govaert (1997), Biernacki et al. (1998) and Tantrum et al. (2003) are more cautious to use the BIC in the clustering context. The last group of authors stress that the key modelling principle in clustering is that the mixture components are distinct representatives of cluster structures in the data, meaning that each such cluster should be approximated by a single mixture density component. However, if one of these clusters is not well suited for a single mixture component and is therefore modelled by several components, the correspondence between clusters and mixture components will no longer be distinct. As a result, the estimated number of mixture components will always be higher than the true number of clusters. To address the latter issue, Biernacki et al. (1998) proposed the Integrated Classification Likelihood (ICL) criterion, which is more attuned with the clustering goals of mixture models by favouring models that have well separated clusters. In other words, the ICL penalises model complexity more severely.

Noting a model with Q blocks as M_Q, the starting point of the ICL derivation is the integrated classification likelihood $f(\boldsymbol{x}, \boldsymbol{z}|\mathrm{m}_Q)$ which considers the evidence of the clustering in the data. Provided that the priors for the model parameters can be

factored as $p(\boldsymbol{\pi}, \boldsymbol{\alpha}|\mathrm{m}_Q) = p(\boldsymbol{\pi}|\mathrm{m}_Q)p(\boldsymbol{\alpha}|\mathrm{m}_Q)$, the integrated classification likelihood (see Appendix A.2) can be written as

$$\log f(\boldsymbol{x}, \boldsymbol{z}|\mathrm{m}_Q) = \log f(\boldsymbol{x}|\boldsymbol{z}, \mathrm{m}_Q) + \log f(\boldsymbol{z}|\mathrm{m}_Q). \tag{2.26}$$

In particular, Biernacki et al. (1998) suggested the BIC as an approximation for $\log f(\boldsymbol{x}|\boldsymbol{z}, \mathrm{m}_Q)$. Hence,

$$\log f(\boldsymbol{x}|\boldsymbol{z}, \mathrm{m}_Q) \approx \log f(\boldsymbol{x}|\boldsymbol{z}, \mathrm{m}_Q, \hat{\boldsymbol{\pi}}) - \frac{1}{2}\frac{Q(Q+1)}{2}\log\left[\frac{n(n-1)}{2}\right], \tag{2.27}$$

where $\frac{Q(Q+1)}{2}$ denotes the total number of parameters in $\boldsymbol{\pi}$ and $\frac{n(n-1)}{2}$ denotes the total number of data points in $\boldsymbol{x}$. However, the BIC approximation is not valid for the term $\log f(\boldsymbol{z}|\mathrm{m}_Q)$ since the elements of $\boldsymbol{\alpha}$ can be on the boundary of the parameter space. This happens in scenarios when a mixture model has too many components and the elements of $\boldsymbol{\alpha}$ may shrink to zero. Nevertheless, $\log f(\boldsymbol{z}|\mathrm{m}_Q)$ can be calculated directly by assuming a Dirichlet prior $\mathrm{D}(\delta, \ldots, \delta)$ for $\boldsymbol{\alpha}$ (Diebolt and Robert, 1994) and a Jeffreys non-informative distribution $\mathrm{D}(1/2, \ldots, 1/2)$ (Robert, 2001). Biernacki et al. (1998) showed (see Appendix A.3) that

$$\begin{aligned}\log f(\boldsymbol{z}|\mathrm{m}_Q) = &\log\Gamma\left(\frac{Q}{2}\right) - \log[Q\Gamma\left(\frac{1}{2}\right)] \\ &+ \sum_{q=1}^{Q}\log[\Gamma\left(n_q + \frac{1}{2}\right)] - \log[\Gamma\left(n + \frac{Q}{2}\right)],\end{aligned} \tag{2.28}$$

where $\Gamma(\cdot)$ is the gamma function and $n_q = \sum_{i=1}^{n} z_{iq}$, the estimated number of nodes in block q. Furthermore, under the assumption that n and n_q take large values, Daudin et al. (2008) used Stirling's approximation of Eq. (2.28) (see Appendix A.4) to get

$$\log f(\boldsymbol{z}|\mathrm{m}_Q) \approx \log f(\boldsymbol{z}|\mathrm{m}_Q, \hat{\boldsymbol{\alpha}}) - \frac{Q-1}{2}\log[n]. \tag{2.29}$$

Finally, combining the results in Eq. (2.27) and (2.29), and replacing $\boldsymbol{z}$ by its estimate $\hat{\boldsymbol{z}}$, yields the ICL criterion

$$\begin{aligned}\mathrm{ICL}(\mathrm{m}_Q) = &\log f(\boldsymbol{x}, \hat{\boldsymbol{z}}|\mathrm{m}_Q, \hat{\boldsymbol{\pi}}, \hat{\boldsymbol{\alpha}}) - \frac{1}{2}\frac{Q(Q+1)}{2}\log\left[\frac{n(n-1)}{2}\right] \\ &- \frac{Q-1}{2}\log[n].\end{aligned} \tag{2.30}$$

Degree Distribution and Clustering Coefficient

A basic property of the SBM is that the degree distributions are well approximated by a mixture of Poisson distributions. To show this, we treat the degree of node V_i as a random variable, which is defined as $\rho(V_i) = \sum_{j=1}^{n} X_{ij}$. The degree value of $\rho(V_i)$ increases only when a new edge is added and the probability of such edge depends on the block assignment of node V_i. Hence, for a node V_i in a block q, the updating probability $\mathrm{P}(X_{ij} = 1 | Z_{iq} = 1)$ is given as

$$\begin{aligned} \mathrm{P}(X_{ij} = 1 | Z_{iq} = 1) &= \sum_{l=1}^{Q} \mathrm{P}(X_{ij} = 1 | Z_{iq} = 1, Z_{jl} = 1) \mathrm{P}(Z_{jl} = 1) \\ &= \sum_{l=1}^{Q} \pi_{ql} \alpha_l = \overline{\pi}_q, \end{aligned} \tag{2.31}$$

a weighted average of the internal and external connection rates associated with the block q. Since, $((X_{ij}))_{1 \leq i \neq j \leq n}$ are conditionally independent given the blocks of nodes V_i and V_j, this implies that

$$\rho(V_i) | Z_{iq} = 1 \sim \text{Binomial}(n - 1, \overline{\pi}_q). \tag{2.32}$$

For large counts, a Binomial density can be approximated by a Poisson density with rate equal to the number of trials times the probability of success. This indicates that

$$\rho(V_i) | Z_{iq} = 1 \sim \text{Poisson}(\lambda_q), \quad \text{where} \quad \lambda_q = (n - 1)\overline{\pi}_q, \tag{2.33}$$

$$\mathrm{P}(\rho(V_i) = k) = \sum_{q=1}^{Q} \alpha_q \frac{e^{-\lambda_q} \lambda_q^k}{k!}. \tag{2.34}$$

In addition to approximating the empirical degree distribution, Daudin et al. (2008) also showed how the fitted SBM can be used to estimate Newman's clustering coefficient (see Eq. (2.3)). This estimate is given as

$$\hat{C} = \frac{\sum_{q,l,s}^{Q} \hat{\alpha}_q \hat{\alpha}_l \hat{\alpha}_s \hat{\pi}_{ql} \hat{\pi}_{qs} \hat{\pi}_{ls}}{\sum_{q,l,s}^{Q} \hat{\alpha}_q \hat{\alpha}_l \hat{\alpha}_s \hat{\pi}_{ql} \hat{\pi}_{qs}}. \tag{2.35}$$

Further details about this estimate can be found in Appendix A.5.

2.4.2 The Stochastic Blockmodel of Mariadassou, Robin and Vacher

Mariadassou et al. (2010) proposed an extension of the SBM which is applicable to weighted network data where each edge carries a covariate value. As a motivating example, the authors considered the phylogenetic relatedness between tree species.

In this undirected network, the nodes represent different tree species and the edges represent the number of their common fungal species. Such weighted edges are clearly discrete (i.e. the edges carry positive integer weights). In addition, each edge carry a single covariate value d_{ij} which indicates, on a discrete scale from 0 to 5, the position of each tree pair in a classification tree (i.e. taxonomy). For example, two trees may either be of the same: species (0), genus (1), family (2), order (3), class (4) or phylum (5). Here, the conventional wisdom is that the more taxonomically closer the trees are, the more likely it is that they share common fungal species. Thus, this covariate is expected to be an informative factor in the estimation of the cluster structure.

To model this type of networks, the authors followed the same specification of nodes as in the SBM (see Eq. (2.11)) but in addition they assume that the distribution of edges is Poisson. In particular, they also considered three types of relationships between the covariate of interest and the block structure, yielding three different models. The first model is a Poisson mixture (PM):

$$X_{ij}|Z_{iq} = 1, Z_{jl} = 1 \sim \text{Poisson}(\lambda_{ql}), \tag{2.36}$$

where $((\lambda_{ql}))_{1\leq q,l\leq Q} = \boldsymbol{\lambda}$ is the mean number of common fungal species. This model is trivial as it has no covariate. The second model is a Poisson regression mixture with homogeneous effects (PRMH):

$$X_{ij}|Z_{iq} = 1, Z_{jl} = 1 \sim \text{Poisson}(\lambda_{ql}e^{\beta d_{ij}}), \tag{2.37}$$

where β represents the global network regression coefficient. In this model, the covariate is assumed to have a global effect on the network's edges. The third model is a Poisson regression mixture with inhomogeneous effects (PRMI):

$$X_{ij}|Z_{iq} = 1, Z_{jl} = 1 \sim \text{Poisson}(\lambda_{ql}e^{\beta_{ql} d_{ij}}), \tag{2.38}$$

where β_{ql} represents the regression coefficient associated with block (q, l). In this model, the covariate is assumed to directly interact with each element of the block structure.

The estimation strategy of these models follows the same approach as in the SBM and the model selection is performed with the ICL criterion. Note that the authors only allowed one covariate per model, and did not provide any guidance about the estimation of the regression coefficients and how this can be combined with a variational estimation. Finally, although the authors discussed the PRMH in detail through different simulation scenarios and real data application, the behaviour of the PRMI was not investigated in the original work, neither in the real data analysis

nor in the simulations.

2.4.3 The Stochastic Blockmodel of Zanghi, Volant and Ambroise

Zanghi et al. (2010) proposed an extension of the SBM which is applicable to a binary undirected network data but with the addition of feature values on the nodes. The idea of this model is to incorporate nodal features as an additional clustering factor, so that the nodes are not only grouped to homogenise their connection profiles within blocks but also their nodal features. This model retains the complete structure of the SBM described in Section 2.4.1, but introduces nodal features as a $n \times P$ matrix of random variables $\boldsymbol{Y}$ such that the features pertaining to a node V_i are distributed according to a Multivariate Normal density,

$$\boldsymbol{Y}_i | Z_{iq} = 1 \sim \text{MVN}(\boldsymbol{\mu}_q, \Sigma_q), \tag{2.39}$$

where $\boldsymbol{\mu}_q = (\mu_{q1}, \ldots, \mu_{qP})$ is the vector of population means related to block (q, q) and the variance Σ_q is a $P \times P$ diagonal matrix related to block (q, q). There are a number of typographical errors and inconsistencies in this work that we have resolved (see Appendix A.6). Using Monte Carlo evaluations, they demonstrated how the model could successfully estimate blocks even when there is no obvious cluster structure in the distribution of edges. In such circumstances, the model can switch gears and use the nodal features as the primary clustering source.

2.5 Evaluation Methods

In this section, we review some tools that can be used to compare clustering methods or models with a particular focus on how we can measure the accuracy of their clustering estimates. A first way to achieve this is to compare the estimated partition (i.e. the complete segmentation of the network into a set of clusters) with already known partitions of the nodes (e.g., known biological classifications). A measure of similarity of two partitions is the Adjusted Rand Index (ARI; Handl et al., 2005; Hubert and Arabie, 1985). This measure is a modification of the Rand Index (RI; Rand, 1971), expressed as the fraction of node pairs that are consistent: a node pair is consistent between two partitions if either (i) the node pair is within the same cluster in both partitions, or (ii) the node pair is split between two clusters in both partitions. The interpretation of the RI depends on the number of groups (Morey and Agresti, 1984) whereas the ARI is adjusted for chance agreement and number of clusters (Hubert and Arabie, 1985). The ARI is defined as

$$\text{ARI} = \frac{\text{RI} - \mathbb{E}(\text{RI})}{\max(\text{RI}) - \mathbb{E}(\text{RI})}, \tag{2.40}$$

where the expectation is computed assuming a hypergeometric distribution of the counts of consistent vertex pairs. The ARI scores range from 0 to 1 and indicate the proportion of overlap; for example, if two partitions have an ARI score of 0.6, this means that 60% of their node pairs are classified in the same groups.

A second approach is to assess the quality of an estimated partition with respect to quantitative biological features. To be a useful evaluation, these features should of course not be used in the estimation of the clusters. For this assessment, we can use the Intra-class Correlation Coefficient (ICC). The ICC measures the variance that a partition explains in a continuous variable. As per best practice, we estimate the ICC with a mixed effects model (Dobson, 2001). For a node-wise feature, if we denote Y_{qi} as the measure on the V_i-th node in the q-th group, the mixed effects model is

$$Y_{qi} = \mu + a_q + \epsilon_{qi}, \tag{2.41}$$

where a_q is the random effect of the q-th group, ϵ_{qi} is the random error term and μ is the population mean. The random terms a_q and ϵ_{qi} are mutually independent and each are independently and identically distributed normal random variables: $a_q \sim \mathcal{N}(0, \sigma_a^2)$ and $\epsilon_{qi} \sim \mathcal{N}(0, \sigma_\epsilon^2)$. The ICC is defined as the proportion of total variance explained by the between group variance,

$$\text{ICC} = \frac{\text{Var}(a_q)}{\text{Var}(Y_{qi})} = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_\epsilon^2}. \tag{2.42}$$

In other words, the ICC tells us how homogeneous the biological feature is within the partitions of a proposed network decomposition. Note that, here, we defined the ICC for node-wise measures, but it can be also defined for edge-wise measures. While edge-wise measures may violate the independence assumption of the mixed effect model, the ICC will still be a useful metric to compare the biological validity of different partitions.

CHAPTER 3

Stochastic Blockmodelling of the Modules and Core of the *Caenorhabditis elegans* Brain Network

Caenorhabditis elegans (*C. elegans*) is a tiny roundworm that has become a vital tool in biological research due to its many favourable properties, including relatively short life cycle, constant number of cells, largely invariant neuronal structures across different worms as well as biological mechanisms that are similar to many other animal species, including humans (Hope, 1999). The *C. elegans* worm is lacking vision and hearing but, nevertheless, it is still capable of sensing the environment trough its receptors which respond to chemical, thermal and tactile stimulations. Its behaviour ranges from simple activities, including locomotion or swimming, to more complex activities such as reproduction and even some rudimentary forms of social behaviour (Sporns, 2010). It also has the capacity to adapt and learn, where the latter is related to the ability to change its food search habits as a response to changes in its surrounding environment. In comparison to many other animal species, the nervous system of *C. elegans* is much smaller and less complex, and, as such, it can serve as a reductionist model from which ultimately complex questions related to memory, intelligence, emotions and cognition in the human brain can be researched (Kim, 2008).

Despite the relative simplicity of the *C. elegans* nervous system, it is still difficult to give detailed insights into the way its synaptic connections facilitate functions and how this gives rise to the animal's behaviour. A relatively recent perspective on this problem draws from the theory of networks to examine topological organisations of nervous systems in different animal species. With this approach, it has been demonstrated that other species, such as jellyfish, exhibit a fairly regular network topology which is generally characterised by an overall lack of long distance

connections and the presence of a lattice structure (Watanabe et al., 2009). However, as pointed out by Kaiser and Varier (2011), this organisation is unlikely to be a characteristic of functionally specialised nervous systems and, in the case of *C. elegans* as well as other more complicated nervous systems, an alternative topological structure is expected to yield a more plausible fit.

For more than a decade now, neuroscientists and biologists have hypothesised that highly specialised brain networks are organised in terms of a modular structure which is defined as a subset of nodes, called modules, each of which consists of highly intra-connected nodes which are sparsely inter-connected to nodes in other modules. According to Meunier et al. (2010), modular organisation is expected to be topologically advantageous due to its adaptivity and robustness along with the evolvability of its network function. On these lines, modular organisations have been reported in many different species, including *C. elegans*, macaque and human brains (Bullmore and Sporns, 2009; Rubinov and Sporns, 2010; Pan et al., 2010; Sporns et al., 2004; Hilgetag et al., 2000; Achard et al., 2006; Bassett and Bullmore, 2009) and this structure has been formally verified with the use of the deterministic clustering algorithms described in Section 2.3.

However, one significant weakness of these studies is that they have not considered any other possible topological structure in the networks. Therefore, if some kind of clustering exists in the data and a researcher simply applies a modular algorithm, it does not come as a great surprise that that the resulting decomposition is modular. To compensate for this methodological weakness, other studies turned their attention to the identification of a rich club. Van den Heuvel and Sporns (2011) describe a rich club as a 'tendency of high-degree nodes to be more densely connected among themselves than nodes of a lower degree'. They stress that a rich club plays a key role in the functional integration of different parts of the network, but they fail to offer any explanation as to why such an important topology is not reflected by specific modules and, in particular, why these nodes are blended in with other less important nodes in the network.

To formally investigate these research gaps, we focus on the network of *C. elegans* which has been reported to have a rich club (Towlson et al., 2013) but also a modular structure (Pan et al., 2010). Using the Stochastic Blockmodel (SBM) discussed in Section 2.4.1, we demonstrate its advantages, including its ability to detect a range of cluster structures that are not necessarily modular, to approximate empirical degree distributions and to estimate a clustering coefficient. Using the wealth of information about the neurons of *C. elegans*, we compare the SBM decomposition against the modular decompositions estimated by the popular Spectral and Fast Louvain algorithms. The goodness of fit of each decomposition is assessed according to how much variability within the biological features has been

explained by each fit. We show that the 'cores-in-modules' decomposition of the worm brain network estimated by the SBM is more compatible with prior biological knowledge about the *C. elegans* nervous system than the purely modular decompositions defined deterministically. We also show that the SBM can be used both to generate stochastic realisations (simulations) of the biological connectome, and to compress a network into a small number of super-nodes and their connectivity.

The outline of this chapter is as follows. To keep this material self contained, we start by the detail descriptions of the *C. elegans* network and its external biological information associated with the neurons in the network, and we establish the standard system of nomenclature for the labelling of the neurons. The next section discusses some practical aspects of the analysis related to different software and computational challenges. The next two sections present the results, which detail the estimated structure of the *C. elegans* brain network found by the SBM and compare it to the ones obtained by the Spectral and Fast Louvain algorithms. The final part considers the generative properties of the SBM with respect to the network's degree distribution and clustering coefficient.

3.1 Data

The phylum of Nematoda, also known as nematodes, represents a group of smooth skinned and unsegmented round worms. They consist of more than 25,000 different species (Zhang, 2013) which can be found in diverse habitats either as free-living or parasitic. *Caenorhabditis elegans* (*C. elegans*) is a small, free-living, terrestrial nematode which, in the conditions of generous food supply, tends to grow up to a length of 1.3 mm and with a diameter of up to 0.08 mm.

In its most common hermaphrodite form, every *C. elegans* has 302 neurons in their nervous system, which can be dichotomised into the pharyngeal nervous system, containing 20 neurons, and the somatic nervous system, containing 282 neurons. Each neuron has its own particular combination of characteristics related to morphology, connectivity and position which allows them to be classified into 118 distinct classes that vary in sizes from 1 to 13 neurons. Finally, the 118 classes of neurons are grouped into 10 ganglia, (White et al., 1986). This diverse class structure is very interesting, particularly when contrasted to the mammalian cerebellum, a 'mini brain' primarily responsible for motor coordination and learning, which contains only 5 classes of neurons.

The neurons of *C. elegans* are labelled according to a uniform system of nomenclature where the labels consists of upper case letters with the last entry that can be a number of up to two digits. Further symmetries descriptors are added to the labels to ease the within-class interpretation of the nodes. Thus, D or V

stands for dorsal or ventral[1] and L or R stands for left or right. For example, IL1 is a class of neurons whose members are: IL1DL, IL1DR, IL1L, IL1R, IL1VL and IL1VR. However, when neurons are not part of a set of symmetrical cells, than the third label entry is numerical and describes the anterior or posterior location relative to the other members of the same class. For example, VA03 is the third VA motoneuron.

The neuronal network of the adult nematode *C. elegans* was first described in the publication by White et al. (1986) and was recently revised by Chen et al. (2006) and Varshney et al. (2011). It expresses the regime of connections between the animal's 282 somatic neurons and classifies them with respect to their type and direction (`http://www.wormatlas.org/neuronalwiring.html`, accessed on the 15th June 2013).

In our analysis, we consider a subset of this data where three disconnected neurons (VC06, CANL and CANR) are excluded from the set and we take all connections to be undirected. Furthermore, while the connections are distinguished in terms of their type (chemical synapses, gap junctions and neuromuscular junctions), we treat all connections as binary, that is, we assign the value 1 if some type of connection exist and 0 otherwise. This yields a 279×279 binary and symmetric adjacency matrix with 2287 edges that defines the *C. elegans* network.

For an external evaluation of the cluster estimates, we use *categorical* and *quantitative* characteristics of the neurons (node-wise features) and *quantitative* characteristics of the edges (edge-wise features) as summarised in Table 3.1. There is a

Name	Type	Node-wise	Edge-wise
Ganglion Classification (10 groups)	Categorical	$\surd$	-
Neuron Type (4 groups)	Categorical	$\surd$	-
Neuron Class (103 groups)	Categorical	$\surd$	-
Locomotion Circuit (84 nodes)	Categorical	$\surd$	-
Anatomical Location	Quantitative	$\surd$	-
Anatomical Distance	Quantitative	-	$\surd$
Birth Time	Quantitative	$\surd$	-
Birth Time Difference	Quantitative	-	$\surd$
Lineage Distance	Quantitative	-	$\surd$

Table 3.1: **Prior biological features of the *C. elegans* network**.

large body of knowledge on the individual neurons producing node-wise features. For example, we use the classification of neurons into ten anatomically defined ganglia ('Ganglion classification'), the classification of neurons by their circuitry ('Neuron type') defined by four groups (sensory, motor, interneurons and polymodal neurons),

[1]Dorsal refers to the direction towards the top of the worm, ventral refers to the bottom, or ground-facing part.

as well as topological and synaptic division of neurons ('Neuron class') defined by 103 groups (Varshney et al., 2011; Altun and Hall, 2005). We also consider the ventral nerve cord motor neurons involved in locomotion, egg-laying and possibly avoidance (broadly labelled as 'Locomotion circuit' in Table 3.1) which was described by Haspel et al. (2010) using the connection data from Wood et al. (1987), Von Stetina et al. (2006), Altun ZF (2009), and Chen et al. (2006). Explicitly, this circuit is composed of 84 neurons, of which 74 are motor neurons (excluding VC06) and they comprise eight neuron classes. Four of these classes are connected to ventral muscles (VA, VD,VB and VC) while the other four classes are connected to dorsal muscles (AS, DA, DD and DB). The remaining 10 neurons are interneurons (AVA, AVD and AVE; AVB and PVC) promoting the backward and forward motion. Although the connection data used in our analysis do not include neuromuscular connections, the circuit presented by Haspel et al. (2010) provides some invaluable insights that are beneficial to the evaluation and comparison of the results obtained in our analysis. The remaining set of the node-wise features includes the 'Anatomical Location' (longitudinal and sectional positions) of the cell body (soma) and the 'Birth Time' of each neuron (`http://www.biological-networks.org/?page_id=25`, accessed on the 15th June 2013; Varier and Kaiser, 2011).

The edge-wise features include the 'Anatomical Distance' (Euclidean distance between each neuron pair), the 'Birth Time Difference' (for each neuron pair, we take an absolute difference in their birth times) and the ' Lineage Distance' (for each neuron pair, this is the sum of total divisions to the most recent common ancestor cell; Brenner, 1974).

3.2 Practical Considerations

Cluster estimation methods are notoriously sensitive to the initial starting conditions (see e.g., Mukherjee and Hill, 2011). Each method begins with some sort of random initialisation that typically leads to a local optimum of the objective function (i.e. ICL or modularity). Thus, for all three methods considered, we use multiple random restarts of the algorithm and take the solution that provides the greatest value of the objective function. In particular, we fitted the SBM with the R package **Mixer** (Daudin et al., 2008; Latouche et al., 2008; Zanghi et al., 2008a,b). The `mixer` function specifies default values for the maximum number of iterations, and we found improved performance by increasing these (`nbiter=80` up from `10`, `fpnbiter=40` up from `5`). We found that $1,000$ random restarts was sufficient to visit the optimal solution multiple times, but, to be exhaustive, we also considered up to $100,000$ random restarts. The Fast Louvain and Spectral algorithms were fitted using the Matlab **Brain Connectivity Toolbox** (`http://www.brain-connectivity-toolbox.net/`, ac-

cessed on the 15th June 2013; Rubinov and Sporns, 2010). For the Fast Louvain algorithm, we used the function `modularity_louvain_und` using $20,000$ restarts. For the Spectral algorithm, we used the function `modularity_und`. This function is initiated on a randomly permuted adjacency matrix and although, in theory, all permutations of the adjacency matrix should provide the same result, some numerical discrepancies may occur during the spectral decomposition, subsequently leading to slightly different modularity fits. Specifically, the variability in the fits is driven by numerical errors in the estimation of the elements of eigenvectors $\boldsymbol{u}_i$ (Eq. (2.7)), which can erroneously change the sign of its element. For example, if the true value of an element of $\boldsymbol{u}_i$ is 5×10^{-20} and the error is -1×10^{-19}, the estimated value would be -5×10^{-20}. Indeed, this has an immediate impact on the classification vector $\boldsymbol{s}$ (Eq. (2.7)) which will classify the corresponding node in the wrong group. To be exhaustive, we have therefore used $20,000$ restarts.

Also, we report the computational times obtained on a 2.7 GHz quad-core Intel Core i7 linux host with 16 GB. The SBM, with the default parameters setting (i.e. `nbiter=10`, `fpnbiter=5`) and the range $Q \in \{2, \ldots, 50\}$, took 186 seconds while the Louvain and Spectral algorithms took 0.07 and 0.62 seconds, respectively. Both the SBM and Louvain methods required multiple restarts to find the optimal model, while restarts for the Spectral algorithm were only needed due to numerical errors. For the SBM, on average, $1,000$ restarts were needed to visit the optimal model 12 times (see Figure 3.1) while, for the Louvain algorithm, over the total of $20,000$ restarts, the optimal model was visited only once (see Figure 3.2). To measure the similarity between a partition (i.e. the complete segmentation of a network into a set of groups) and some known biological classifications, we use the Adjusted Rand Index (ARI; see Section 2.5). The ARI scores were obtained using the function `adjustedRandIndex` in R software (Fraley and Raftery, 2006, 2002) and, for the ICC (see Section 2.5), we use the R function `lmer` Bates et al. (2012) that employs a Restricted Maximum Likelihood procedure (Harville, 1977) to obtain the parameter estimates.

3.3 Results

We first detail the estimated cluster structure of the *C. elegans* brain network found by the SBM, and then we proceed to compare these results with the ones obtained by the Fast Louvain and Spectral algorithms. The final part considers the generative properties of the SBM with respect to the network's degree distribution and clustering coefficient.

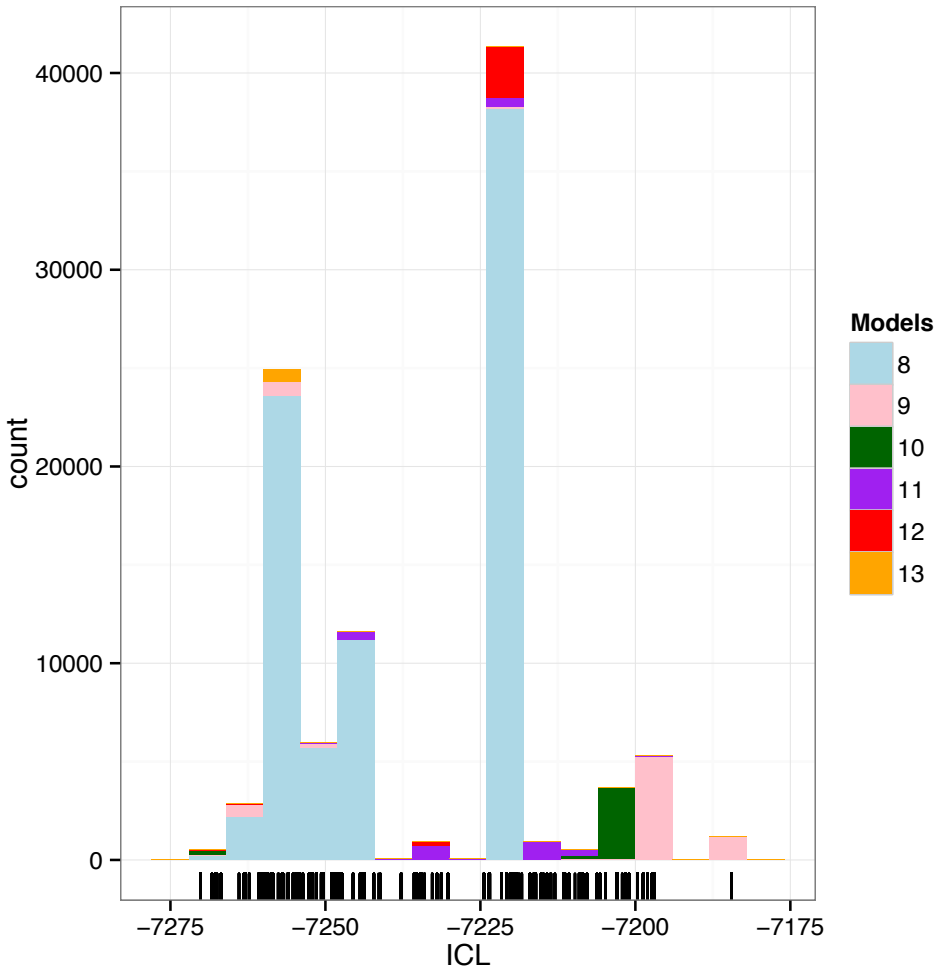


Figure 3.1: **Model selection for SBM**. Histogram of the maximal ICL scores taken over 100,000 restarts and given in terms of the number of blocks in the corresponding fitted model. The number of blocks within the models that maximised the ICL scores ranged from 8 to 13. The top ICL range (above -7218) was attained by the models with 9, 10 and 11 blocks. Out of these, the model with 9 blocks achieved the overall maximal ICL score of -7184.5 and this identical partition was repeated 1167 times.

3.3.1 Stochastic Blockmodel Estimate of the *C. elegans* Cluster Structure

The optimal SBM fit of the *C. elegans* brain network consists of 9 blocks, each of which is listed in Figure 3.3. In addition, the anatomical locations of neurons in each block are given in Figure 3.4.

Broadly speaking, we found that approximately 70% of the neurons in Block 1 are head sensory neurons involved in chemo/thermotaxis or chemo/thermosensation which modulate body movement. About 20% of the neurons in Block 2 are also head

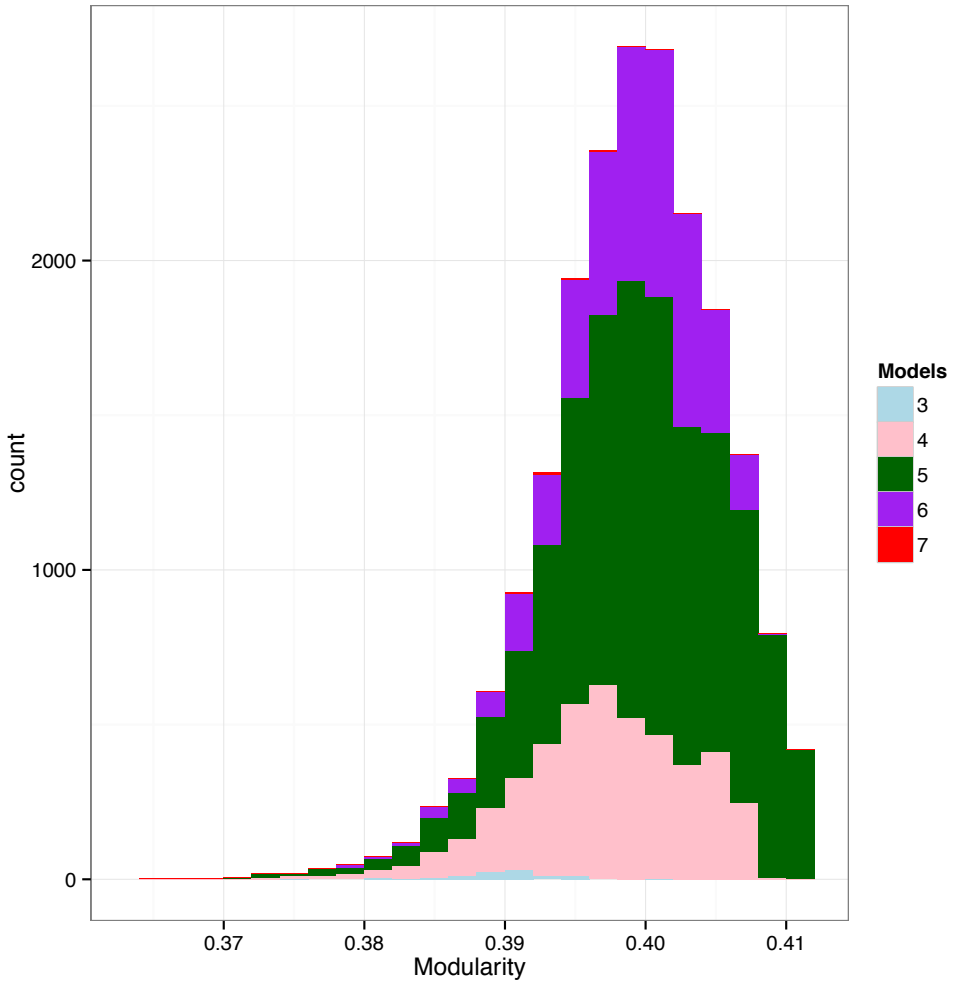


Figure 3.2: **Selection of the optimal number of modules for the Louvain algorithm**. Histogram of modularity scores taken over 20,000 restarts and given in terms of the number of modules in the corresponding model. The maximal modularity score 0.411 (rounded to 3 decimals) was achieved by models with 5 modules and this was repeated 94 times. Note that the models with 5 modules were mutually different and that the best fit (i.e. the fit that achieves the global maximum of the modularity score) occured only once.

sensory neurons but involved in more direct, reflex like and deterministic effects on body movements such as escape or avoidance behaviour. Almost 60% of the remaining neurons in Block 2 are ring interneurons (ADA, AIB, AVK, RIA, RIB, RIC, RIG, RIS, RMG, URX), about half of which have unknown functions. Note that we characterised this block as 'escape/avoidance' even though its function or perhaps functional homogeneity is not entirely clear. Next, more than half of the neurons in Block 3 (55%) consists of mid-body and posterior ventral cord motor neurons while almost all of the remaining neurons are posteriorly located sensory neurons (PDE,

PHA, PHB, PHC, PLM, PVD, PVM), known to have quite a direct effect on motor neurons (e.g., PHA and PHB control the extent of reversals in chemo-repulsion). We have labeled this group 'motor (posterior)', but we will revisit the possible causes for their inclusion. Similarly, close to 90% of neurons in Block 4 are made up of anteriorly located ventral cord motor neurons (AS, DA, DB, VA, VB, VD) and it is therefore labeled as 'motor (anterior)' group. The next two Blocks (5 & 6) are among the smallest in size, each with only 6 neurons. In particular, 4 neurons in Block 5 are command interneurons for (backward) locomotion (AVD, AVE), while the remaining 2 neurons are DVA (mechanosensory integration) and PVR (unknown function); whereas all 6 neurons in Block 6 are locomotion command interneurons. Next, Block 7 is mostly (about 65%) composed of neurons with unknown functions. However, as 15% of the neurons seem to be involved in egg-laying and defecation, we have labeled Block 7 as 'unknown/egg-laying/defecation' group. The largest number of neurons is found in Block 8 which appears to be predominantly (about 60%) composed of head motor neurons and nose touch mechanoreceptors (mainly located in the head) as well as numerous ring motor neurons. Many of these neurons are involved in both local search behaviour (RIV, SMDD, SMDV) and avoidance or aversive head withdrawal (ALN, IL1D, IL1V, OLQD, OLQV, RMD). We labeled this block as 'nose-touch/head motor', but it may be interesting to further investigate whether this block could be subdivided into more specialised subunits. The remaining Block 9 is composed entirely (100%) of anterior ventral cord motor neurons of class DB, DD, VB, VC and VD - as previously discussed, Block 4 contains the remaining of anterior ventral cord motor neurons of type AS, DA, DB, VA, VB, VD. Overall, these observations indicate that the SBM partition highlights functionally meaningful features of the system's topological organization. In Figure 3.5 (a), we show the optimal SBM fit as a reorganised adjacency matrix. Note that the SBM fit demonstrates the dense connections between - as well as within - certain groups. This is in stark contrast to traditional deterministic methods that seek to find modules with dense intra-modular connectivity and sparse connections between modules. Instead, the SBM classifies neurons into separate groups according to their individual connectivity profile to other groups, regardless of where connectivity happens to be dense. For example, Block 6 (AVA, AVB, PVC) comprises neurons with maximal interconnections (i.e. clique), which are also fairly densely connected to the rest of the network. Note that Blocks 5 (AVE, AVD) and 6 (AVA, AVB, PVC) are separate groups because of differing internal connection rates (i.e. $\pi_{55} = 40\%$ vs. $\pi_{66} = 100\%$) and external connection rates (e.g., $\pi_{54} = 30\%$ vs. $\pi_{64} = 45\%$). Also, while Blocks 1 (chemo/thermosensation or chemo/thermotaxsis) and 2 (escape/avoidance) seems to have similar internal and external block connections, they differ by the fact that Block 1 has virtually no connections with Block 8 (nose touch

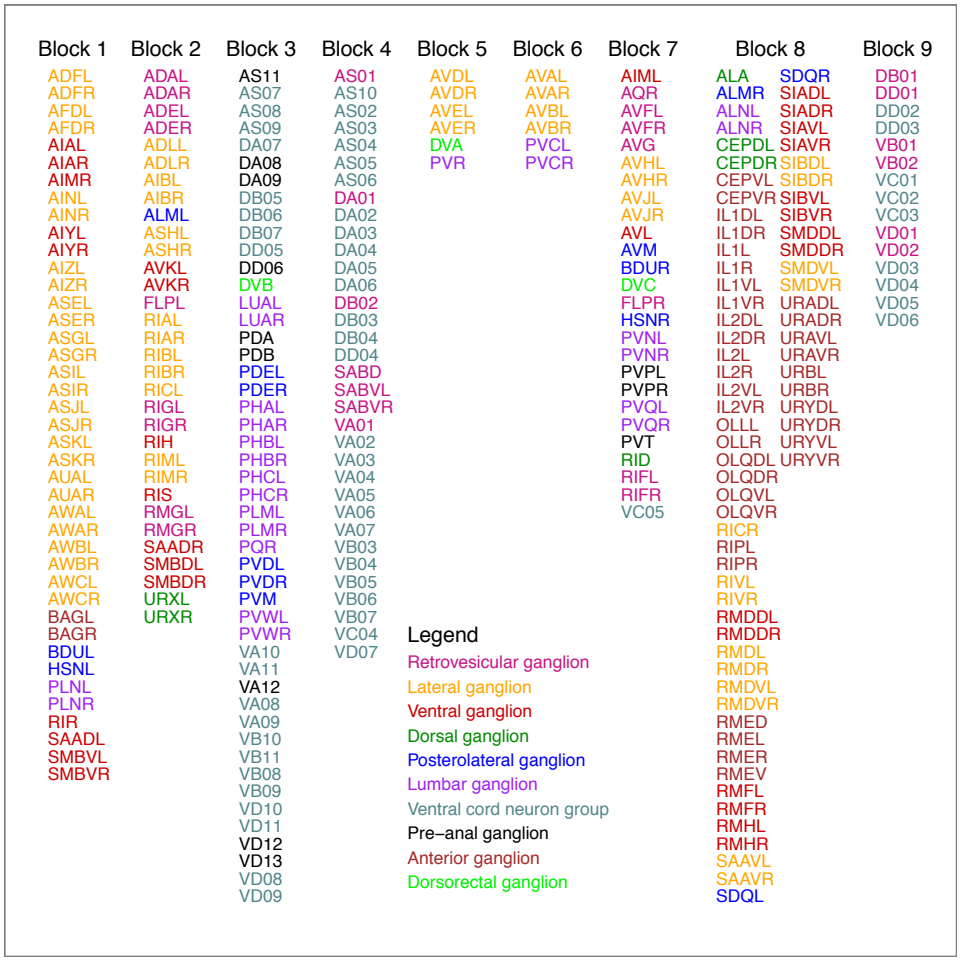
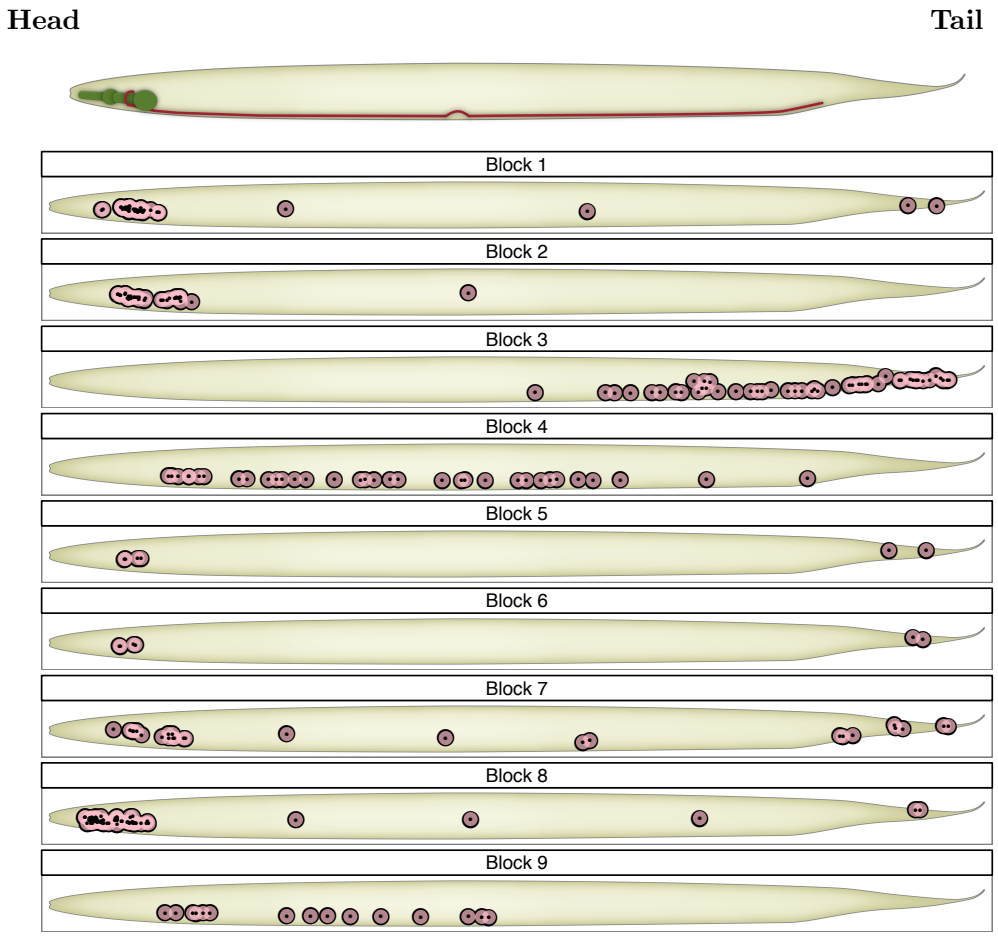


Figure 3.3: **Classification of neurons for each Block of the SBM fit**. The corresponding neuron labels are colour coordinated according to their ganglion type.

mechanoreceptors and head motor neurons) while Block 2 is densely connected to Block 8.

A concomitant advantage of the SBM approach to cluster estimation is its ability to provide a compressed view of the original *C. elegans* network. As shown in Figure 3.6, this compressed view serves as a summary of the network's topology and reveals diverse patterns of connectivity between the blocks. Here, some blocks, such as Blocks 1, 6 and 8, appear to fit the standard definition of a 'module' with high internal connectivity and sparse external connectivity. However, other structures which are characterised by strong communications between blocks are also present in the network; for example, Blocks 5 & 6 and Blocks 6 & 7, which may suggest that these are involved in the same functional circuit. In particular, Block 6 (command interneurons) - previously identified as a clique - maintains relatively strong ties

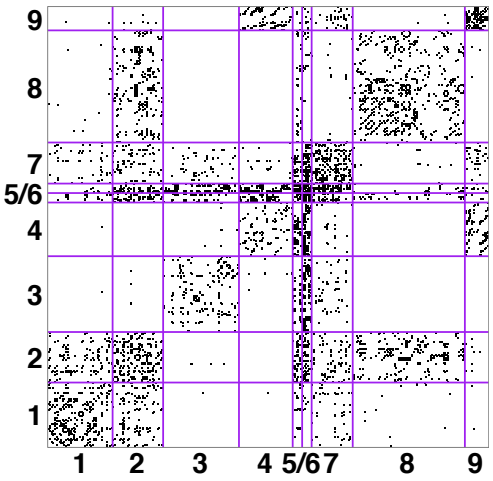


(a)

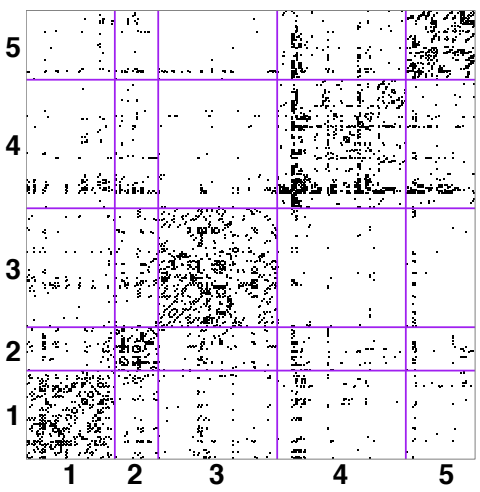
Figure 3.4: **Anatomical locations of neurons (cell body) according to the SBM fit**. Each block is shown on an approximate template, obtained from `http://www.wormatlas.org/`, last accessed on the 9th October 2013.

with Blocks 3 and 4 (motor neurons) whose internal connections are sparse. This structure is known as the 'core-periphery' and has been shown to be a functionally significant organisational structure in various real-world networks such, as social networks or power grids (Borgatti and Everett, 2000; Ravasz and Barabási, 2003).

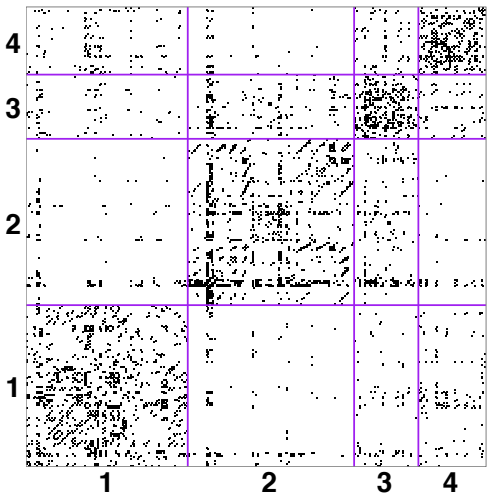
For further qualitative evaluation of the SBM cluster estimate, we use the Locomotion circuit as described in the accompanying text of Table 3.1. Figure 3.7 (a) shows a simplified diagram of this circuit (originally presented by Haspel et al., 2010) with the neuron block membership indicated by colour. The SBM isolated the command interneurons into Blocks 5 and 6 while Blocks 4, 3 and 9 are fairly uniformly spread over all motor neurons. The distinction between these three blocks of motor neurons appears to be, at least partially, anatomically motivated,



(a) SBM



(b) Louvain



(c) Spectral

Figure 3.5: **Reorganised adjacency matrices for each approach**. The groups are ordered arbitrarily; within each group, the neuron labels are sorted in alphabetical order. (a) The SBM fit demonstrates the dense connections between - as well as within - certain blocks. This is in stark contrast to traditional deterministic methods (b) and (c), that seek to find modules with dense intra-modular connectivity and sparse connections between modules. In addition, the SBM fit defines blocks according to their internal and external connections. Thus, although Block 1 and 2 have similar within block connections, they are split because of their different connectivity with Block 8.

with the neurons in Block 3 being more posterior while the neurons in Blocks 4 and 9 are mainly found in the mid-section and anterior parts of the animal. Another noteworthy point is that the neurons VC04 and VC05, both implicated in egg-laying, are assigned separately to Blocks 4 and 7. The principal justification of this

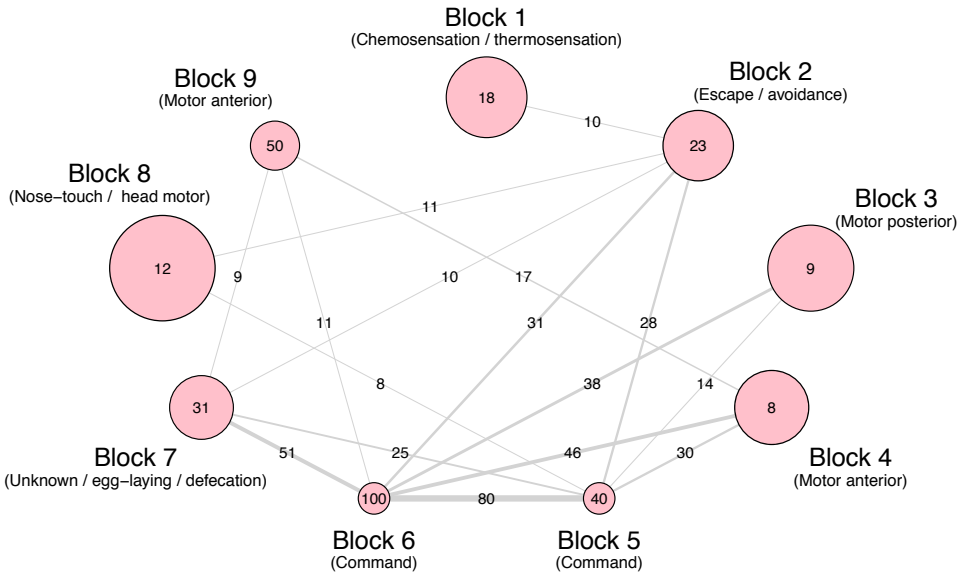
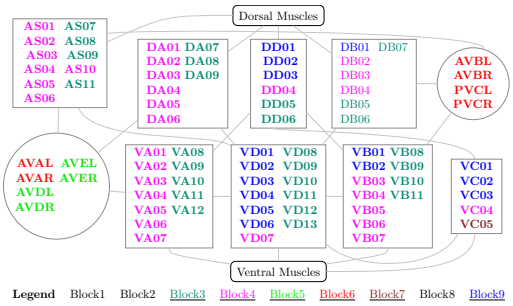
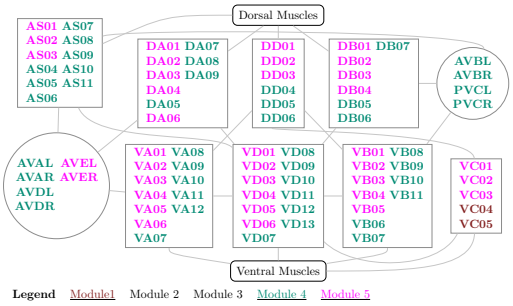


Figure 3.6: **Compressed view of the *C. elegans* network, in terms of between/within block connection probability rates of the SBM fit**. The relative size of each circle indicates the number of neurons in that block. The number inside the circle is the within-block connection probability in percent. The relative thickness of each line indicates the between-group connection strength, while the number on the edge gives the connection probability in percent (those less than 7 % are omitted). Each block is broadly characterised by its most representative function. Note how Blocks 1, 2 and 9 are 'modules' with internal connectivity greater than external connectivity, while other structures are characterised by strong inter-block connectivity (e.g., Blocks 5 & 6 and Blocks 6 & 7). In addition, Block 6 (command interneurons) maintains relatively strong ties with Blocks 3 and 4 (motor neurons) whose internal connections are sparse, an example of core-periphery.

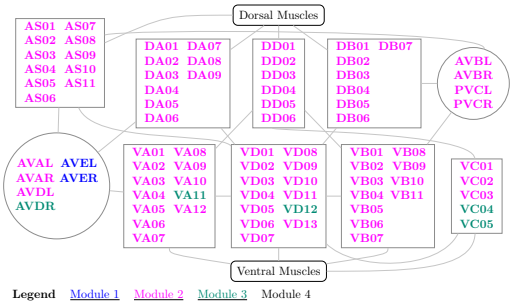
separation can be traced back to the network data used in this analysis where, for example, VC04 maintains connections to the locomotion neurons AVB and AVH while VC05 does not, and, moreover, VC05 maintains connections to the egg-laying neurons AVFL, AVFR, HSNR and PVT (Block 7) while VC04 does not. Given such differences in connection profiles between these two neurons, it is not surprising that they are separated. Relating to this, it is also worth mentioning that our network data excludes neuromuscular connections to the vulval muscles, made by both VC04 and VC05, which are the primary reason why these neurons are implicated in egg-laying behaviour.



(a) SBM



(b) Louvain



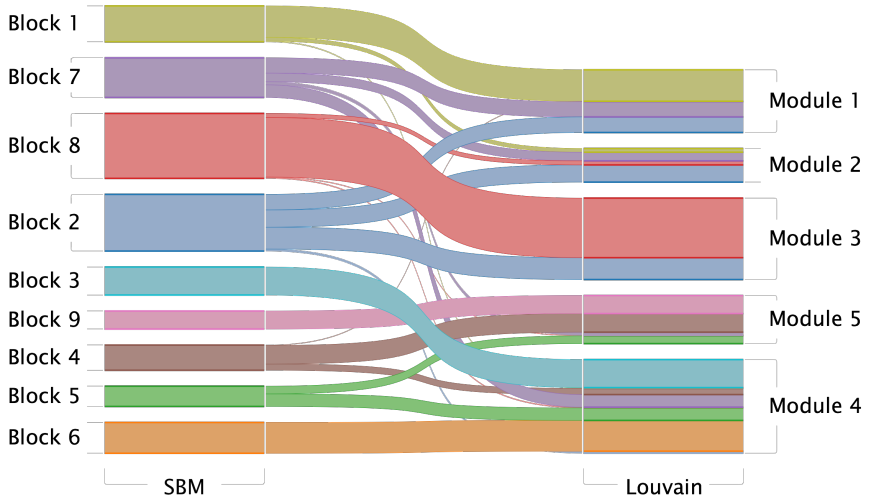
(c) Spectral

Figure 3.7: ***C. elegans* locomotion network and estimated cluster structure for each approach**. Each subfigure shows the 74 Motor neurons (shown in rectangles) that support the animal's motion, which are divided into eight distinct groups. Four of these groups are connected to the ventral muscles (neuron labels: VA, VD, VB and VC; VC06 is omitted in our analysis), while the remaining four groups are connected to the dorsal muscles (neuron labels starting AS, DA, DD and DB). The remaining neurons (command neurons; shown in circles) belong to the category of interneurons; some are primarily required for promoting forward movements (labels starting PVC and AVB) while others promote backward movements (labels starting AVA, AVD and AVE). The colour of each neuron indicates the group membership from a particular method's partition. The SBM fit (a) isolates the command neurons in Blocks 5 & 6 and distinguishes the posterior (Block 3) from the more anterior motor neurons (Blocks 4 & 9).

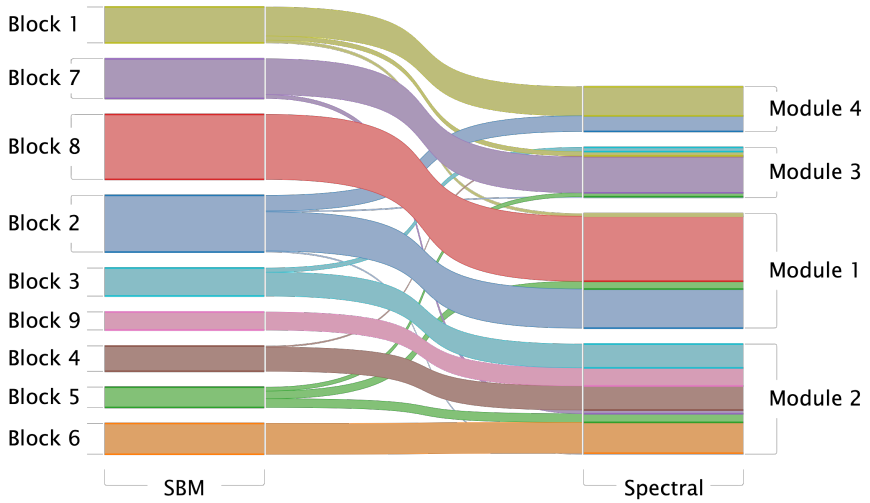
3.3.2 Comparative Assessment of the *C. elegans* Cluster Structure Estimates between the Stochastic Blockmodel and the Fast Louvain and Spectral Algorithms

The Spectral and Fast Louvain algorithms decompose the *C. elegans* network into 4 and 5 modules with maximal modularity scores of 0.402 and 0.411, respectively (Eq. (2.7) and (2.5)), indicating that both algorithms detect a prominent modular structure. As shown in the adjacency matrices in Figure 3.5 (b) and (c), both the Spectral and Louvain algorithms produced partitions with strong within-group connections and relatively sparse inter-group connections, as expected by definition.

In order to compare the cluster structures obtained via all three methods, we plot an alluvial diagram (see Figure 3.8 (a) and (b)) showing each block of the SBM (on the left) and how these merge and split in order to make up the modules of the Louvain and Spectral partition. Strands of the alluvial diagram are coloured according to the block decomposition of the SBM. The first thing to note when



(a)



(b)

Figure 3.8: **Correspondence between the SBM fit and the estimates of Louvain (a) and Spectral (b) algorithms**. The strands of the alluvial diagram show each block of the SBM fit (on the left) and how these merge and split to form the modules of the Louvain and Spectral algorithms (on the right). The functional labels for the SBM blocks are as follows. Block 1 (chemosensation/ thermosensation), Block 7 (unknown/egg-laying/defecation), Block 8 (nose touch/head/motor), Block 2 (escape/avoidance), Block 3 (motor posterior), Block 9 (motor anterior), Block 4 (motor anterior), Block 5 (command) and Block 6 (command).

observing this diagram is that the blocks obtained with the SBM often roughly correspond to modules obtained via the other methods, with the Louvain and Spectral algorithm merging progressively more blocks into fewer modules. Secondly, we note that Blocks 3, 4, 5, 6 and 9 (mainly ventral cord motor neurons and interneurons controlling locomotion) are fairly well separated from Blocks 1, 2, 7 and 8 by all algorithms, so we will discuss these two subsets separately below.

Most of the nodes in Block 1 (chemo/thermo sensation) are also classed together by the other two algorithms, although they are also merged with some nodes from Blocks 2 (escape/avoidance) and 7 (mainly unknown function) in Module 1 of the Louvain algorithm. In contrast, the nodes in Block 2 are fairly dispersed in the Louvain algorithm (equally distributed between Modules 1, 2 and 3), while they are split between Module 1 and 4 (mainly amphid neurons) by the Spectral algorithm. As noted earlier, the neurons in Blocks 1 (chemo/thermo sensation) and 2 (escape/avoidance) are also tightly interconnected and they only fall as separate blocks because of their differential connectivity to nodes in Block 8 (nose-touch/head motor). The functional relevance of this finding is yet unclear but this pattern is biologically plausible and is a particularly striking aspect of the SBM result (as shown in Figure 3.5 (a)). We also note that, while Block 8 seems to lump together many of the non-sensory neurons in the head, these neurons are also all grouped together by both the Spectral (Module 1) and the Louvain algorithm (Module 3). Nevertheless, it may be interesting to further investigate whether this block could be subdivided into more specialised subunits. One such approach could be to include in the analysis virtual nodes for various external cues (chemical attractants, olfactory cues, temperature, touch, osmolarity, etc) or to include virtual nodes for various muscle groups controlled by the motor neurons, as this information has recently been shown to be useful in understanding the connectivity of motor neurons in the Locomotor system (Haspel and O'Donovan, 2011).

Looking at Block 7, we note that it corresponds quite well to Module 3 in the Spectral algorithm, but it is split between all modules (and mainly Modules 1 and 4) in the Louvain algorithm. From Figure 3.5 (a), it is clear that Block 7 has a very specific connectivity pattern. We therefore predict that this is likely to correspond to a biologically relevant functional grouping. This is particularly interesting because many of the neurons in this block have unknown functions and because these neurons are not anatomically co-located. Thus, in investigating the functional relevance of this block, it will be important to consider its particularly strong relationships to Blocks 2 and 6.

Now, turning our attention to Blocks 3, 4, 5, 6 and 9, we note the following observations. Block 4 is made up almost entirely of anteriorly located ventral cord motor neurons (AS, DA, DB, VA, VB, VD) while Block 9 is composed entirely of

anterior ventral cord motor neurons (DB, DD, VB, VC and VD). These two blocks are merged by both the Spectral and Louvain algorithms and, looking at Figure 3.5 (a), their separation into two different blocks does not seem to be a strong feature of the SBM method either. It seems to be based by a differential connectivity to Block 7, but the effect is not very strong.

As previously mentioned, Block 3 is composed mostly of mid-body and posterior ventral cord motor neurons and almost all of the remaining neurons are posteriorly located sensory neurons (PDE, PHA, PHB, PHC, PLM, PVD, PVM). Almost all neurons in this block (including the posteriorly located sensory neurons listed) are also grouped together in Module 4 of the Louvain algorithm and almost all of them are in Module 2 of the Spectral algorithm. We however note that, in the Spectral (but not Louvain) partition, these neurons are also grouped together with the anterior ventral cord motor neurons of Blocks 4 and 9. While the roughly anatomical split between ventral cord motor neurons in the SBM and Louvain method may not lead to new biological insights, it is certainly driven by a strong lack of connectivity between Blocks 4 (anterior) and 3 (posterior) which is a true feature of the data. It is worth noting that the connectivity data for *C. elegans* are known to be partial or missing for 39 of 302 neurons, including 21 of the 75 locomotor motoneurons (Haspel and O'Donovan, 2011) and the data for the posterior parts of the nerve cords are especially sparse and uncertain. It is therefore unclear whether this split between Blocks 3, 4 and 9 contains biological information or whether a more complete mapping of connections in the posterior part of the ventral cord would alter these results. Note, for example, that the split does not correspond to a division between neurons involved in forward and backward locomotion (Haspel et al., 2010).

Finally, Blocks 5 and 6 are also merged with ventral cord motor neurons from Block 3 in both the Louvain and Spectral algorithms. This is driven by the dense inter-connectivity between these nodes, however, the separation of Blocks 5 and 6 from the rest of the networks is one of the key features of the SBM decomposition. Indeed, these blocks correspond almost exactly to a rich-club (core-periphery structure) whose functional importance has recently been confirmed (Towlson et al., 2013).

It is also worth noting that compressed views of the network - see the SBM fit in Figure 3.6 - are not available for the Fast Louvain and Spectral algorithms since these, by definition, decompose the network into modules with minimal connectivity between them.

Figure 3.7 (b) and (c) show the Locomotion circuit and the partitions found by the Spectral and Fast Louvain algorithms. In contrast to the SBM model, both of these algorithms failed to distinguish the command neurons from the motorneurons.

In the case of the Spectral algorithm, some of the command neurons like AVEL and AVER are isolated but the rest are mixed with the motorneurons. This effect may be explained by the rigid definition of 'cluster structure' that is common to both algorithms. As we can observe, this particular *a priori* assumption does not allow the network's topology to dictate the form of the cluster structure, resulting in functionally less meaningful decompositions. Similar observations can be made about the neurons VC04 and VC05, which are merged by both algorithms despite their different connectivity profiles, inherent to the data. As we saw previously, these neurons are split in the SBM partition.

Further quantifications of the solutions in terms of the separation of L/R (left/right symmetric) neurons of the same class are presented in Table 3.2. Here, we note that, out of 92 L/R neuron pairs contained in this data set, the SBM and Spectral algorithm partitions assigned 85 such neuron pairs in the same groups and misclassified 7 pairs while the Louvain partitions misclassified 5 pairs. In general, ALM and SAAD are separated by all methods, while other misclassified neurons appear to be distinct.

SBM			Spectral			Louvain		
AIMR	AIML	Block (1,7)	ALMR	ALML	Module (1,3)	RMGR	RMGL	Module (1,3)
BDUL	BDUR	Block (1,7)	RMGL	RMGR	Module (1,4)	SMBVL	SMBVR	Module (1,2)
HSNL	HSNR	Block (1,7)	PLNR	PLNL	Module (1,4)	ADER	ADEL	Module (2,3)
SADDL	SADDR	Block (1,2)	RMGL	RMGR	Module (1,4)	ALMR	ALML	Module (3,4)
ALML	ALMR	Block (2,8)	SAADR	SAADL	Module (1,4)	SDQR	SDQL	Module (3,4)
FLPL	FLPR	Block (2,8)	AVDL	AVDR	Module (2,3)			
RICL	RICR	Block (2,8)	SABVR	SABVL	Module (2,3)			

Table 3.2: **Bilateral pairs of neurons split by each method**. The neurons are listed according to their group assignment.

Although the clustered neurons in the SBM partition appear to be functionally related, this overall partition does not closely correspond to the anatomical partition of neurons in 10 groups called ganglions (see Figure 3.3). As shown in Figures 3.9 and 3.10 this is also true for the Spectral and Louvain partitions. However, to more formally assess the level of similarity between each of the estimated and known biological partitions (i.e. ganglion, neuron classes and neuron types), we use the ARI scores. As shown in Figure 3.11 (a), we observe that the ARI scores are small and no greater than 0.26 for all 3 comparisons, with the ganglion based partition being matched by the SBM and Spectral algorithm with 0.25 ARI units, while the Louvain algorithm scores slightly lower. Regarding the partition with 103 classes of neurons, the SBM partition exhibits a relatively low ARI scores, but slightly higher ARI score than those of the Louvain and Spectral algorithms. For the ARI scores by neuron type, the Spectral cluster structure estimate seems to

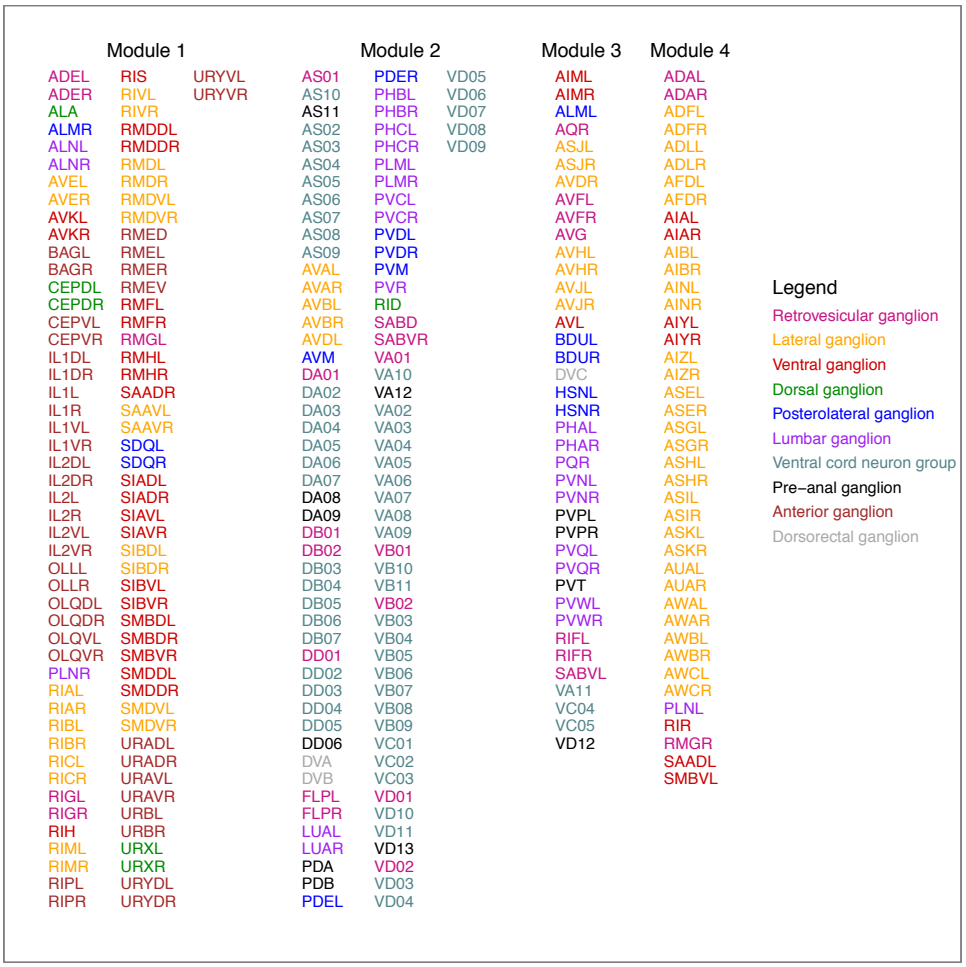


Figure 3.9: **Membership structure of the neurons in the Spectral fit**. Neurons are coloured coded according to their ganglion type.

be slightly more compatible than the other two fits, which tend to assign different neuron types to the same groups. These findings suggest that in general all three solutions are fairly different from the known biological partitions.

However, it has to be noted that, although the ARI score can quantify the level of similarity between partitions of interest, it cannot infer (i.e. assign a $p-$value) on whether one partition has a significantly better fit than another partition. This is statistically challenging as the solutions (fits) are sourced from the three different methods and, therefore, simple validation strategies like the one presented in Pan et al. (2010) would not be appropriate. Nevertheless, using the ICC (Eq. (2.42)), we can compare qualitatively all three estimates. For this, Figure 3.11 (b) shows the ICC performance of the estimated partitions with respect to six quantitative biological features. The hypothesis here is that neurons that are implicated in the

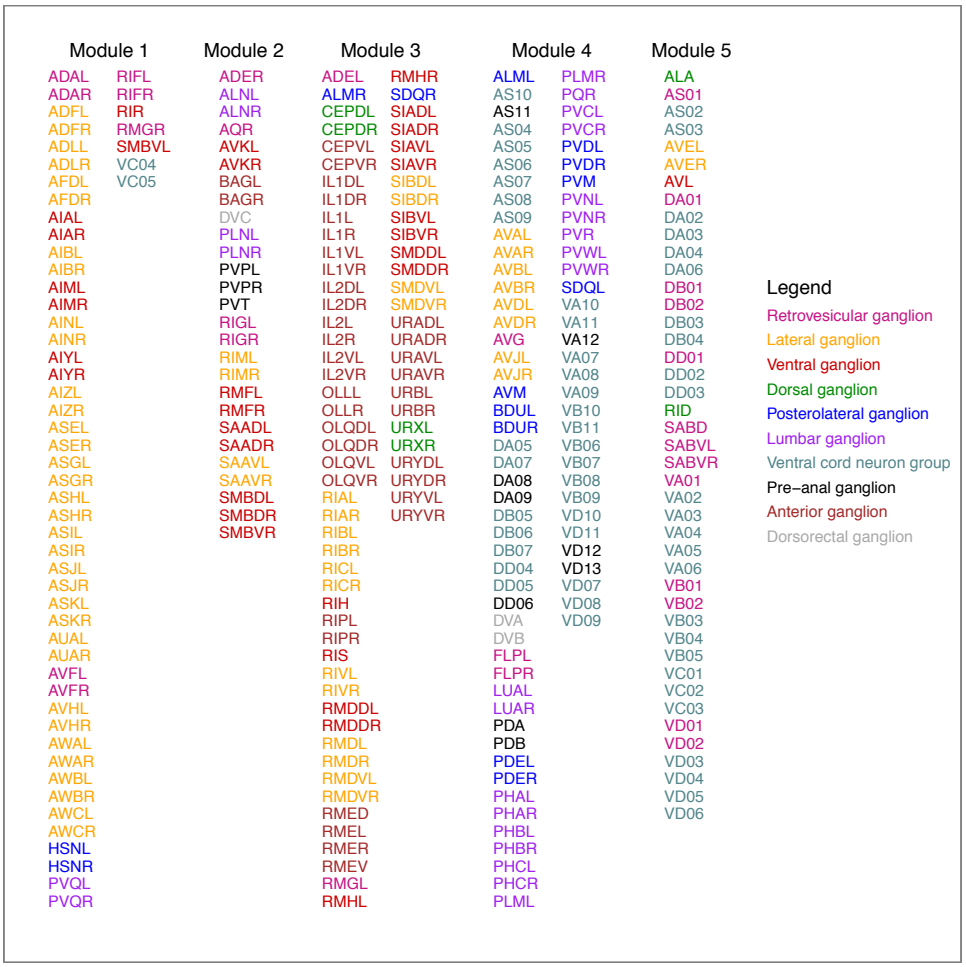
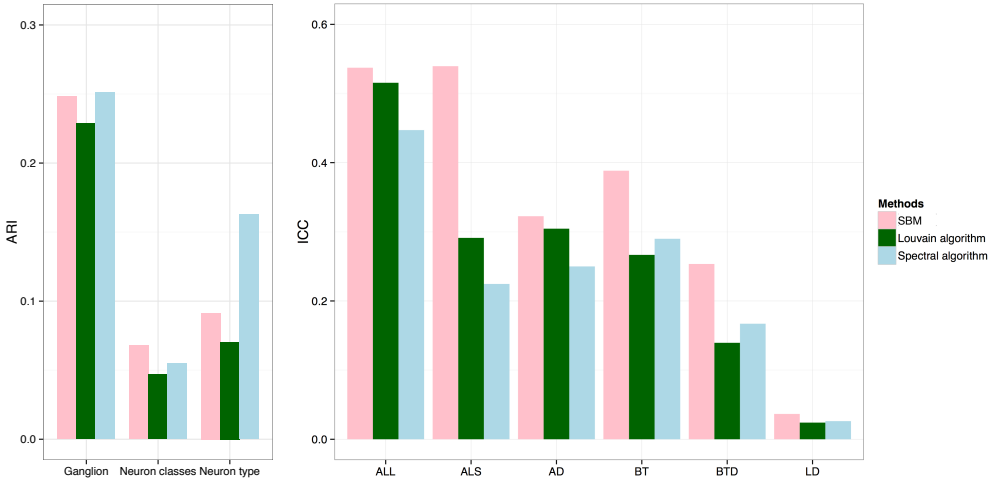


Figure 3.10: **Membership structure of the neurons in the Louvain fit**. Neurons are coloured to match their ganglion type.

same function or behaviour might be similar in terms of these quantitative features, so biologically meaningful cluster structures should be composed of groups that are relatively homogeneous in terms of these metrics. We note that the six quantitative biological features were selected due to their availability, but not all are expected to be equally useful or revealing. For example, although one might expect the lineage distance (LD) to be relevant, in fact neurons of the same class (typically involved in the same function) develop around the same time and usually have no immediate common precursors. This is reflected in low ICC scores for all three methods for lineage distance. The birth time (or BTD) is therefore expected to be more representative of function, and this is confirmed by higher ICC scores across all methods, with the SBM showing particularly good performance. Similarly, the anatomical location (especially in the longitudinal direction) is expected to cluster



(a) (b)

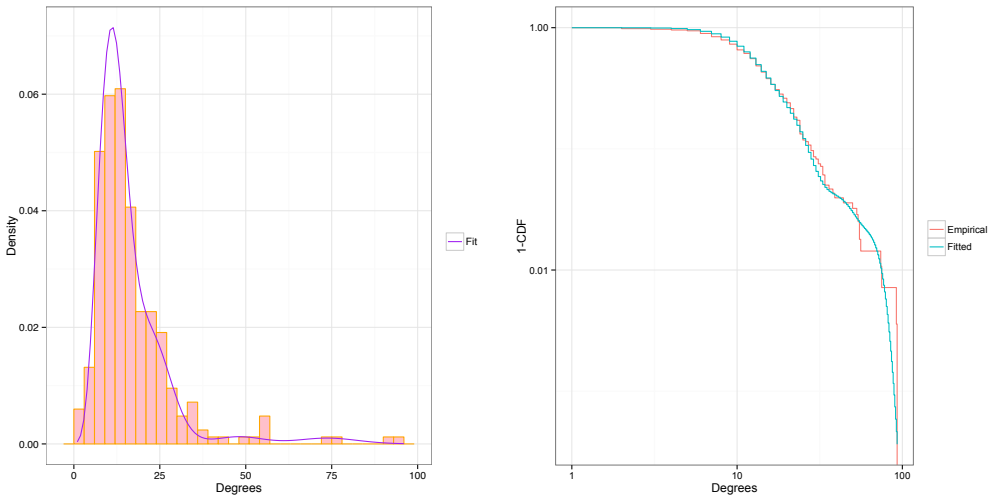
Figure 3.11: **Comparisons in terms of prior biological features**. (a) ARI scores computed between three known biological classifications - Ganglion, Neuron classes (103 groups) and Neuron type (sensory, motor, interneuron and polymodal) - and the fits of each method. Collectively, the ARI scores are small and no greater than 0.26 for all three methods, suggesting that all three solutions are fairly different from the known biological partitions. (b) ICC scores for the Longitudinal anatomical location (ALL), Sectional anatomical location (ALS), Anatomical distance (AD), Birth time (BT), Birth time difference (BTD) and Lineage distance (LD). The ICC results indicate that the SBM partition explains more biological variance than either of the other two methods. Compared to each other, the ICC scores of the Spectral and Louvain fits are largely similar.

functionally related neurons close together. This is because neuronal placement tends to minimise wiring (Alexander-Bloch et al. (2013)) and neurons involved in the same function therefore benefit by being close together both for efficient interconnections between these neurons and because they are likely to be receiving similar (localised) external cues or controlling similar (localised) muscle groups. Again, the SBM shows superior ICC for all the distance related metrics.

Overall, the SBM partition provides the best ICC scores on all six biological features. For example, the SBM partition explains over 50 % of the variance for the sectional anatomical location (ALS) while the other two methods explain only about half as much variability. Also, we note that neither the Louvain or Spectral measures dominate one another on the basis of the ICC scores.

3.3.3 Generative Modelling of the *C. elegans* Brain Network with Stochastic Blockmodel

In addition to estimating the cluster structure of a network, the SBM also provides a generative model of the *C. elegans* nervous system which provides estimates of other network characteristics such as the empirical clustering coefficient (Eq. (2.3)) and degree distribution. Figure 3.12 shows the observed and fitted degree distribution, demonstrating that the SBM provides a faithful approximation of the empirical degree distribution. The fit is based on a Poisson mixture (Eq. (2.34)), and Table 3.3 gives the estimated Poisson means ($\hat{\boldsymbol{\lambda}}$) and weights ($\hat{\boldsymbol{\alpha}}$). Notably, Block 6 (command neurons) has the greatest connectivity with an average degree of 74.23.



(a) Histogram of observed degrees and fit.

(b) 1-CDF plot of observed degrees and fit on log scale.

Figure 3.12: **Observed and SBM-based fit of the degree distribution**. (a) Histogram of the empirical distribution with the SBM fit. (b) Complementary cumulative distribution function (CDF) (i.e. 1-CDF) of the degrees and SBM fit on the log-log scale. The SBM-fitted distribution captures the large-scale features of the degree distribution extremely well, as well as most of the fine-scale features.

Block	1	2	3	4	5	6	7	8	9
$\hat{\boldsymbol{\alpha}}$	0.15	0.11	0.17	0.12	0.02	0.02	0.09	0.26	0.05
$\hat{\boldsymbol{\lambda}}$	12.69	25.48	8.46	10.54	48.53	74.23	22.16	13.26	17.59

Table 3.3: **Poisson mixture parameters $\hat{\boldsymbol{\lambda}}$ and mixture weights $\hat{\boldsymbol{\alpha}}$ in the SBM.**

The model-based clustering coefficient from the SBM is $\hat{C} = 0.154$, which

is somewhat less than the empirical clustering coefficient $C = 0.213$. The likely explanation for this difference is that the SBM specifies each edge as an independent Bernoulli trial (edges are formed with a given probability, independently from one another), which may underestimate the actual rate at which the triangles occur (two neurons connected to the same neighbour are also likely to connect to each other). To assess this, we conducted a small simulation, creating 100 adjacency matrices that followed the SBM assumptions, using the *C. elegans* estimated parameters $\hat{\boldsymbol{\alpha}}$ and $\hat{\boldsymbol{\pi}}$ as the baseline truth. Based on these 100 realisations, the two clustering coefficients were quite similar, with $\hat{C} = 0.152$ (standard deviation 0.005) and $C = 0.154$ (standard deviation 0.004), verifying that $\hat{C}$ is a reasonable estimate of C when the SBM assumptions are satisfied.

3.4 Discussion

Our results highlight the advantages in the use of the SBM over the deterministic clustering algorithms. The SBM decomposed the network into an interpretable set of 9 blocks, comprising 2 small blocks that correspond to the command interneurons (cores), and 7 larger blocks that approximately correspond to the modules defined by the deterministic algorithms. Contrary to the general expectations and conclusions of the previous literature (Meunier et al., 2010; Bullmore and Sporns, 2009; Rubinov and Sporns, 2010; Pan et al., 2010; Sporns et al., 2004; Hilgetag et al., 2000; Achard et al., 2006; Bassett and Bullmore, 2009), our analysis showed that the 'cores-in-modules' decomposition of the worm brain network estimated by the SBM was more compatible with prior biological knowledge about the *C. elegans* nervous system than the purely modular decompositions defined deterministically. Furthermore, the 'cores-in-modules' fit topologically captured both aspects of brain's functional hierarchy, as the modular part reflects functional segregation, while the densely connected core clearly reflects functional integration. Considering also the work of Towlson et al. (2013) which points to the command interneurons of the *C. elegans* nervous system as a topological rich club, it seems that the SBM decomposition has been able to capture both modular and 'core-periphery' aspects (Borgatti and Everett, 2000; Holme, 2005) of the topological organization of the network. This conceptual scope, which can reconcile modular and 'core-periphery' views of cluster structure, is a clear advantage of the SBM compared to deterministic algorithms which are limited to an exclusive selection of one form over the other.

The block decomposition of the SBM was also more successful at accounting for prior biological data than the Spectral and Louvain algorithms. Using the ICC metric to quantify the percentage of variance in a biological variable that is explained by any cluster structure, we found that the SBM decomposition accounted for more

than 50% of the variance for the anatomical location of the neurons, and more than 20% of the variance for the anatomical distance of connections between neurons and neuronal birth times. Also, the SBM explained a greater proportion of the variance in all biological variables than either of the deterministic algorithms. Although, in a more general context, our analysis has demonstrated that the cluster structure can indeed serve as the optimal summary statistic, it is essential that the chosen approach is rich enough to detect a wide range of different cluster structures. In that aspect, the SBM has showed to be reliable and free from the assumptions about a particular type of cluster structure, which is not the case with other deterministic clustering methods. This aspect is especially valuable in the analysis of human data in which the ground truth is generally unknown and, therefore, it is very difficult to benchmark the resulting cluster estimates.

In addition to its advantages as an estimator, the SBM also has technical advantages as a generative model (to simulate the network) and as a network compression basis to shrink the scale of a system. For example, we showed that the SBM generated a good fit of the degree distribution and clustering coefficient of the *C. elegans* network. We also illustrated how the SBM could be used to compress a network into a set of super nodes, allowing a clearer view of the topology with fewer connections. In this sense, the SBM provides a compression of a network similar to power network analysis methods (Royer et al., 2008), but it relaxes the condition for grouping nodes together, which allows for a more efficient and realistic network compression.

We have found some shortcomings of the SBM. For example, the mismatch between the empirical and model based clustering coefficients suggests that the stochastic model does not exactly match the data generating mechanism represented by the *C. elegans* nervous system. The SBM can be extended by seeing it as a mixture Exponential Random Graph Model (ERGM) (Vu et al., 2012) where, conditional on the assignment, the ERGM summary statistics are the edge counts in each of the $Q(Q+1)/2$ unique block pairs. Additional summary statistics can then be considered; for example, the triangle counts in each group pair. However, this will create a more complicated likelihood and necessitate new and yet more involved estimation procedures.

Nevertheless, the general practical advantage of the SBM is that it leaves room for other distributional characterisations of edges which appear to be more in agreement with the network's specific type. Thus, for example, if this approach is used for the analysis of the weighted *C. elegans* network (i.e. with edge weights corresponding to the total number of synaptic connections between a neuron pair), then the assumption that the edges are following a Binomial distribution would be more appropriate. Furthermore, given that our study have used only a simple

unweighted *C. elegans* network, it is interesting to compare our results to the 6 modules decomposition of the weighted *C. elegans* network reported by Pan et al. (2010) who used a modified version of the Spectral algorithm. The corresponding extended results of this comparison (Figure 3.13) show that the SBM decomposition, again, explained more variance in the prior biological information, with the exception of the lineage distance where the SBM explained 1.5% less than the 6 modules decomposition (Figure 3.13).

Lastly, we found that the SBM computational times are reasonable and depend on the range of blocks, the values of the internal parameters (`nbiter`, `fpnbiter`) and, also, the size of the network. It has been reported in Daudin et al. (2008) that this approach can handle networks with several thousands of vertices, which is particularly impressive given the challenging likelihood optimisation. However, the problem of finding the global maximum is heavily dependent on the initialisation and, hence, we require restarts in order to carefully search the state space. While we used a cautious approach of running a large number of restarts (100, 000), we found that 1,000 was sufficient to reliably identify the optimal model.

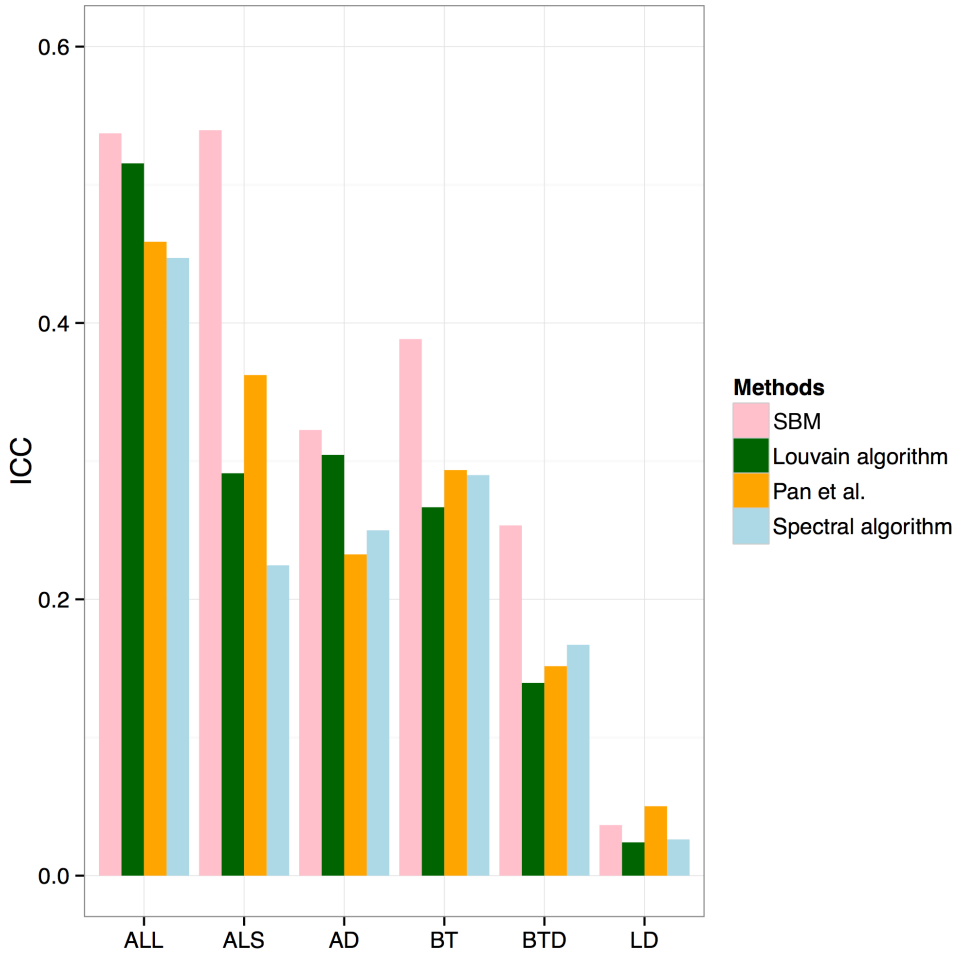


Figure 3.13: **ICC scores of the partitions obtained with the SBM, Louvain and Spectral algorithms on the unweighted *C. elegans* network and the 6 modules partition reported by Pan et al. (2010), who used a modified Spectral algorithm to analyse a weighted *C. elegans* network.** The ICC scores are measured over the range of biological features such as: ALL (Anatomical location longitudinal), ALS (Anatomical location sectional), AD (Anatomical distance), BT (Birth time), BTD (Birth time difference), LD (Lineage distance). The SBM partition obtains the highest ICC scores, apart from the Lineage Distance, where the 6 module decomposition explained more variance. Also, it is interesting to observe the level of similarity between the results of the two Spectral algorithms, where the inclusion of the full weighted network seems to show a clear improvement of the ICC scores on the ALS and LD, but not on the other biological features.

CHAPTER 4

Generalised Linear Stochastic Blockmodelling and Inference in Multi-subject Networks

In the analysis of human brain connectivity data, one of the most significant current discussions is how to make comparisons between the brains in a state of health and pathology. On these lines, Van Den Heuvel and Pol (2010) highlighted that such comparative network studies could make valuable contributions to the understanding of mind disorders like schizophrenia, for which it has been long hypothesised that it is a disease characterised by a lack of connections between brain regions, especially those regions which overlap with the Default Mode Network. In addition to this, numerous other studies (Weinberger et al., 2001; Ellison-Wright et al., 2008; Kubickia et al., 2007; Kubicki et al., 2005) reported altered white matter integrity in schizophrenic patients, particularly in the cingulum tract which is known to mediate connections between the components of the Default Mode Network. These findings were also corroborated by Bassett et al. (2008), whose large multi-subject study (involving 259 healthy subjects and 203 patients with schizophrenia) suggested numerous between-group differences in local clustering properties, including cingulate and insular cortex. Other authors (Liang et al., 2006; Liu et al., 2008; Micheloyannis et al., 2006) also suggested that functional disconnectivity in schizophrenia can be reflected in topological alterations of functional brain networks, which is one of the reasons why clustering approaches could be of potential value. However, despite the high necessity for such comparative network studies, up until now, there has not been much progress in extending single network clustering models to multi-subject network models. The reasons for this can be found in a set of challenges brought in by the multi-subject nature of the data, including:

(i) The need to estimate a common network decomposition over subjects while accounting for between-subject variability in connectivity rates.

(ii) How to use such a network decomposition to infer differences between populations (e.g., patients vs. controls) or effects of covariates.

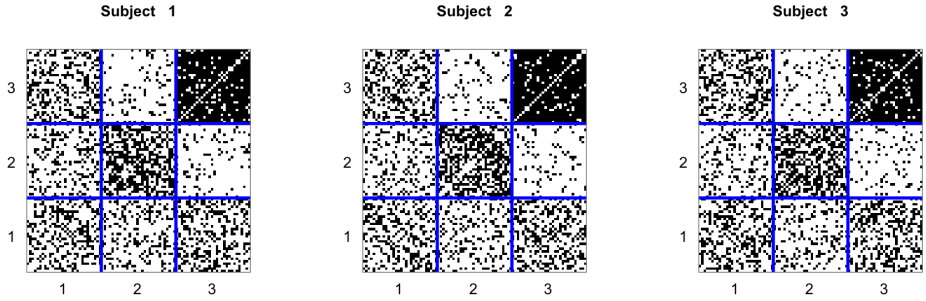
In this chapter, we address these questions by developing a Stochastic Blockmodel for multi-subject binary network data. Our approach is inspired by the extensions of Mariadassou et al. (2010) (see Section 2.4.2) who showed how an edge-based covariate can be used to estimate the cluster structure in a single network whose edges are assumed to follow a mixture of Poisson densities. While those authors discussed other types of distributions as well as potential examples of such network data, it is important to stress out that their model is appropriate only for a single network and situations where a particular edge-based covariate is a significant clustering factor. The level of importance of such covariates is not investigated through a formal hypothesis test but, instead, it is gauged by the complexity of the cluster structure before and after the inclusion of a covariate in the model. This is justified by the fact that the cluster structure is estimated on the residuals from an underlying generalised linear model and, thus, if the addition of a particular edge-based covariate leads to a less complex cluster structure (e.g., larger block sizes and a smaller number of blocks), then this is taken as an indicator of the covariate significance. It is also interesting to take note of the fact that in both their real data analyses and simulation experiments, the authors only investigated the 'homogenous effect' version of their model (see Eq. (2.37)) and, while the 'heterogenous effect' model (see Eq. (2.38)) was formally defined, the behaviour of this model and the domains of its applications were not investigated neither in the real data analysis nor in the simulations.

In contrast to the type of network data considered in Mariadassou et al. (2010), our applications are concerned with multi-subject data, in which each of K subjects in a fMRI study has its own binary and undirected network (see Sections 2.1.3 and 2.1.4). Thus, direct applications of any version of the SBM discussed in Sections 2.4.1 and 2.4.2 would open up two possible choices for the analysis. The first option, allowing the most inter-subject variability, would be to fit an SBM model on each subject in the study. However, this approach would be cumbersome as each subject could have a different cluster structure and, therefore, there would be no clear way for making comparisons between the fitted cluster structure estimates. The second option would be to average the data across all subjects and then fit an SBM on a representative binary and undirected network. However, this approach would tend to overlook the variability between subjects and, in addition, it would be still difficult to include and compare different groups of subjects.

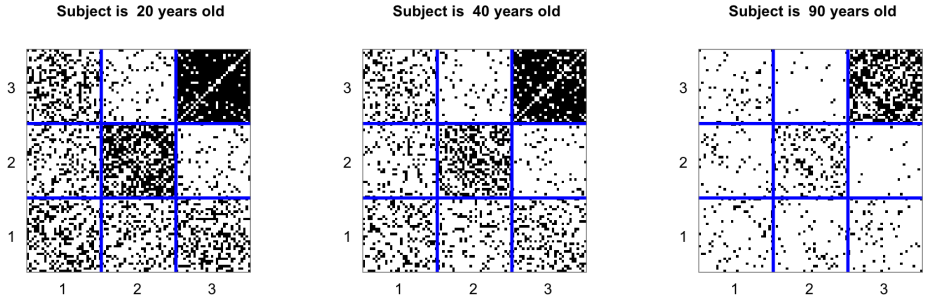
To strike a balance between these two approaches, we consider several extensions of the SBM that are built upon the assumption that the cluster structure is common to all subjects in the study and the variability in connection rates between subjects is modelled through subject-specific covariates in a logistic regression model. It is worth highlighting that our approach does not focus on edge-based covariates like the models in Mariadassou et al. (2010) but on subject-specific covariates, making these models a separate and a novel class. Depending on the type of interaction between the subject-specific covariates and the cluster structure in the data, we propose three multi-subject extensions of the SBM: the Binomial SBM (Bin-SBM), the Homogeneous SBM (Hom-SBM) and the Heterogenous SBM (Het-SBM), where the last two models are a combination of an SBM and a logistic regression model for binary network data. Our approach allows the inclusion of subject specific covariates (e.g., age, gender, patient/control) to model the variability of connectivity between subjects while estimating the common cluster structure in the multi-subject networks. In order to understand the conceptual differences between these models, we introduce three simple running examples, each of which is featuring three binary and undirected networks with 90 nodes which are decomposed into three blocks and the assignment of nodes is common across all subjects, whose ages are 20, 40 and 90 years.

As shown in Figure 4.1 (a), the Bin-SBM does not use the subject's age to model the variability in their individual cluster structures. Although there are some minor differences in the concentration of edges across the subjects, this variability is only modelled by the variance in that particular block. For example, the intensity of connections in the block $(2, 3)$ is varying across the subjects according to its binomial variance. For the Hom-SBM, the effect of age is decreasing across the entire cluster structure for all subjects (see Figure 4.1 (b)) and, consequently, the variability of the cluster structure is tuned to tolerate minor variations across the subjects. In contrast, the Het-SBM allows the effect of age in each block to decrease or increase the connectivity independently and, because of this, we can see different types of decompositions across the subjects (see Figure 4.1 (c)). For example, the first subject has a modular structure, while the third subject exhibits a structure which is generally known as 'disassortative mixing' (Hu and Wang, 2009).

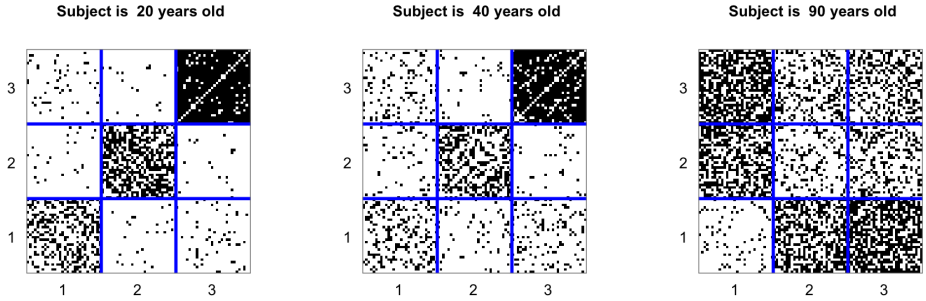
In general, depending on the multi-subject model and the goals of the study, the covariates can be regarded as nuisance or we can conduct hypothesis tests on the estimated regression coefficients. For example, we may be interested in estimating a cluster structure while controlling for different groups of subjects or other factors which are not necessarily study objectives, and, in other situations, we may be only interested in a group analysis or in determining the significance of a particular covariate.



(a) Binomial SBM (Bin-SBM)



(b) Homogeneous SBM (Hom-SBM)



(c) Heterogenous SBM (Het-SBM)

Figure 4.1: **Conceptual differences between three multi-subject models**. Each subject's network is given as a reorganised adjacency matrix comprised of three blocks, labelled numerically (1-3) on both x and y axis. (a) The Binomial SBM (Bin-SBM) does not use the subject's age to model the variability between the three subjects. (b) In the Homogenous SBM (Hom-SBM), the effect of age is seen as an overall decrease of connectivity with increasing age. The direction of the covariate effect is the same across all blocks and all subjects. (c) In the Heterogenous SBM (Het-SBM), the effect of age is seen as an increase or decrease of connectivity in each block. For example, the connectivity in Block $(1, 1)$ is decreasing across subjects and, in Block $(1, 2)$, the connectivity is increasing across the subjects.

The general estimation framework of the multi-subject models follows the structure of the classical SBM described in Section 2.4.1, including the variational

estimation and model selection with the ICL criterion. In addition to this, in the Hom-SBM, the estimation of the regression coefficients is based on a classical maximum likelihood procedure with the exception of the intercepts for which we use Firth type estimates (Firth, 1993; Heinze and Schemper, 2002) in order to correct for small-sample bias. As we will see later in more details in Section 4.1.2 the Firth type estimates are only applied to the block level parameters (i.e. the intercepts in the Hom-SBM) and, for the slope parameters (i.e. the global network parameters), we use Ordinary Maximum Likelihood Estimates (MLEs). In contrast to this, in the case of the Het-SBM, we only use Firth type estimates as all its regression parameters are entirely block dependent.

The remainder of this chapter is organised as follows. First, we describe the three multi-subject Stochastic Blockmodels (Bin-SBM, Hom-SBM and Het-SBM), along with their estimation and inference strategies. Then, we describe the Monte Carlo simulations that are used to assess the quality of the parameter estimates. For this, we also make comparisons between the estimates based on the Ordinary and Firth MLEs. After this, we describe the multi-subject resting state fMRI dataset to which we fit the Het-SBM, and we conclude with a discussion and future work.

4.1 Multi-subject Stochastic Blockmodels

In this section, we use a binary random variable X_{ijk} to denote the presence/absence of the k-th subject's edge between the nodes V_i and V_j. For each of K subjects ($k = 1, \ldots, K$) we use $\boldsymbol{X}_k = ((X_{ijk}))_{1 \leq i \neq j \leq n}$ to denote its $n \times n$ random and symmetric adjacency matrix and, we use $\boldsymbol{X}$ to denote the set of independent random matrices such that $\boldsymbol{X} = \{\boldsymbol{X}_1, \ldots, \boldsymbol{X}_K\}$. The individual random networks $\boldsymbol{X}_k$ are assumed to be simple, i.e. without self-connected nodes, and undirected. To be consistent with the general notation introduced in Section 2.4, the realisations of these random variables are written in lower cases.

Since the goal of each model is to estimate a common cluster structure among K subjects, the assumptions on the nodes are similar to those already discussed in Section 2.4.1. For the sake of completeness, we will briefly restate these assumptions. In particular, the set of nodes is assumed to be allocated into one of Q unknown (latent) blocks and, thus, for a node V_i, there exists a $1 \times Q$ dimensional random vector $\boldsymbol{Z}_i = (Z_{i1}, \ldots, Z_{iQ})$ whose element Z_{iq} takes value 1 if V_i is in the q-th group and 0 otherwise. Finally, we assume that these assumptions hold for each subject. Formally, the random variables in the $n \times Q$ matrix $\boldsymbol{Z}$ are assumed to be independent and to follow a Categorical distribution with Q possible outcomes,

$$\boldsymbol{Z}_i \sim \text{ Categorical}(Q, \boldsymbol{\alpha}), \tag{4.1}$$

with individual probabilities of success $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_Q)$ such that $\sum_{q=1}^{Q} \alpha_q = 1$. Hence, the probability mass function of $\boldsymbol{Z}$ can be noted as

$$f(\boldsymbol{z}; \boldsymbol{\alpha}) = \prod_{q=1}^{Q} \prod_{i=1}^{n} \alpha_q^{z_{iq}}. \tag{4.2}$$

4.1.1 The Binomial Stochastic Blockmodel (Bin-SBM)

In this model, we assume that all subjects have the same expected edge frequencies (see Figure 4.1 (a)). Consequently, for the k-th subject, the edges are assumed to follow a Bernoulli distribution

$$X_{ijk} | Z_{iq} = 1, Z_{jl} = 1 \ \sim \ \text{Bernoulli}(\pi_{ql}), \tag{4.3}$$

where π_{ql} is the connectivity rate, expressing the probability that the nodes in block (q, l) are connected. As the connectivity rates are taken to be constant across the subjects, the random variable $X_{ij} = \sum_{k}^{K} X_{ijk}$ follows a Binomial distribution,

$$X_{ij} | Z_{iq} = 1, Z_{jl} = 1 \ \sim \ \text{Binomial}(K, \pi_{ql}). \tag{4.4}$$

In the context of the Bin-SBM, $\boldsymbol{X} = ((X_{ij}))_{1 \leq i \neq j, \leq n}$. Observe that this trivial model is equivalent to a single network model with binomially distributed edges whose values indicate the total number of observed edges across K subjects. The latter model has already been considered in Mariadassou et al. (2010), but, for the sake of completeness, we give below the details of its estimation and model selection procedure.

Estimation and Model Selection

The optimisation strategy follows the variational approximation already given in Section 2.4.1. The model parameters are estimated by maximising the variational bound

$$\begin{aligned}
\mathcal{J}(f^*(\boldsymbol{z}; \boldsymbol{\tau}); \boldsymbol{\pi}, \boldsymbol{\alpha}) &= \mathbb{E}\big[\log f(\boldsymbol{x}, \boldsymbol{Z}; \boldsymbol{\pi}, \boldsymbol{\alpha})\big] - \mathbb{E}\big[\log f^*(\boldsymbol{Z}; \boldsymbol{\tau})\big] \\
&= \frac{1}{2} \sum_{i=1}^{n} \sum_{j \neq i}^{n} \sum_{q,l}^{Q} \tau_{iq} \tau_{jl} \log \left[\binom{K}{x_{ij}} \pi_{ql}^{x_{ij}} (1 - \pi_{ql})^{K - x_{ij}} \right] \\
&\quad + \sum_{i=1}^{n} \sum_{q=1}^{Q} \tau_{iq} \log[\alpha_q] - \sum_{i=1}^{n} \sum_{q=1}^{Q} \tau_{iq} \log[\tau_{iq}],
\end{aligned} \tag{4.5}$$

with respect to the variational parameters $\boldsymbol{\tau}$ and the model parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\pi}$, subject to the constrains $\sum_{q=1}^{Q} \tau_{iq} = 1$ and $\sum_{q=1}^{Q} \alpha_q = 1$. After maximisation, the

optimal variational parameter $\hat{\boldsymbol{\tau}}$ and the optimal model parameters $\hat{\boldsymbol{\alpha}}$ and $\hat{\boldsymbol{\pi}}$ satisfy the fixed point relations

$$\hat{\tau}_{iq} \propto \hat{\alpha}_q \prod_{j \neq i}^{n} \prod_{l=1}^{Q} [\hat{\pi}_{ql}^{x_{ij}} (1 - \hat{\pi}_{ql})^{K - x_{ij}}]^{\hat{\tau}_{jl}}, \tag{4.6}$$

$$\hat{\alpha}_q = \frac{1}{n} \sum_{i=1}^{n} \hat{\tau}_{iq}, \tag{4.7}$$

$$\hat{\pi}_{ql} = \frac{\sum_{i=1}^{n} \sum_{j \neq i}^{n} \hat{\tau}_{iq} \hat{\tau}_{jl} x_{ij}}{K \sum_{i=1}^{n} \sum_{j \neq i}^{n} \hat{\tau}_{iq} \hat{\tau}_{jl}}. \tag{4.8}$$

As before, the estimates of the classification vector $\hat{\boldsymbol{z}}_i$ are found by

$$\hat{z}_{iq} = \begin{cases} 1 & \text{if } q = \text{argmax}_{q'} \, \hat{\tau}_{iq'} \\ 0 & \text{otherwise.} \end{cases} \tag{2.25 revisited}$$

In order to estimate the optimal number of blocks, we use the ICL criterion described in Section 2.4.1, given as

$$\begin{aligned} \text{ICL}(\text{m}_Q) = & \log f(\boldsymbol{x}, \hat{\boldsymbol{z}} | \text{m}_Q, \hat{\boldsymbol{\pi}}, \hat{\boldsymbol{\alpha}}) - \frac{1}{2} \frac{Q(Q+1)}{2} \log \left[\frac{n(n-1)}{2} \right] \\ & - \frac{Q-1}{2} \log[n]. \end{aligned} \tag{4.9}$$

Hence, the optimal number of blocks (Q) in a fitted model is found by selecting the model which maximises the ICL score.

4.1.2 The Homogeneous Stochastic Blockmodel (Hom-SBM)

In this model, the variability between subjects is assumed to be a global feature of the observed networks (see Figure 4.1 (b)). Thus, conditional on the node assignments, the edges are assumed to follow a Bernoulli distribution, whose rates of connection depend on the subject covariates via a logistic regression model

$$X_{ijk} | Z_{iq} = 1, Z_{jl} = 1 \ \sim \ \text{Bernoulli}(\pi_{qlk}), \tag{4.10}$$

$$\log \left(\frac{\pi_{qlk}}{1 - \pi_{qlk}} \right) = \theta_{ql} + \boldsymbol{d}_k^\top \boldsymbol{\beta}, \tag{4.11}$$

where θ_{ql} is the intercept of the block (q, l), $\boldsymbol{d}_k^\top$ is a $1 \times P$ dimensional vector of covariates associated with the k-th subject and $\boldsymbol{\beta}$ is the $P \times 1$ vector of regression parameters.

Estimation and Model Selection

In this model, we use the variational optimisation strategy to estimate the parameters $(\boldsymbol{\tau}, \boldsymbol{\alpha}, \boldsymbol{\theta}, \boldsymbol{\beta})$ with a Newton-Raphson algorithm for the parameters $(\boldsymbol{\theta}, \boldsymbol{\beta})$. In addition to this, we also consider a Firth type estimation (Firth, 1993) for the intercept parameters in $\boldsymbol{\theta}$, which prevents small sample biases that may easily occur in a combination of several practical situations. First, the number of data points pertaining to the block (q, l) can be very small because either the number of nodes in the block is small or the number of subjects in the study is small. Second, even if the sample size is moderate, there may be situations in which the covariate values can be completely divided between two parts, one which corresponds to the presence of edges and the other which corresponds to the absence of edges. For example, let us suppose that we have a block with 10 nodes in which we observe only edges for the subjects who are smokers. In such a scenario, we would get a convergent likelihood but the MLE of θ_{ql} would diverge. This type of behaviour has been well documented in the literature (Albert and Anderson, 1984; Santner and Duffy, 1986; Lesaffre and Albert, 1989; Hirji et al., 1989; Clarkson and Jennrich, 1991; Kolassa, 1997; Lesaffre and Marx, 1993) and it is broadly labeled as the problem of separation. In particular, Heinze and Schemper (2002) showed that the Firth type estimates are also effective against this problem as they introduce a shrinkage to the estimate.

As noted above, the starting point for the estimation of the parameters is the variational bound,

$$\begin{aligned}\mathcal{J}(f^*(\boldsymbol{z};\boldsymbol{\tau});\boldsymbol{\alpha},\boldsymbol{\theta},\boldsymbol{\beta}) = \frac{1}{2}\sum_{k=1}^{K}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\sum_{q,l}^{Q}\tau_{iq}\tau_{jl}\log[f(x_{ijk}|z_{iq},z_{jl};\theta_{ql},\boldsymbol{\beta})] \\ + \sum_{i=1}^{n}\sum_{q=1}^{Q}\tau_{iq}\log[\alpha_q] - \sum_{i=1}^{n}\sum_{q=1}^{Q}\tau_{iq}\log[\tau_{iq}],\end{aligned} \tag{4.12}$$

where

$$f(x_{ijk}|z_{iq},z_{jl};\theta_{ql},\boldsymbol{\beta}) = \left(\pi_{qlk}\right)^{x_{ijk}}\left(1-\pi_{qlk}\right)^{1-x_{ijk}}, \tag{4.13}$$

$$\pi_{qlk} = \frac{e^{\theta_{ql}+\boldsymbol{d}_k^\top\boldsymbol{\beta}}}{1+e^{\theta_{ql}+\boldsymbol{d}_k^\top\boldsymbol{\beta}}}. \tag{4.14}$$

Similarly to the previous model, we obtain the point estimating equations for $\hat{\boldsymbol{\tau}}$ and $\hat{\boldsymbol{\alpha}}$ by maximising the variational bound in Eq. (4.12), which yields

$$\hat{\tau}_{iq} \propto \hat{\alpha}_q \prod_{k=1}^{K}\prod_{j\neq i}^{n}\prod_{l=1}^{Q}\left[\frac{e^{x_{ijk}(\hat{\theta}_{ql}+\boldsymbol{d}_k^\top\hat{\boldsymbol{\beta}})}}{1+e^{\hat{\theta}_{ql}+\boldsymbol{d}_k^\top\hat{\boldsymbol{\beta}}}}\right]^{\hat{\tau}_{jl}}, \tag{4.15}$$

$$\hat{\alpha}_q = \frac{1}{n}\sum_{i=1}^{n}\hat{\tau}_{iq}. \tag{4.16}$$

However, in order to obtain the estimates of the parameters in the logistic regression model $(\boldsymbol{\theta}, \boldsymbol{\beta})$, we use the Newton-Raphson algorithm. In what follows next, we first fix a general notation and we state the implicit forms of various preliminary quantities that are needed for the subsequent discussion. Thus, we note that the first order partial derivatives of the variational bound $\mathcal{J}(f^*(\boldsymbol{z};\boldsymbol{\tau});\boldsymbol{\alpha},\boldsymbol{\theta},\boldsymbol{\beta})$ with respect to θ_{ql} and $\boldsymbol{\beta}$ as $\mathrm{U}(\theta_{ql})$ and $\mathbf{U}(\boldsymbol{\beta})$ and note that they are explicitly given as

$$\mathrm{U}(\theta_{ql}) = \frac{1}{2}\sum_{k=1}^{K}\sum_{i\neq j}^{n}\sum_{j\neq i}^{n}\tau_{iq}\tau_{jl}(x_{ijk}-\pi_{qlk}), \tag{4.17}$$

$$\mathbf{U}(\boldsymbol{\beta}) = \frac{1}{2}\sum_{k=1}^{K}\sum_{i\neq j}^{n}\sum_{j\neq i}^{n}\sum_{q,l}^{Q}\tau_{iq}\tau_{jl}(x_{ijk}-\pi_{qlk})\boldsymbol{d}_k. \tag{4.18}$$

Similarly, we note the negative second order partial derivatives of the variational bound yield the observed Fisher Information matrix $\boldsymbol{\mathcal{I}}(\boldsymbol{\theta},\boldsymbol{\beta})$. In particular, $\boldsymbol{\mathcal{I}}(\boldsymbol{\theta},\boldsymbol{\beta})$ consists of three sub-matrices

$$\boldsymbol{\mathcal{I}}(\boldsymbol{\theta},\boldsymbol{\beta}) = \left[\begin{array}{c|c} \boldsymbol{\mathcal{I}}_1(\boldsymbol{\theta}) & \boldsymbol{\mathcal{I}}_2^{\top}(\boldsymbol{\theta},\boldsymbol{\beta}) \\ \hline \boldsymbol{\mathcal{I}}_2(\boldsymbol{\theta},\boldsymbol{\beta}) & \boldsymbol{\mathcal{I}}_3(\boldsymbol{\beta}) \end{array}\right], \tag{4.19}$$

such that

- $\boldsymbol{\mathcal{I}}_1(\boldsymbol{\theta})$ is a diagonal $Q(Q+1)/2 \times Q(Q+1)/2$ matrix,
- $\boldsymbol{\mathcal{I}}_2(\boldsymbol{\theta},\boldsymbol{\beta})$ is a $P \times Q(Q+1)/2$ matrix, and
- $\boldsymbol{\mathcal{I}}_3(\boldsymbol{\beta})$ is a $P \times P$ matrix.

Below, we give their specific elements

$$\mathcal{I}_1(\theta_{ql},\theta_{q'l'}) = \begin{cases} \frac{1}{2}\sum_{k=1}^{K}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\tau_{iq}\tau_{jl}\pi_{qlk}(1-\pi_{qlk}) & \text{if } q,l = q',l' \\ 0 & \text{otherwise ,} \end{cases} \tag{4.20}$$

$$\boldsymbol{\mathcal{I}}_2(\theta_{ql},\boldsymbol{\beta}) = \frac{1}{2}\sum_{k=1}^{K}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\tau_{iq}\tau_{jl}\pi_{qlk}(1-\pi_{qlk})\boldsymbol{d}_k \tag{4.21}$$

$$\boldsymbol{\mathcal{I}}_3(\boldsymbol{\beta}) = \frac{1}{2}\sum_{k=1}^{K}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\sum_{q,l}^{Q}\tau_{iq}\tau_{jl}\pi_{qlk}(1-\pi_{qlk})\ \boldsymbol{d}_k\boldsymbol{d}_k^{\top}. \tag{4.22}$$

However, in order to obtain the Firth type estimates of $\boldsymbol{\theta}$, we want to maximise

$$\mathcal{J}^*(f^*(\boldsymbol{z};\boldsymbol{\tau});\boldsymbol{\alpha},\boldsymbol{\theta},\boldsymbol{\beta}) = \mathcal{J}(f^*(\boldsymbol{z};\boldsymbol{\tau});\boldsymbol{\alpha},\boldsymbol{\theta},\boldsymbol{\beta}) + \frac{1}{2}\log\Big[\mathrm{Det}(\boldsymbol{\mathcal{I}}(\boldsymbol{\theta},\boldsymbol{\beta}))\Big], \tag{4.23}$$

whose partial derivatives $U^*(\theta_{ql})$ are given as

$$U^*(\theta_{ql}) = U(\theta_{ql}) + \frac{1}{2}\mathrm{Tr}\Big[\boldsymbol{\mathcal{I}}^{-1}(\boldsymbol{\theta},\boldsymbol{\beta})\frac{\partial}{\partial\theta_{ql}}\,\boldsymbol{\mathcal{I}}(\boldsymbol{\theta},\boldsymbol{\beta})\Big], \tag{4.24}$$

and the derivatives of the observed Fisher Information matrix are given as

$$\frac{\partial\mathcal{I}_1(\theta_{q'l'},\theta_{q''l''})}{\partial\theta_{ql}} = \begin{cases} \frac{1}{2}\sum_{k=1}^{K}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\tau_{iq}\tau_{jl}\pi_{qlk}(1-\pi_{qlk})(1-2\pi_{qlk}) & q,l=q',l'=q'',l'' \\ 0 & \text{otherwise} \end{cases} \tag{4.25}$$

$$\frac{\partial\boldsymbol{\mathcal{I}}_2(\boldsymbol{\theta},\boldsymbol{\beta})}{\partial\theta_{ql}} = \frac{1}{2}\sum_{k=1}^{K}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\tau_{iq}\tau_{jl}\pi_{qlk}(1-\pi_{qlk})(1-2\pi_{qlk})\boldsymbol{d}_k, \tag{4.26}$$

$$\frac{\partial\boldsymbol{\mathcal{I}}_3(\boldsymbol{\beta})}{\partial\theta_{ql}} = \frac{1}{2}\sum_{k=1}^{K}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\tau_{iq}\tau_{jl}\pi_{qlk}(1-\pi_{qlk})(1-2\pi_{qlk})\boldsymbol{d}_k\boldsymbol{d}_k^\top. \tag{4.27}$$

Finally, the estimating equations for $\boldsymbol{\theta}$ and $\boldsymbol{\beta}$ are found using the Newton-Raphson formula which, for the (m)-th step, updates the parameters according to the following expression

$$\begin{pmatrix}\boldsymbol{\theta}\\ \boldsymbol{\beta}\end{pmatrix}^{(m)} = \begin{pmatrix}\boldsymbol{\theta}\\ \boldsymbol{\beta}\end{pmatrix}^{(m-1)} + \boldsymbol{\mathcal{I}}^{-1}(\boldsymbol{\theta}^{(m-1)},\boldsymbol{\beta}^{(m-1)})\begin{pmatrix}\mathbf{U}^*(\boldsymbol{\theta})\\ \mathbf{U}(\boldsymbol{\beta})\end{pmatrix}^{(m-1)}, \tag{4.28}$$

where $\mathbf{U}^*(\boldsymbol{\theta})$ is the column vector of length $Q(Q+1)/2$ composed of the elements given in Eq. (4.24). It is important to highlight that the $\boldsymbol{\theta}$ parameters are based on the Firth type MLEs and thus they are updated with $\mathbf{U}^*(\boldsymbol{\theta})$, while the $\boldsymbol{\beta}$ parameters are based on the Ordinary MLEs and thus are updated with $\mathbf{U}(\boldsymbol{\beta})$.

Finally, we discuss the ICL criterion. As noted in Section 2.4.1, the ICL criterion is constructed from $\log f(\boldsymbol{x},\boldsymbol{z}|\mathrm{m}_Q) = \log f(\boldsymbol{x}|\boldsymbol{z},\mathrm{m}_Q) + \log f(\boldsymbol{z}|\mathrm{m}_Q)$. The approximation of $\log f(\boldsymbol{z}|\mathrm{m}_Q)$ is the same as in Section 2.4.1, while for $\log f(\boldsymbol{x}|\boldsymbol{z},\mathrm{m}_Q)$, we simply apply the BIC. For this, we note that the total number of parameters is $\frac{Q(Q+1)}{2}+P$, since we have $\frac{Q(Q+1)}{2}$ intercepts in $\boldsymbol{\theta}$ and P regression coefficients in $\boldsymbol{\beta}$, and the total number of data points in $\boldsymbol{x}$ is $\frac{n(n-1)}{2}K$. With this, the ICL criterion is stated as

$$\mathrm{ICL}(\mathrm{m}_Q) = \log f(\boldsymbol{x},\hat{\boldsymbol{z}}|\mathrm{m}_Q,\hat{\boldsymbol{\alpha}},\hat{\boldsymbol{\theta}},\hat{\boldsymbol{\beta}}) - \frac{1}{2}\left(\frac{Q(Q+1)}{2}+P\right)\log\left[\frac{n(n-1)}{2}K\right]$$

$$- \frac{Q-1}{2} \log[n]. \tag{4.29}$$

Estimation Algorithm

In this section, we describe the proposed estimation algorithm. Starting with some initial values for $\boldsymbol{\tau}^0$, we iteratively update the model parameters according to the two steps

1. $(\boldsymbol{\alpha}^{(h+1)}, \boldsymbol{\theta}^{(h+1)}, \boldsymbol{\beta}^{(h+1)}) = \underset{(\boldsymbol{\alpha},\boldsymbol{\theta},\boldsymbol{\beta})}{\operatorname{argmax}} \left[\mathcal{J}(f^*(\boldsymbol{z}; \boldsymbol{\tau}^{(h)}); \boldsymbol{\alpha}, \boldsymbol{\theta}, \boldsymbol{\beta}) + \frac{1}{2} \log \left[\operatorname{Det}(\boldsymbol{\mathcal{I}}(\boldsymbol{\theta}, \boldsymbol{\beta})) \right] \right]$,
2. $\boldsymbol{\tau}^{(h+1)} = \underset{\boldsymbol{\tau}}{\operatorname{argmax}} \left[\mathcal{J}(f^*(\boldsymbol{z}; \boldsymbol{\tau}); \boldsymbol{\alpha}^{(h+1)}, \boldsymbol{\theta}^{(h+1)}, \boldsymbol{\beta}^{(h+1)}) \right]$,

until the convergence is obtained. In the first step, we update $\boldsymbol{\alpha}$ according to Eq. (4.16) and we update $(\boldsymbol{\theta}, \boldsymbol{\beta})$ according to the Newton-Raphson algorithm (see Eq. (4.28)). In the second step, we update $\boldsymbol{\tau}$ according to Eq. (4.15). The convergence is measured by the relative changes of the parameter estimates and the improvement of the variational bound. Note that, in the first step, the parameters $(\boldsymbol{\theta}, \boldsymbol{\beta})$ are updated according to the Newton-Raphson algorithm (see Eq. (4.28)) which is initialised with zero starting values for the parameters $(\boldsymbol{\theta}, \boldsymbol{\beta})$. Also, note that, for each iteration of the Newton-Raphson algorithm, we set a maximal absolute change of value 5 for all the parameters (e.g., if the change for a parameter is estimated to be -7.2, it is forced to be -5). Moreover, when the variational bound is smaller than the value obtained with the previous parameter estimates, a halving procedure is executed. This procedure iteratively reduces the parameter updates by half until an improvement is observed or until a maximal number of halving steps is reached, in which case we keep the previous parameter estimates. Note that the halving step is performed block-wise for the parameter $\boldsymbol{\theta}$, which also indicates that the improvement of the bound is checked block-wise.

4.1.3 The Heterogenous Stochastic Blockmodel (Het-SBM)

In this model, the variability between subjects is assumed to independently influence the connectivity of each block and block-to-block relationships (see Figure 4.1 (c)). Formally this can be written as

$$X_{ijk} | Z_{iq} = 1, Z_{jl} = 1 \ \sim \ \text{Bernoulli}(\pi_{qlk}) \tag{4.30}$$

$$\log \left(\frac{\pi_{qlk}}{1 - \pi_{qlk}} \right) = \boldsymbol{d}_k^\top \boldsymbol{\beta}_{ql}, \tag{4.31}$$

where $\boldsymbol{\beta}_{ql}$ is a $1 \times P$ dimensional vector of regression parameters and, as before, $\boldsymbol{d}_k^\top$ is a $1 \times P$ dimensional vector of the k-th subject's covariates (typically the

first element will be 1, representing the intercept). In total, there are $Q(Q+1)/2$ individual regression vectors $\boldsymbol{\beta}_{ql}$ which we will collectively denote as $\boldsymbol{\beta}$.

Estimation and Model Selection

Similarly to the estimation procedure outlined in Section 4.1.2, in this model, we use the variational optimisation strategy to estimate the parameters $(\boldsymbol{\tau}, \boldsymbol{\alpha}, \boldsymbol{\beta})$ with a Newton-Raphson algorithm for the parameters in $\boldsymbol{\beta}$. However, unlike the Hom-SBM where we used Firth type estimates (Firth, 1993) only for the intercept parameters and not for the global regression coefficients, in this model, we use Firth type estimates for every regression parameter vector $\boldsymbol{\beta}_{ql}$ in $\boldsymbol{\beta}$.

In particular, we note that the variational bound is given as

$$\begin{aligned}\mathcal{J}(f^*(\boldsymbol{z};\boldsymbol{\tau});\boldsymbol{\alpha},\boldsymbol{\beta}) =& \frac{1}{2}\sum_{k=1}^{K}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\sum_{q,l}^{Q}\tau_{iq}\tau_{jl}\log\left[f(x_{ijk}|z_{iq},z_{jl};\boldsymbol{\beta}_{ql})\right] \\ &+\sum_{i=1}^{n}\sum_{q}^{Q}\tau_{iq}\log[\alpha_q]-\sum_{i=1}^{n}\sum_{q}^{Q}\tau_{iq}\log[\tau_{iq}],\end{aligned} \tag{4.32}$$

but, since $\boldsymbol{\beta}_{ql}=\boldsymbol{\beta}_{lq}$, it is convenient to use the notation γ_{ijql}, defined as

$$\gamma_{ijql}=\begin{cases}\tau_{iq}\tau_{jq} & \text{if } q=l \\ \tau_{iq}\tau_{jl}+\tau_{il}\tau_{jq} & \text{if } q<l,\end{cases} \tag{4.33}$$

which points to the equivalence between $\sum_{i=1}^{n}\sum_{j\neq i}^{n}\sum_{q,l}^{Q}\tau_{iq}\tau_{jl}$ and $\sum_{i=1}^{n}\sum_{j\neq i}^{n}\sum_{q\leq l}^{Q}\gamma_{ijql}$ and, hence, the variational bound given in Eq. (4.32) can be also stated as

$$\begin{aligned}\mathcal{J}(f^*(\boldsymbol{z};\boldsymbol{\tau});\boldsymbol{\alpha},\boldsymbol{\beta}) =& \frac{1}{2}\sum_{k=1}^{K}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\sum_{q\leq l}^{Q}\gamma_{ijql}\log[f(x_{ijk}|z_{iq},z_{jl};\boldsymbol{\beta}_{ql})] \\ &+\sum_{i=1}^{n}\sum_{q}^{Q}\tau_{iq}\log[\alpha_q]-\sum_{i=1}^{n}\sum_{q}^{Q}\tau_{iq}\log[\tau_{iq}].\end{aligned} \tag{4.34}$$

Here,

$$f(x_{ijk}|z_{iq},z_{jl};\boldsymbol{\beta}_{ql})=\left(\pi_{qlk}\right)^{x_{ijk}}\left(1-\pi_{qlk}\right)^{1-x_{ijk}}, \tag{4.35}$$

$$\pi_{qlk}=\frac{e^{\boldsymbol{d}_k^\top\boldsymbol{\beta}_{ql}}}{1+e^{\boldsymbol{d}_k^\top\boldsymbol{\beta}_{ql}}}. \tag{4.36}$$

Optimising Eq. (4.32) for $\boldsymbol{\tau}$ and $\boldsymbol{\alpha}$ yields the following point estimating equations

$$\hat{\tau}_{iq} \propto \hat{\alpha}_q \prod_{k=1}^{K} \prod_{j \neq i}^{n} \prod_{l=1}^{Q} \left[\frac{e^{x_{ijk} \boldsymbol{d}_k^\top \hat{\boldsymbol{\beta}}_{ql}}}{1 + e^{\boldsymbol{d}_k^\top \hat{\boldsymbol{\beta}}_{ql}}} \right]^{\hat{\tau}_{jl}}, \tag{4.37}$$

$$\hat{\alpha}_q = \frac{1}{n} \sum_{i=1}^{n} \hat{\tau}_{iq}. \tag{4.38}$$

However, for the same reasons as discussed in Section 4.1.2, we use the Firth type estimates for $\boldsymbol{\beta}_{ql}$ which are based on the optimisation of

$$\mathcal{J}^*(f^*(\boldsymbol{z}; \boldsymbol{\tau}); \boldsymbol{\alpha}, \boldsymbol{\beta}) = \mathcal{J}(f^*(\boldsymbol{z}; \boldsymbol{\tau}); \boldsymbol{\alpha}, \boldsymbol{\beta}) + \frac{1}{2} \sum_{q \leq l}^{Q} \log[\text{Det}(\boldsymbol{\mathcal{I}}_{ql}(\boldsymbol{\beta}_{ql})], \tag{4.39}$$

where $\mathcal{J}(f^*(\boldsymbol{z}; \boldsymbol{\tau}); \boldsymbol{\alpha}, \boldsymbol{\beta})$ is given in Eq. (4.34) and its first order partial derivatives with respect to $\boldsymbol{\beta}_{ql}$ can be written as $\mathbf{U}_{ql}(\boldsymbol{\beta}_{ql})$ such that

$$\mathbf{U}_{ql}(\boldsymbol{\beta}_{ql}) = \frac{1}{2} \sum_{k=1}^{K} \sum_{i=1}^{n} \sum_{j \neq i}^{n} \gamma_{ijql} (x_{ijk} - \pi_{qlk}) \boldsymbol{d}_k, \tag{4.40}$$

and similarly, the negative second order partial derivatives with respect to $\boldsymbol{\beta}_{ql}$ yield the observed Fisher Information matrix $\boldsymbol{\mathcal{I}}(\boldsymbol{\beta})$ which appears as a $Q(Q+1)/2P \times Q(Q+1)/2P$ block diagonal matrix of individual sub-matrices $\boldsymbol{\mathcal{I}}_{ql}(\boldsymbol{\beta}_{ql})$ defined as

$$\boldsymbol{\mathcal{I}}_{ql}(\boldsymbol{\beta}_{ql}) = \frac{1}{2} \sum_{k=1}^{K} \sum_{i=1}^{n} \sum_{j \neq i}^{n} \gamma_{ijql} \pi_{qlk} (1 - \pi_{qlk}) \boldsymbol{d}_k \boldsymbol{d}_k^\top. \tag{4.41}$$

To find the estimates of $\boldsymbol{\beta}_{ql}$, we want to optimise the modified bound given in Eq. (4.39), whose (r)-th element is given as

$$\mathbf{U}_{ql}^{*(r)}(\boldsymbol{\beta}_{ql}) = \mathbf{U}_{ql}^{(r)}(\boldsymbol{\beta}_{ql}) + \frac{1}{2} \text{Tr}[\boldsymbol{\mathcal{I}}_{ql}^{-1}(\boldsymbol{\beta}_{ql}) \frac{\partial}{\partial \boldsymbol{\beta}_{ql}^{(r)}} \boldsymbol{\mathcal{I}}_{ql}(\boldsymbol{\beta}_{ql})], \tag{4.42}$$

and the (r)-th partial derivative of the of the observed Fisher Information matrix is given as

$$\frac{\partial \boldsymbol{\mathcal{I}}_{ql}(\boldsymbol{\beta}_{ql})}{\partial \boldsymbol{\beta}_{ql}^{(r)}} = \frac{1}{2} \sum_{k=1}^{K} \sum_{i=1}^{n} \sum_{j \neq i}^{n} \gamma_{ijql} \pi_{qlk} (1 - \pi_{qlk}) (1 - 2\pi_{qlk}) \boldsymbol{d}_k \boldsymbol{d}_k^\top \boldsymbol{d}_k^{(r)}. \tag{4.43}$$

Lastly, applying the Newton-Raphson formula, the Firth type estimates are found iteratively using the formula

$$\boldsymbol{\beta}_{ql}^{(m)} = \boldsymbol{\beta}_{ql}^{(m-1)} + \boldsymbol{\mathcal{I}}_{ql}^{-1}(\boldsymbol{\beta}_{ql}^{(m-1)}) \mathbf{U}_{ql}^{*}(\boldsymbol{\beta}_{ql}^{(m-1)}). \tag{4.44}$$

Similarly to the Hom-SBM (see Eq. 4.29), the part of the ICL criterion which depends on $\log f(\boldsymbol{z}|\mathrm{m}_Q)$ is the same as in Section 2.4.1. For $\log f(\boldsymbol{x}|\boldsymbol{z}, \mathrm{m}_Q)$, we use the BIC, where we note that the total number of parameters in $\boldsymbol{\beta}$ is $\frac{Q(Q+1)}{2}P$ and the total number of data points in $\boldsymbol{x}$ is $\frac{n(n-1)}{2}K$. Finally, the ICL criterion of the Het-SBM is

$$\begin{aligned}\mathrm{ICL}(\mathrm{m}_Q) =& \log f(\boldsymbol{x}, \hat{\boldsymbol{z}}|\mathrm{m}_Q, \hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}) - \frac{1}{2}\left(\frac{Q(Q+1)}{2}P\right)\log\left[\frac{n(n-1)}{2}K\right] \\ &- \frac{Q-1}{2}\log[n]. \end{aligned} \tag{4.45}$$

Estimation Algorithm

In this model, the proposed estimation algorithm follows the same structure as the algorithm described in Section 4.1.2. Starting with some initial values for $\boldsymbol{\tau}^0$, we iteratively update the model parameters according to the two steps

1. $(\boldsymbol{\alpha}^{(h+1)}, \boldsymbol{\beta}^{(h+1)}) = \underset{(\boldsymbol{\alpha},\boldsymbol{\beta})}{\operatorname{argmax}} \left[\mathcal{J}(f^*(\boldsymbol{z}; \boldsymbol{\tau}^{(h)}); \boldsymbol{\alpha}, \boldsymbol{\beta}) + \frac{1}{2}\sum_{q\leq l}^{Q}\log[\mathrm{Det}(\boldsymbol{\mathcal{I}}_{ql}(\boldsymbol{\beta}_{ql})]\right],$
2. $\boldsymbol{\tau}^{(h+1)} = \underset{\boldsymbol{\tau}}{\operatorname{argmax}} \left[\mathcal{J}(f^*(\boldsymbol{z}; \boldsymbol{\tau}); \boldsymbol{\alpha}^{(h+1)}, \boldsymbol{\beta}^{(h+1)})\right],$

until the convergence is obtained. In the first step, we use Eq. (4.38) to update $\boldsymbol{\alpha}$ and the Newton-Raphson algorithm to update $\boldsymbol{\beta}$ (see Eq. (4.44)). In particular, the Newton-Raphson algorithm is initialised with zero starting values for the parameters ($\boldsymbol{\beta}$) and, to keep the convergence in check, we use the step halving procedure as described in Section 4.1.2. In the second step, we update $\boldsymbol{\tau}$ according to Eq. (4.37).

4.2 Inference in Multi-subject Stochastic Blockmodels

As mentioned earlier in Section 4.1, the multi-subject models offer the possibility to estimate a common cluster structure across subjects that can serve as a common ground for making comparisons between the subjects. In that regard, the Homogeneous and Heterogeneous SBMs pose a logistic regression model on each element of a block structure and thus the methodological framework of logistic regression model can be used to estimate differences between groups of subjects or effects of covariates on the connectivity rate at each block. While these can be interpreted as quantities which can summarise the networks like other summary metrics commonly used in the literature, the logistic regression model offers also the possibility to statistically test if these quantities are different from 0 or, in a more general sense, if linear combinations of these quantities are different from a specified constant. As far as we are aware of, these two models are the firsts to combine clustering and hypothesis testing within the same network model.

In the continuation of this section, we discuss several inference strategies in the context of parametric (i.e. Wald and likelihood ratio test) and non-parametric (i.e. permutation test) approaches. It is important to highlight that each of these tests is assumed to be conditional on $\boldsymbol{\tau}$ and $\boldsymbol{\alpha}$ (i.e. the block assignments). While this may seem like a significant limitation, it is important to note that unconditional inferences would be impractical as the very membership of the blocks (i.e. that each regression coefficient refers to) would be uncertain. While the uncertainty of the unconditional approach would be somewhat of a problem for the Hom-SBM, it would be a catastrophe for the Het-SBM.

4.2.1 Wald Test

In the general context of logistic regression analysis, the Wald test has been commonly used to make inferences on the estimates of the regression coefficients. For the Hom-SBM, we write the set of parameters $\{\boldsymbol{\theta}, \boldsymbol{\beta}\} = \boldsymbol{\psi}$ as a column vector. To perform inference on a combination of the parameters, $\mathcal{H}_0 : \boldsymbol{L}\boldsymbol{\psi} = \boldsymbol{b}_0$, where $\boldsymbol{L}$ is a matrix (or a vector) which defines the combination of the parameters (i.e. linear contrast) tested and $\boldsymbol{b}_0$ are some constants under the null hypothesis, the Wald statistic is

$$W = (\boldsymbol{L}\hat{\boldsymbol{\psi}} - \boldsymbol{b}_0)^\top (\boldsymbol{L}\boldsymbol{\mathcal{I}}^{-1}(\hat{\boldsymbol{\psi}})\boldsymbol{L}^\top)^{-1}(\boldsymbol{L}\hat{\boldsymbol{\psi}} - \boldsymbol{b}_0)/c, \tag{4.46}$$

where c is the rank of $\boldsymbol{L}$. Asymptotically, W follows a χ^2_c distribution. Alternatively, if $\boldsymbol{L}$ is a vector, then the Wald statistic can be noted as

$$W^* = \frac{\boldsymbol{L}\hat{\boldsymbol{\psi}} - \boldsymbol{b}_0}{\sqrt{\boldsymbol{L}\boldsymbol{\mathcal{I}}^{-1}(\hat{\boldsymbol{\psi}})\boldsymbol{L}^\top}}, \tag{4.47}$$

which asymptotically follows a Standard Normal distribution. The standard errors of the model parameters depend on the observed Fisher Information matrix given by Eq. (4.19). Note that Eq. (4.19) applies for the Ordinary and Firth MLEs (Heinze and Schemper, 2002; Firth, 1993).

Similarly, in the context of the Het-SBM, due to the special block diagonal structure of the Fisher Information matrix, we can write the null hypothesis as $\mathcal{H}_0 : \boldsymbol{L}_{ql}\boldsymbol{\beta}_{ql} = \boldsymbol{b}_{ql0}$. The Wald statistic can take the form

$$W_{ql} = (\boldsymbol{L}_{ql}\hat{\boldsymbol{\beta}}_{ql} - \boldsymbol{b}_{ql0})^\top (\boldsymbol{L}_{ql}\, \boldsymbol{\mathcal{I}}_{ql}^{-1}(\hat{\boldsymbol{\beta}}_{ql})\boldsymbol{L}_{ql}^\top)^{-1}(\boldsymbol{L}_{ql}\hat{\boldsymbol{\beta}}_{ql} - \boldsymbol{b}_{ql0})/c_{ql} \tag{4.48}$$

where $\boldsymbol{L}_{ql}$ is a matrix (or a vector) defining the combination of the parameters (contrast) tested and c_{ql} is the rank of $\boldsymbol{L}_{ql}$. Asymptotically, W_{ql} follows a χ^2_c distribution.

If $\boldsymbol{L}_{ql}$ is a vector, then

$$W^*_{ql} = \frac{\boldsymbol{L}_{ql}\hat{\boldsymbol{\beta}}_{ql} - \boldsymbol{b}_{ql0}}{\sqrt{\boldsymbol{L}_{ql}\boldsymbol{\mathcal{I}}^{-1}_{ql}(\hat{\boldsymbol{\beta}}_{ql})\boldsymbol{L}^{\top}_{ql}}}, \tag{4.49}$$

asymptotically follows a Standard Normal distribution. As in the case of Hom-SBM, the standard errors of the model parameters are estimated with the Fisher Information matrix (see Eq. (4.41)) and, as noted previously, this equation applies for both the Ordinary and Firth MLE approaches.

Multiple testing Inference procedures for the block level parameters comprise a multiple testing problem as we are effectively making $Q(Q+1)/2$ individual tests. To control the family wise error rate (FWE), defined as the probability of making at least one Type I error, we use the Bonferroni correction (Holm, 1979). This correction is valid for any dependence structure and is easy to apply: Instead of using a nominal α_0 significance value (e.g., 0.05), α_0/n_T is used instead, where n_T is the number of tests (here, $n_T = Q(Q+1)/2$).

4.2.2 Likelihood Ratio Test

In the case of the Hom-SBM, to perform inference on a combination of parameters $\mathcal{H}_0 : \boldsymbol{L}\boldsymbol{\psi} = \boldsymbol{b}_0$, we can use the likelihood ratio (LR) test. For the Firth approach, we base the likelihood ratio test statistic on the modified variational bound given by Eq. (4.23), and thus the test statistic can be stated as

$$\Lambda = 2\Big[\mathcal{J}^*(f^*(\boldsymbol{z};\hat{\boldsymbol{\tau}});\hat{\boldsymbol{\alpha}},\hat{\boldsymbol{\psi}}) - \mathcal{J}^*(f^*(\boldsymbol{z};\hat{\boldsymbol{\tau}});\hat{\boldsymbol{\alpha}},\tilde{\boldsymbol{\psi}})\Big], \tag{4.50}$$

where $\tilde{\boldsymbol{\psi}}$ is a column vector containing the parameter estimates under the null hypothesis. In particular, Λ is assumed to follow a χ^2_c distribution where c is the rank of $\boldsymbol{L}$. Note that the $\tilde{\boldsymbol{\psi}}$ is derived from a restricted model whose penalisation term is different from the penalisation term associated with the full model. Hence, to derive Λ, we do not use the bound associated with the restricted model, but we use instead the bound of the full model evaluated with $\tilde{\boldsymbol{\psi}}$. Note that for the Ordinary MLEs, the likelihood ratio is computed in a similar manner, but it is based on the non-modified variational bound given by Eq. (4.12).

Similarly, in the case of the Het-SBM, to perform inference on a combination of parameters $\mathcal{H}_0 : \boldsymbol{L}_{ql}\boldsymbol{\beta}_{ql} = \boldsymbol{b}_{ql0}$, we can use the likelihood ratio test. For the Firth MLEs, using the modified variational bound given by Eq. (4.39), we can formulate

the likelihood ratio statistic as

$$\Lambda = 2\Big[\mathcal{J}^*(f^*(\boldsymbol{z};\hat{\boldsymbol{\tau}});\hat{\boldsymbol{\alpha}},\hat{\boldsymbol{\beta}}) - \mathcal{J}^*(f^*(\boldsymbol{z};\hat{\boldsymbol{\tau}});\hat{\boldsymbol{\alpha}},\tilde{\boldsymbol{\beta}})\Big], \tag{4.51}$$

where $\tilde{\boldsymbol{\beta}}$ is the set of all parameters containing the parameter estimates under the null hypothesis. As before, Λ is assumed to follow a χ^2_c distribution where c is the rank of $\boldsymbol{L}_{ql}$. In particular, $\tilde{\boldsymbol{\beta}}$ is derived from a restricted model whose penalisation term is different from the penalisation term associated with the full model. Hence, to derive Λ, we do not use the bound associated with the restricted model, but we use instead the bound of the full model evaluated with $\tilde{\boldsymbol{\beta}}$. Note that for the Ordinary MLEs, the likelihood ratio is computed in a similar manner, but it is based on the non-modified variational bound given by Eq. (4.34).

Note that the multiple testing procedures outlined in the previous section (Section 4.2.1) also apply for the likelihood ratio test.

4.2.3 Permutation Test

As both the Wald and likelihood ratio tests depend on asymptotic sampling distributions, we also consider a non-parametric testing procedure. Permutation tests are based on the premise that, under the null hypothesis, the data can be exchanged without altering its distribution (Good, 2000). As a result of this, the distribution of any test statistics can be found empirically through sequential evaluations of the rearranged (or permuted) data. In the context of the Het-SBM model, we use permutation tests to make inferences on the $\boldsymbol{\beta}_{ql}$, where we use a null hypothesis of no association between edge occurrence and the covariate tested. We only consider tests of the entire parameter vector $\boldsymbol{\beta}_{ql}$, although permutation tests allowing for nuisance variables are possible (for a review, see Winkler et al., 2014).

The p-value of the observed Wald test statistic w_{ql0} is computed by a Monte Carlo sampling scheme, so that, for a sequence of random permutations indexed by t ($t = 1, \ldots, M$), we obtain M Wald statistics labelled as $w_{ql1} \ldots, w_{qlM}$. Thus, including the observed Wald statistic w_{ql0} into the permuted scores, the Monte Carlo p-value is computed as

$$\mathrm{P}(W_{ql} \geq w_{ql0}|H_0) = \frac{\sum_{t=0}^{M} I(w_{qlt} \geq w_{ql0})}{M+1}, \tag{4.52}$$

where $I(\cdot)$ is the indicator function. The same approach can be used with the likelihood ratio test to obtain p-values without assuming an asymptotic χ^2 distribution.

Improved Multiple testing procedures with permutation The Bonferroni method for controlling the FWE is conservative in the presence of dependence.

Hence, we also consider a permutation based procedure. Specifically, the FWE of a critical value (e.g., W_C) can be expressed as the probability that the maximum statistic (taken over the $Q(Q+1)/2$ blocks) exceeds that critical value under the null hypothesis (Westfall and Young, 1993). Therefore, to control the FWE at a level of significance 5% for the Wald test, we simply need to find the value of W_C such that

$$\mathrm{P}(\max\left[((W_{ql}))_{1\leq q,l\leq Q}\right] \geq W_C|H_0) = 0.05, \tag{4.53}$$

likewise for the likelihood ratio (LR) test. To find W_C, we need to record the maximum statistic across the $Q(Q+1)/2$ blocks for the original and each of the M permutation statistics, and choose W_C as the 95-th percentile of all the maximum statistics. The corresponding FWE corrected p-values, for each of the $Q(Q+1)/2$ blocks, is likewise computed as the proportion of other group maximal statistics which are greater than or equal to it (Westfall and Young, 1993; Nichols and Hayasaka, 2003).

4.3 Simulation Methods

In this section, we use two Monte Carlo simulations, Simulation I and II, to investigate the properties of the proposed multi-subject SBMs. In Simulation I, we address three main points. First, we investigate how closely each multi-subject model recovers the true cluster structure. Second, we investigate the accuracy of the model's parameters under a range of different synthetic data, and we contrast the Ordinary and Firth ML estimates. Third, we investigate the false positive rates (FPRs) of the inference procedures associated with each model. In Simulation II, we focus on specific parameter configurations that are particularly challenging for ML estimation and that may be improved by the Firth approach. We also study the effects of small samples in more detail by varying the block sizes and the number of nodes in networks.

In both sets of simulations, we fix the total number of blocks to be three ($Q = 3$) and order the sizes of each block so that $n_1 \geq n_2 \geq n_3$. To study the effects of block sizes on the accuracy of the parameter estimates, we consider three proportion designs (Balanced, Mildly Unbalanced and Unbalanced) whose details are shown in Table 4.1. The Balanced case maintains an even spread of nodes, while the Mildly Unbalanced and Unbalanced cases tend to shrink the number of nodes in the last two blocks. As the number of edges (or data points) depends on the individual block sizes, we can expect a more accurate fit in the case of the Balanced design than in the cases of the Mildly Unbalanced and Unbalanced designs. However, since the block size increases with increasing network sizes, it is expected

Network Sizes	Proportion Design	Block Sizes n_1	n_2	n_3
$n = 30$	Balanced	10	10	10
	Mildly Unbalanced	18	19	3
	Unbalanced	21	6	3
$n = 50$	Balanced	17	17	16
	Mildly Unbalanced	30	15	5
	Unbalanced	35	10	5
$n = 100$	Balanced	34	34	32
	Mildly Unbalanced	60	30	10
	Unbalanced	70	20	10
$n = 500$	Balanced	167	167	166
	Mildly Unbalanced	300	150	50
	Unbalanced	350	100	50

Table 4.1: **Proportion Designs**. Number of nodes for each scenario. The Balanced, Mildly Unbalanced and Unbalanced proportion designs are defined as the ratio of individual block sizes and the total number of nodes in a network.

that this effect is ameliorated in the larger networks. For each simulation setting, we generate the true partition as a vector of length n. Comparisons between the estimated clustering and the true partition are carried out using the Adjusted Rand Index, discussed in Section 2.5 (Eq. (2.40)).

For each combination of parameters in Simulations I and II, we generate $S = 1000$ networks, $s = 1, \ldots, S$ and, to this data, we fit the relevant multi-subject model (e.g., in Simulation I, we consider the Bin-SBM, Hom-SBM and Het-SBM while in Simulation II, we consider the Het-SBM). For each simulated network and multi-subject model, we only consider a single restart with $Q = 3$. We compare their resulting parameter estimates using the root mean square error (RMSE) and the bias (Bias), defined as

$$\text{RMSE}(\hat{\alpha}_q) = \left(\frac{1}{S} \sum_{s=1}^{S} (\hat{\alpha}_{qs} - \alpha_q)^2 \right)^{1/2}, \tag{4.54}$$

$$\text{RMSE}(\hat{\theta}_{ql}) = \left(\frac{1}{S} \sum_{s=1}^{S} (\hat{\theta}_{qls} - \theta_{ql})^2 \right)^{1/2}, \tag{4.55}$$

$$\text{RMSE}(\hat{\boldsymbol{\beta}}_{ql}^{(p)}) = \left(\frac{1}{S} \sum_{s=1}^{S} (\hat{\boldsymbol{\beta}}_{qls}^{(p)} - \boldsymbol{\beta}_{ql}^{(p)})^2 \right)^{1/2}, \tag{4.56}$$

$$\text{Bias}(\hat{\boldsymbol{\beta}}_{ql}^{(p)}) = \frac{1}{S} \sum_{s=1}^{S} (\hat{\boldsymbol{\beta}}_{qls}^{(p)} - \boldsymbol{\beta}_{ql}^{(p)}), \tag{4.57}$$

$$\text{Bias}(\widehat{\text{var}}(\hat{\boldsymbol{\beta}}_{ql}^{(p)})) = \frac{1}{S}\sum_{s=1}^{S}(\widehat{\text{var}}(\hat{\boldsymbol{\beta}}_{qls}^{(p)}) - \text{var}(\hat{\boldsymbol{\beta}}_{ql}^{(p)})), \tag{4.58}$$

where $\hat{\boldsymbol{\beta}}_{ql}^{(p)}$ is the (p)-th element of $\hat{\boldsymbol{\beta}}_{ql}$, $\hat{\boldsymbol{\beta}}_{qls}^{(p)}$ is its realisation in simulation s, $\widehat{\text{var}}(\hat{\boldsymbol{\beta}}_{ql}^{(p)})$ corresponds to the (p,p)-th element of the inverse of the observed Fisher Information given in Eq. (4.41), $\widehat{\text{var}}(\hat{\boldsymbol{\beta}}_{qls}^{(p)})$ is its realisation in simulation s and $\text{var}(\hat{\boldsymbol{\beta}}_{ql}^{(p)})$ is the simulation sample variance of $\hat{\boldsymbol{\beta}}_{ql}^{(p)}$. Note that we can use the transformation $\log(\hat{\boldsymbol{\pi}}/(1-\hat{\boldsymbol{\pi}})) = \hat{\boldsymbol{\theta}}$ for the Bin-SBM so that Eq. (4.55) applies.

4.3.1 Simulation I

In this simulation, we consider ten subjects ($K = 10$) and networks of three sizes ($n \in \{50, 100, 500\}$ nodes). Binary network edges are generated according to eight different types of connectivity structures, labelled PI1-8. As shown in Figure 4.2, each such PI portrays various rates of within/between block connectivity and, therefore, different cluster structures in the simulated networks. Furthermore, each PI represents different levels of clustering evidence for three blocks. For example, PI1 illustrates the case when there is no block structure at all because all three blocks rates are the same. Similarly, the combination of PI1 and Unbalanced proportions will also be difficult to fit and the same can be said about PI2 and PI5, which have almost homogenous rates of connections. For the remaining cases, the fitting is expected to be easier as there is clearly more heterogeneity in the data. Note that, apart from the modular structure in PI6, all cases have identical within-block connection rates.

Noting that the labelling of node assignment is random for each fit, we need to correct this by relabelling the numbering of the Block after each fit. For the majority of the scenarios, we resolved this by rearranging the blocks by decreasing sizes, giving the label 'Block 1' to the largest block, the label 'Block 3' for the smallest block and the label 'Block 2' for the remaining block. However, this approach cannot work in the Balanced design as the blocks have the same size (see Table 4.1) and, thus, there is some ambiguity in their ordering. To resolve this, we take advantage of the PI structure, where in all cases (see Figure 4.2), except for PI1 & PI6, the off-diagonal elements are decreasing. For PI6, we assign the labels in such a way that 'Block 1' has the smallest connectivity rate while 'Block 3' has the largest. For PI2-PI5, PI7 & PI8, we rearrange the labels in such a way that the connectivity rates between 'Block 1' and 'Block 2' is the largest between-block connectivity rate and the connectivity rates between 'Block 2' and 'Block 3' is the smallest between-block connectivity rate. However, note that the labelling for PI1 and the Balanced design is not identifiable at all.

To simulate the networks, we also consider the effect of an age covariate.

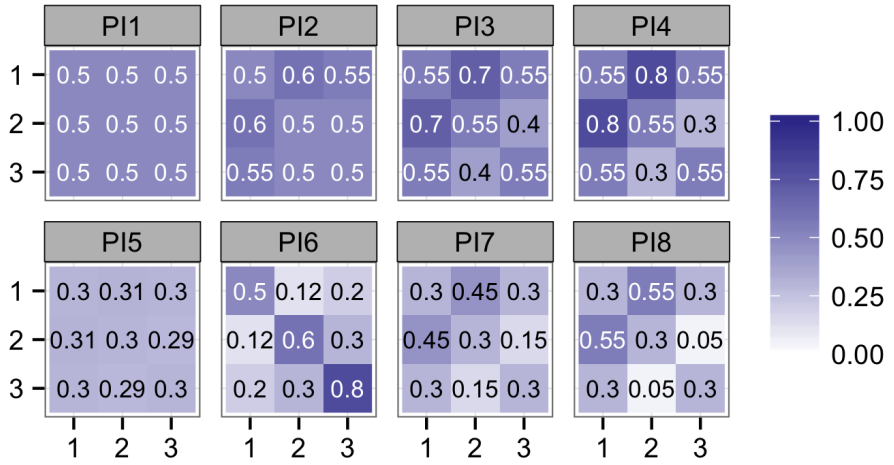


Figure 4.2: **Design of connectivity structures PI1-8**. Each of eight different cases portraits various levels of homogeneity in the cluster structures, starting from PI1, P2 and PI5 as highly homogenous examples to more heterogeneous cases with stronger evidence for three block structures (PI3-4 and PI6-8).

Each of the ten subjects is assigned an age score, between 21 and 49 years, such that their average value is 33.9 (the same as in the real data). We consider two cases: (a) no age effect and hence all subjects have the same PI (see Figure 4.2), and (b) the effect of age is set to -0.025, so that the subject specific block-rates decrease with age. To distinguish between these two cases, hereafter, we use the notation $n50\_0$ and $n50\_0025$ to indicate the size of network (50 nodes) and to state an age effect (i.e. n50_0 stands for no age effect and n50_0025 stands for an age effect of -0.025).

For the Het-SBM, we apply the Wald and permutation inference procedures described in Sections 4.2.1 and 4.2.3. The p-values of the Monte Carlo permutation test statistic were obtained by computing 1000 permutations of the age covariate in each of the simulations. We compare both approaches with respect to their control of the false positive rates (FPR) and their power. In addition, we also consider the control of FWE rate for both approaches when they were using a FWE correction.

4.3.2 Simulation II

In this simulation, we consider samples with 3, 5 and 10 subjects ($K \in \{3, 5, 10\}$) and four network sizes ($n \in \{30, 50, 100, 500\}$ nodes) and the block sizes varying according to Table 4.1. By considering additional networks with 30 nodes, we have

generated blocks which are smaller than those considered in the previous set of simulations. In the combination with other simulation parameters, the blocks with fewer nodes are expected to create small sample effects for which we expect to see differences between the Ordinary and Firth MLEs. Similarly to the previous set of simulations, we generate data considering either the presence or the absence of an age effect, whose respective sizes are set to -0.025 and 0. The binary networks

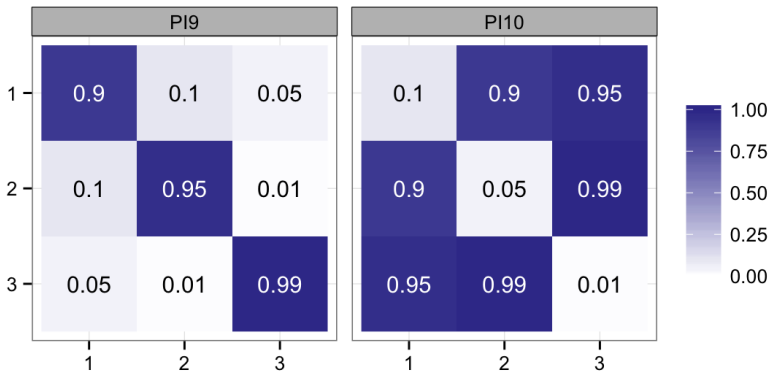


Figure 4.3: **Design of connectivity structures PI9 and PI10**. Connection rates portrait two cases of cluster structure, which are relatively easy to estimate with the multi-subject SBMs. However, the probability rates should pose some challenges in the logistic regression estimation, and this should provide a more interesting basis for the comparisons between the Ordinary and Firth ML estimates.

are generated according to the connection rates given in Figure 4.3. The cluster structures featured in PI9 and PI10 are expected to be relatively easy to estimate by the Het-SBM, so that, regardless of the MLE approach, we will get the same cluster structure estimate, allowing a clear comparison between the two approaches. It is also interesting to note that PI9 and PI10 are characterised by saturated/sparse connection rates. This, in the combination of a small number of nodes and subjects, is expected to pose some interesting estimation challenges.

4.4 Simulation Results

In this section, we detail the results of Simulation I and II. In Simulation I, we show the behaviour of the parameter estimates in the circumstances when the Ordinary and Firth MLEs coincide and, in Simulation II, we focus in more details on the improvements brought in by the Firth MLEs.

4.4.1 Results of Simulation I

We report the quality of the fitted parameter estimates based on the RMSE of $\hat{\boldsymbol{\alpha}}$ and the regression parameters $(\hat{\boldsymbol{\theta}}_{ql}, \hat{\boldsymbol{\beta}}_{ql})$ associated with the Hom-SBM and Het-SBM respectively.

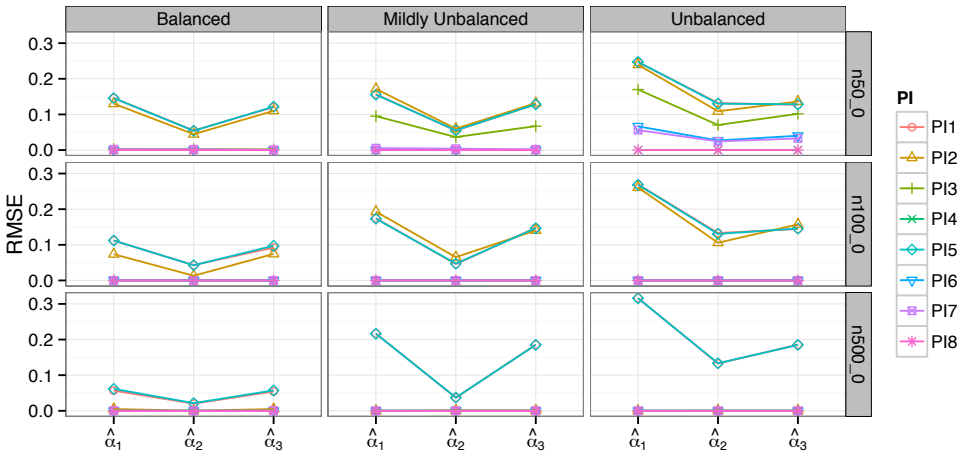


Figure 4.4: **RMSE of $\hat{\boldsymbol{\alpha}}$**. The elements of the vector $\hat{\boldsymbol{\alpha}}$ are given along the x-axis. The RMSE scores are given with respect to varying (i) proportion designs (Balanced, Mildly Unbalanced and Unbalanced, on each column), (ii) network sizes ($n \in \{50, 100, 500\}$ nodes, on each row), and (iii) connectivity structures (PI1-8, plotting symbols and colours).

In Figure 4.4, the highest RMSE values are mostly associated with the challenging types of connectivity structures (PI1-2 & PI5), while the smallest values are mostly associated with PI3-4 and PI6-8 whose connectivity structures show more clustering evidence. The proportion designs (Unbalanced and Mildly Unbalanced) and the network size also contribute to the accuracy of the estimates. Indeed, the accuracy seems to increase with increasing network or block sizes.

In Figure 4.5, we also show the RMSE for $\hat{\boldsymbol{\theta}}$ with respect to the Ordinary and Firth MLEs. In particular, for the connectivity data considered in this simulation (PI1-8), there are no appreciable differences between the two approaches. Overall, we can observe that increasing the network sizes, and hence, block sizes, tends to improve the RMSE values. It is also evident that the proportion design influences the accuracy of the estimates. For example, in the Unbalanced design, the estimates related to Block 3 (i.e. $\hat{\theta}_{23}$ and $\hat{\theta}_{33}$) are less accurate which can be explained by the very small number of nodes in Block 3 (see Table 4.1). In all of these settings, there seems to be little difference between the RMSE of the Firth and Ordinary MLEs even though our simulation consider small block sizes and 10 subjects. These results might suggest that there is no interest in using the Firth MLEs even in small

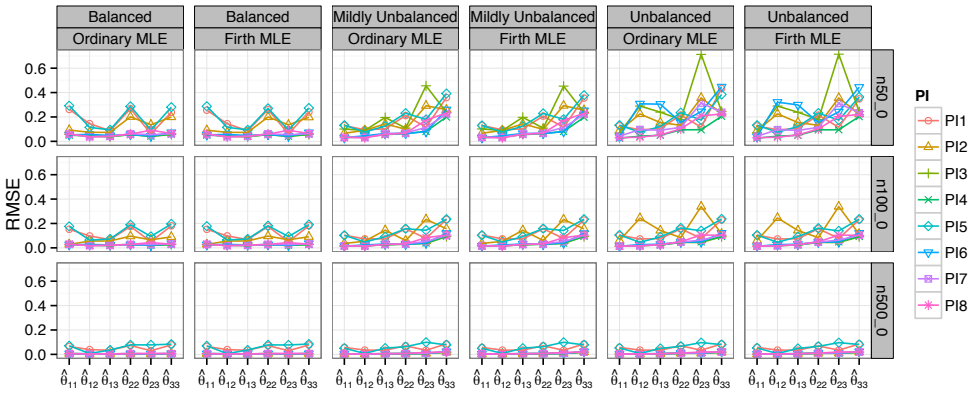


Figure 4.5: **RMSE of $\hat{\theta}$ evaluated in terms of the Ordinary and Firth MLEs**. The RMSE scores are given with respect to varying (i) proportion designs (Balanced, Mildly Unbalanced and Unbalanced, on each column), (ii) network sizes ($n \in \{50, 100, 500\}$ nodes, one each row), and (iii) connectivity structures (PI1-8, plotting symbols and colours). For these PI values, there are no appreciable differences between the Ordinary and Firth MLEs.

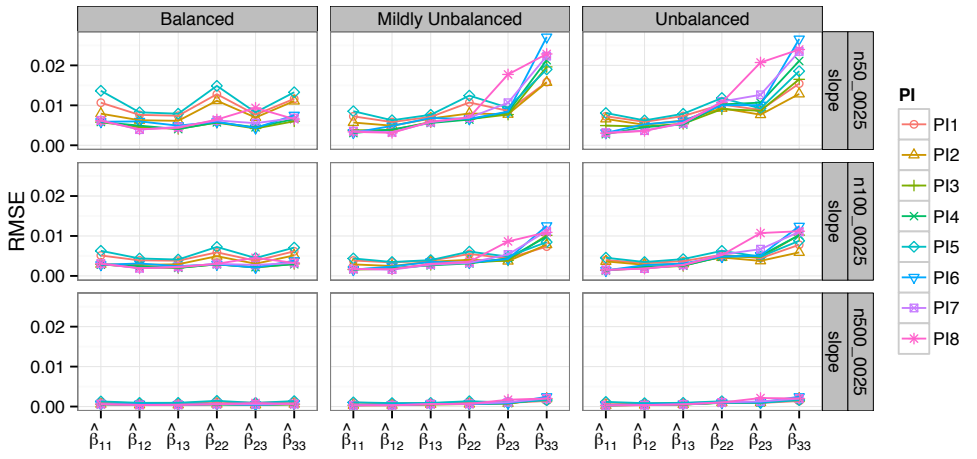


Figure 4.6: **RMSE of $\hat{\beta}_{ql}$ (Het-SBM) with an age effect of -0.025**. The Ordinary and Firth MLE approaches gave identical results, so only one set of the results is shown. The RMSE scores are given with respect to varying (i) proportion designs (Balanced, Mildly Unbalanced and Unbalanced, on each column), (ii) network sizes ($n \in \{50, 100, 500\}$ nodes, one each row), and (iii) connectivity structures (PI1-8, plotting symbols and colours).

samples. However, as we will show in Simulation II, there is a gain in using the Firth MLEs in small samples when the connectivity rates are extreme (i.e. close to 0 or 1). In Figure 4.6, we show only the RMSE for $\hat{\boldsymbol{\beta}}_{ql}$ slope (i.e. the effect of age) of the Het-SBM as the estimates for the intercepts are very similar to the those presented

in Figure 4.5. The estimates associated with the slope are reasonably accurate.

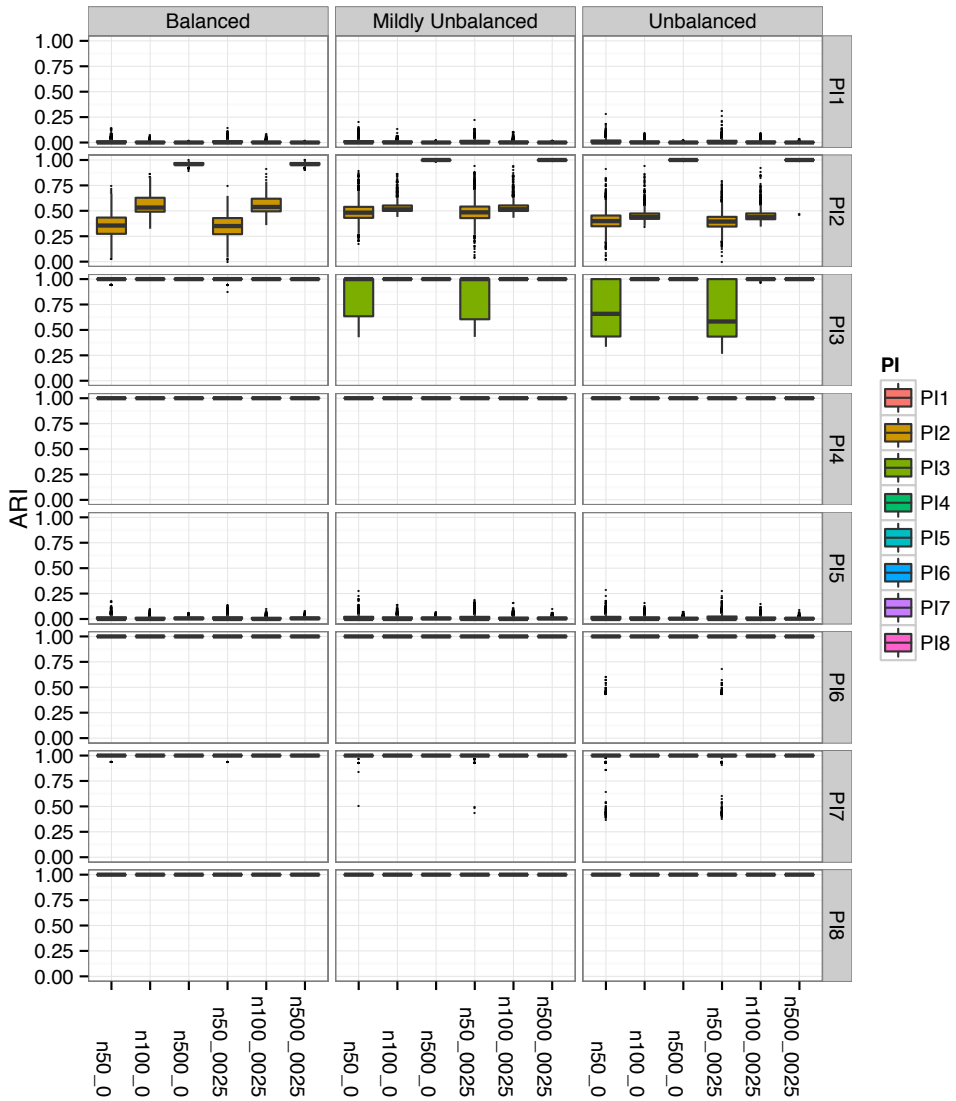
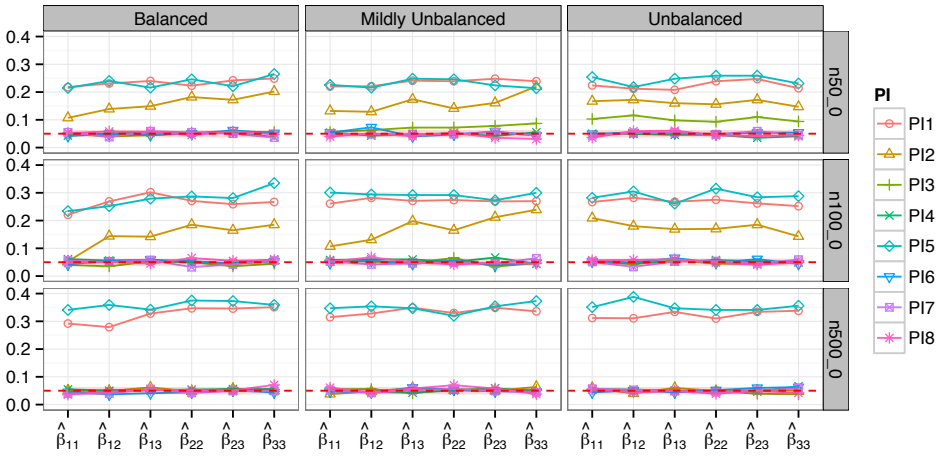
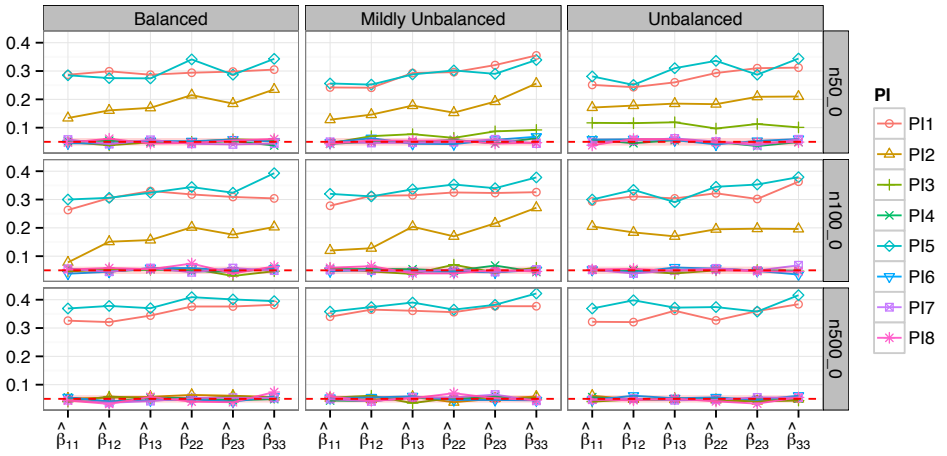


Figure 4.7: **ARI scores for the Het-SBM**. The network sizes and the presence/absence of the effect are given along the x-axis. The Het-SBM is evaluated with respect to varying (i) proportion designs (Balanced, Mildly Unbalanced and Unbalanced, on each column) and (ii) connectivity structures (PI1-8, on each row and plotting colour). These results are consistent with the results of the Hom-SBM and Bin-SBM.

In addition to these results, we also use ARI scores (see Eq. (2.40)) to evaluate how similar the estimated partitions are to the true partitions. In Figure 4.7, we note that, for PI1 and PI5, the estimated partitions never fully match the true partition and the median values are consistently low. For PI2 and PI3, we note that with an



(a) **Wald test**

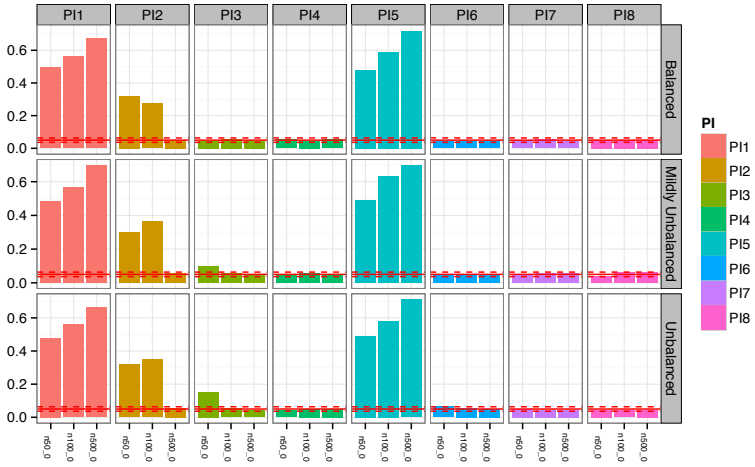


(b) **Monte Carlo Permutation test**

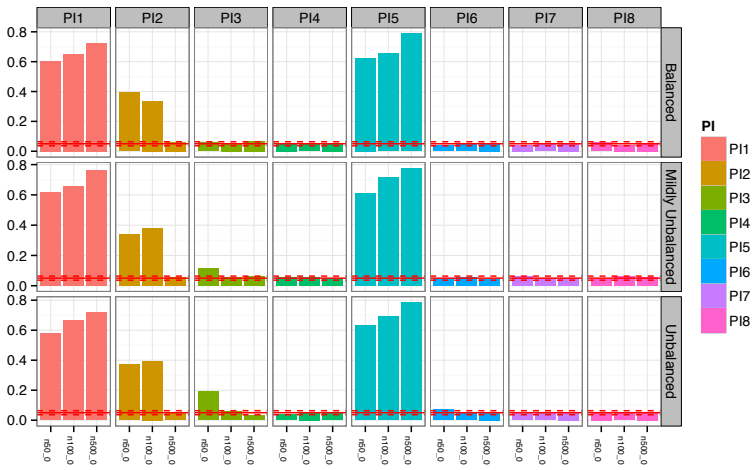
Figure 4.8: **False Positive Rates (FPR) for $\hat{\beta}_{ql}$ slope (age effect)**. Both methods are evaluated with respect to (i) proportion designs (Balanced, Mildly Unbalanced and Unbalanced, on each column), (ii) network sizes (on each row) and (iii) connectivity structures (PI1-8, plotting symbols and colours). Both tests show a similar profile with respect to the simulation conditions and are liberal for the cases in which the model does not accurately recover the true partitions.

increasing network and block sizes, the estimated partitions almost always recover the true partition, suggesting the influence of the proportion design and the network size. The other types of connectivity structure are generally more straightforward to estimate and, thus, in the cases of PI4 and PI6-8, the model consistently recovers the true partitions over all 1000 realisations.

In the final part of our simulation analysis, we investigate the accuracy of



(a) **Wald test/Bonferroni correction**



(b) **Monte Carlo permutation test/Max-based correction**

Figure 4.9: **Control of the FWE after correction**. Network sizes are displayed along the x-axis. Both methods are evaluated with respect to varying (i) proportion designs (Balanced, Mildly Unbalanced and Unbalanced, on each row) and (ii) connectivity structures (PI1-8, on each column). The significance level of 0.05 is given by a red line, while the Monte Carlo confidence interval $(0.04, 0.06)$ is demarcated by red dashed lines. Both tests are accurate for the cases in which the model accurately recovers the true partitions.

parametric and non-parametric inference procedures. In Figure 4.8, we show the FPR for the Wald and Monte Carlo permutation tests, computed at a significance level of 0.05. For each test, we show the confidence interval for a nominal 0.05 rejection rate using a Normal approximation to the Binomial distribution $(0.04, 0.06)$, demarcated by a red shaded strip. It is evident that both tests show remarkable

mutual similarity and also equal sensitivity to the simulation conditions. In that sense, both tests show a good control of the FPR for PI4 and PI6-8. For these cases, the influence of the proportion design and network size seems negligible. However, in the cases of PI1 and PI5, we can see that the inference is always liberal (i.e. FPR>0.05). Nevertheless, for the remaining connectivity structures (PI2 and PI3), we see that both methods fail to control the FPR in small networks but that the quality of their inference is improved when the number of nodes is increased. Thus, the inference becomes more accurate in networks with 100 and 500 nodes. From Figure 4.7, we see that a successful control of the FPR corresponds closely to the cases in which the estimated partitions are the most accurate. For example, for PI4 and PI6-8 the method accurately retrieves the true partitions and this is reflected in the good control of the FPR. Similarly, when the model struggles to estimate the true partition, like in the case of PI2, we see a poor control of the FPR, which becomes instantaneously better when the model recovers the true partition. Thus, we can conclude that the accuracy of the tests is not driven by the logistic regression but by the general framework of the mixture model which struggles to provide good partition estimates in the challenging cases with very little evidence for the three block fit.

In Figure 4.9, we show the control of the FWE rate for the slope estimates after FWE correction (see Sections 4.2.1 and 4.2.3). Notably, both methods show very similar results and the overall improved control of FWE for the well behaved cases. As in the FPR results, we see that both tests struggle with PI1 and PI5 and show improved accuracy for the cases PI2-3, when the estimated partition recovers exactly the true partition.

4.4.2 Results of Simulation II

Due to space considerations, we refrain from reporting on all of the simulation outcomes and, instead, we focus only on the most striking results. Specifically we show only the Unbalanced design since these by construction are expected to exhibit very strong small sample effects (see Table 4.1). Also, we only consider results for the connectivity matrix PI9 (see Figure 4.3). Nevertheless, we note that the full set of results is available in Appendix B.1.

Figures 4.10 (a) and (b) show the bias of the intercept and slope ($\hat{\boldsymbol{\beta}}_{ql}$) estimates according to the Ordinary and Firth MLEs. In several simulation scenarios, the Ordinary and Firth MLEs seem to be equally accurate, however, in the smallest samples (in terms of nodes and subjects), the Ordinary MLE displays the tendency to be severely biased, which is evidently not the case with the Firth MLE which seems to be almost unbiased. This is especially noticeable in the cases with three subjects and small blocks (e.g., in the networks with 30 nodes) for which the bias of

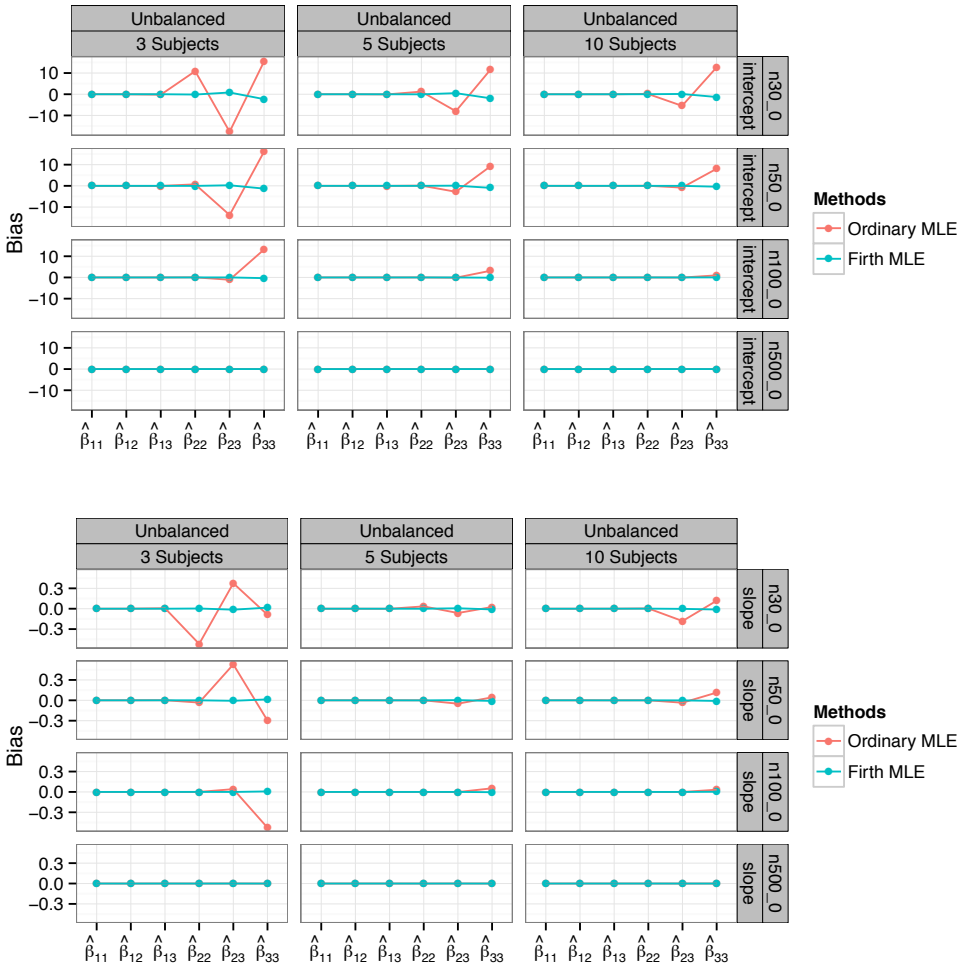
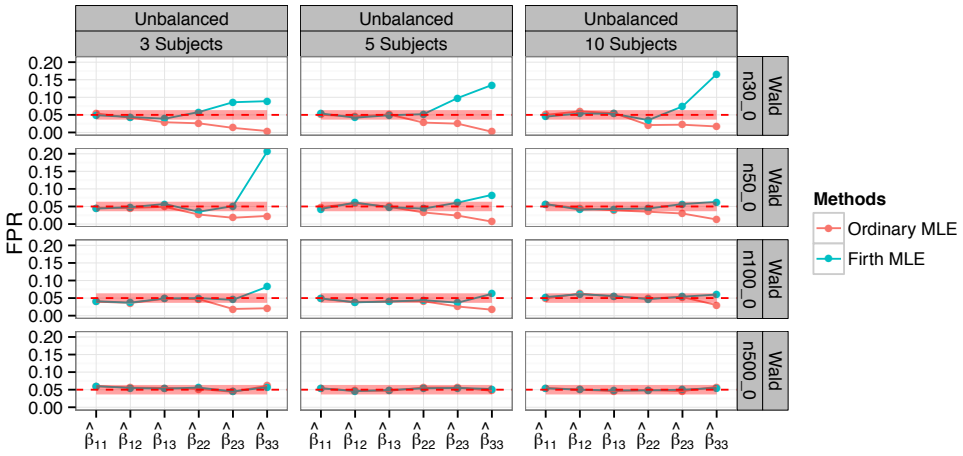


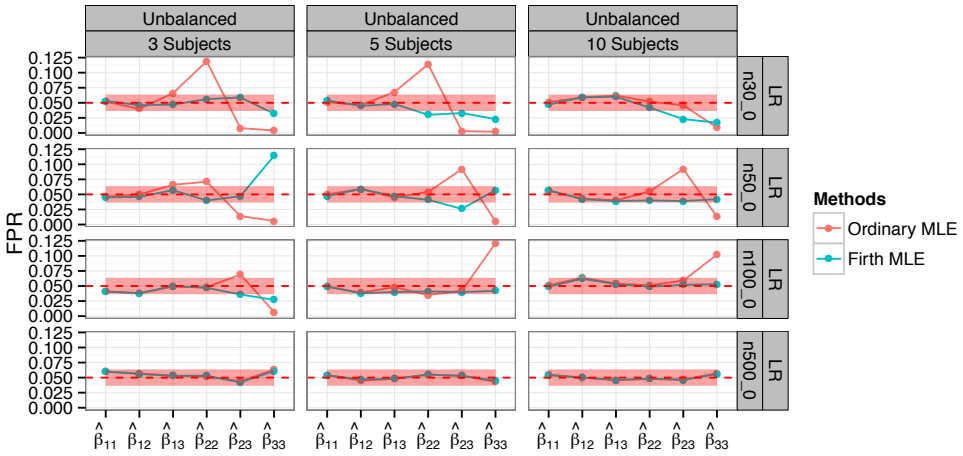
Figure 4.10: **Bias of $\hat{\boldsymbol{\beta}}_{ql}$ of the Het-SBM in terms of the Ordinary and Firth MLEs**. Between/within block $\hat{\boldsymbol{\beta}}_{ql}$ values are displayed along the x-axis. These estimates are shown with respect to the Unbalanced proportion design, varying numbers of subjects ($K \in \{3, 5, 10\}$) and network sizes ($n \in \{30, 50, 100, 500\}$ nodes).

the Ordinary MLE seems to be the largest. However, it is also interesting to point out that these effects are ameliorated in larger blocks and with more subjects.

Next, we report the quality of the inference procedures based on the Wald and likelihood ratio tests, each of which is assessed in terms of the Ordinary and Firth MLE approaches. In Figure 4.11 (a), we show the False Positive Rates (FPR) for the Wald test related to the intercept values. Note that despite the obvious small sample bias of the Ordinary MLEs, it seems that the Wald test based on such estimates are conservative. A possible explanation for this can be obtained in Figure 4.12 which shows the bias of the variance of the intercept ($\widehat{\text{var}}(\hat{\boldsymbol{\beta}}_{ql})$). The



(a) Wald test for $\hat{\boldsymbol{\beta}}_{\boldsymbol{ql}}$ intercepts



(b) Likelihood ratio (LR) test for $\hat{\boldsymbol{\beta}}_{\boldsymbol{ql}}$ intercepts

Figure 4.11: **False Positive Rates (FPR) at 5% significance level for the Wald and likelihood ratio (LR) tests based on the Ordinary and Firth MLEs**. Between/within block intercept values are displayed along the x-axis. The FPR are shown with respect to the Unbalanced proportion design, varying numbers of subjects ($K \in \{3, 5, 10\}$) and network sizes ($n \in \{30, 50, 100, 500\}$ nodes).

variance estimates of the Ordinary MLEs are strongly positively biased and, as a result of this, the Wald scores are deflated and the test is conservative. In contrast to this, the Wald test based on the unbiased Firth estimates (Figure 4.11 (a)) has the tendency to be liberal in the small sample cases. This is especially evident in the cases of blocks with only three nodes (e.g., network with $n = 30$ nodes), for which an increasing number of subjects does not seems to lead to a better control of the FPR. However, in Figure 4.11 (b), we see that the likelihood ratio (LR) test

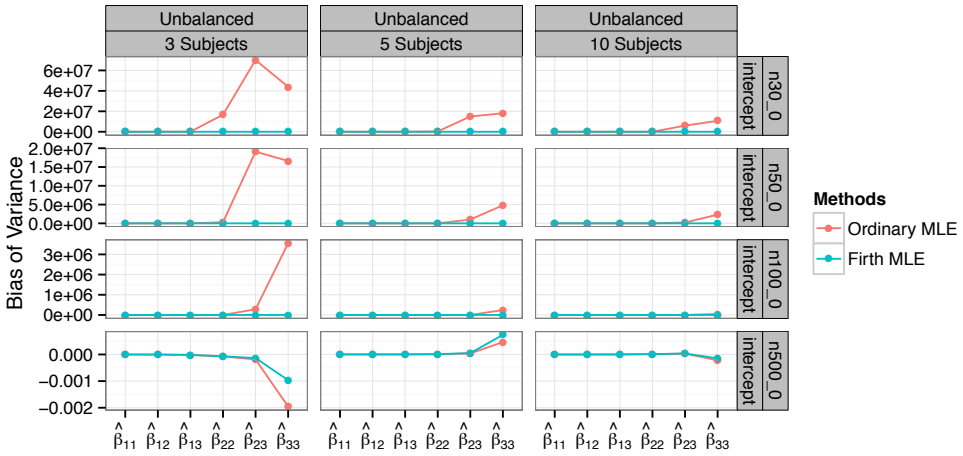
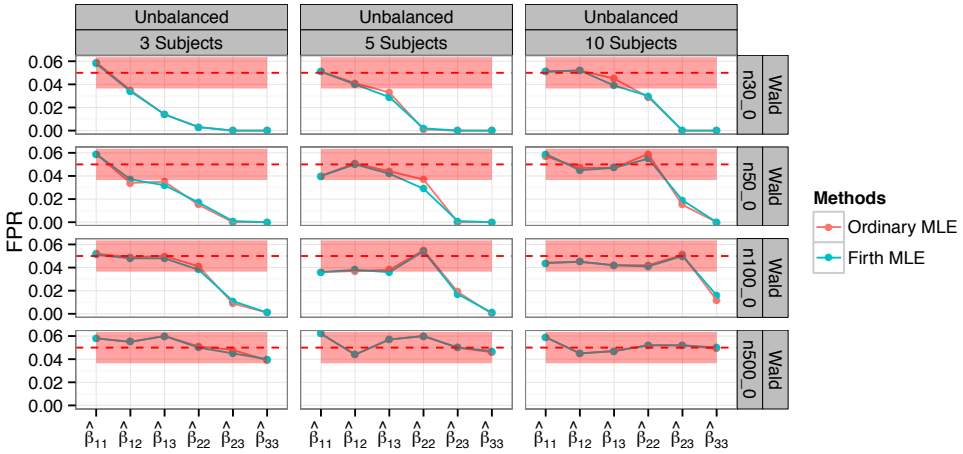


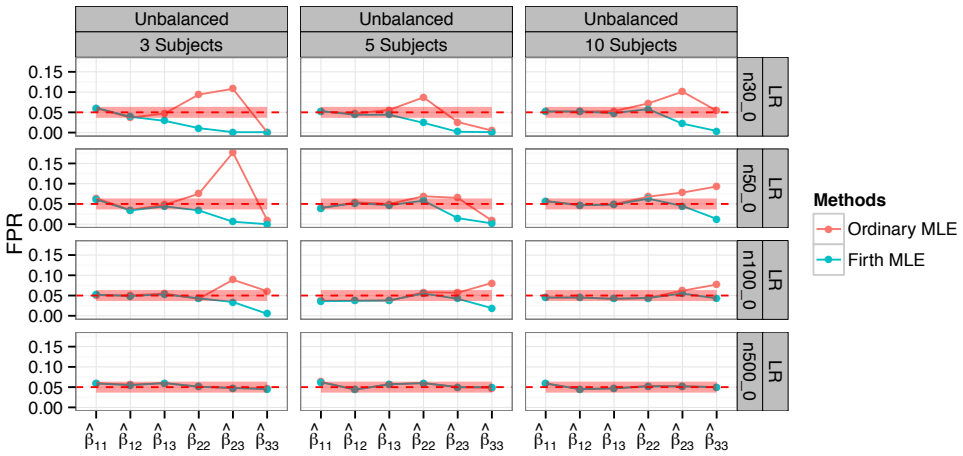
Figure 4.12: **Bias of $\widehat{\text{var}}(\hat{\beta}_{ql})$ intercepts for the Ordinary and Firth MLEs.** Between/within block intercept values are displayed along the x-axis. Bias estimates are evaluated with respect to Unbalanced proportion design, varying numbers of subjects ($K \in \{3, 5, 10\}$) and network sizes ($n \in \{30, 50, 100, 500\}$ nodes). In many scenarios, the Ordinary MLEs present a large bias that is corrected by the the Firth MLEs.

is better behaved for the Firth MLEs than for the Ordinary MLEs. Indeed, the Ordinary MLEs in small samples seem to be either liberal or conservative. On the other hand, the Firth MLEs have the tendency to be slightly conservative in small samples with the exception of the case with three subjects and the network with 50 nodes where it seems to be liberal.

In addition to this, we also investigate the performance of the Wald and likelihood ratio tests based on the slope estimates of the Ordinary and Firth MLE approaches. In Figure 4.13 (a), we note that the Ordinary and Firth MLE approaches are very similar and tend to be accurate in many situations. However, in the most difficult scenarios, both approaches exhibit a conservative behaviour. In contrast, for the likelihood ratio test (see Figure 4.13 (b)), we observe a difference of behaviours between the Ordinary and Firth MLEs such that the Ordinary MLEs are liberal, while the Firth MLEs are conservative.



(a) Wald test for $\hat{\boldsymbol{\beta}}_{\boldsymbol{ql}}$ slope



(b) Likelihood ratio (LR) test for $\hat{\boldsymbol{\beta}}_{\boldsymbol{ql}}$ slope

Figure 4.13: **False Positive Rates (FPR) at 5% significance level for the Wald and likelihood ratio (LR) tests based on the Ordinary and Firth MLEs**. Between/within block slope values are displayed along the x-axis. The FPR are shown with respect to the Unbalanced proportion design, varying numbers of subjects ($K \in \{3, 5, 10\}$) and network sizes ($n \in \{30, 50, 100, 500\}$ nodes).

4.5 Multi-subject Functional Connectivity fMRI Data

Our analysis considers the Placebo treatment data obtained from a crossover resting state fMRI study (Lynall et al., 2010; Towlson et al.), involving Placebo, Aripiprazole and Sulpiride treatment sessions. Within the Placebo treatment session, data was collected for 13 healthy volunteers (Controls) and 12 individuals diagnosed with schizophrenia (Patients). The full details of the fMRI acquisition and pre-processing can be found in the original references. In brief, the subjects were instructed to lay quietly in the scanner with their eyes closed for a duration of 17 minutes and 12 seconds. In each session, a total of 512 scans were taken with a repetition time of 2 seconds. Each such dataset was further corrected for motion artefacts according to the procedures described in Suckling et al. (2006) and then registered to the MNI standard space atlas. A Gaussian kernel of 6 mm was used to spatially smooth the registered images and the time series were high-pass filtered with a cutoff frequency of ≈ 0.008 Hz. Each subject's data was parcellated into 325 anatomically defined regions (ROIs) using the AFNI atlas and 28 regions were discarded due to missing data for some individuals. Averaged voxel time series in each region were further decomposed into four frequency scales by a discrete wavelet transform (Percival and Walden, 2000). Our subsequent analysis considers correlations in the frequency interval of $0.06 - 0.125$ Hz, as it has been shown in Lynall et al. (2010) that, within this frequency range, the differences between Controls and Patients were the most salient. Using the binarisation procedure outlined in Section 2.1.4 with $N = 128$ discrete wavelet coefficients, we obtain a 297×297 undirected and binary adjacency matrix for each subject.

To this data, we fit the Het-SBM, described in Section 4.1.3, for which we consider as covariates: age, premorbid intelligence[1] (Nart IQ; Nelson, 1982), and per-subject motion (see Section 2.1.3). All covariate vector values are column-wise assigned to the design matrix $\boldsymbol{D}$, so that the first two columns represent the intercept for Controls and the intercept for Patients, while the remaining columns are associated with the age, Nart IQ and motion, all of which are centred about their respective mean values.

4.5.1 Procedure for Functional Block Evaluations

In order to make functional interpretations of the estimated partition, we compare each estimated block of nodes against the 20 Intrinsic Connectivity Networks (ICN)[2] reported in the study of Laird et al. (2011). More precisely, we first classify each node in terms of their Intrinsic Connectivity Networks, using the coordinates of its ROI

[1] Subject's intellectual functioning prior to known or suspected brain disfunction or disease.

[2] Intrinsic Connectivity Networks are defined as large scale networks, simultaneously captured in resting-state and task-based neuroimaging data.

centroid, and then, for each estimated block of nodes, we compute the proportion of nodes in each of the 20 ICNs and assign to it the ICN with the highest proportion.

Note that this comparison with the ICNs reported in Laird et al. (2011) is rather rough due to the fact that we only use the coordinates of the ROI centroids and not the whole volumes of the ROIs (the latter information was unfortunately not available to us). Also, for some blocks, it seemed that the proportion of nodes in the selected ICN was relatively small and therefore, the ICN assignment for these blocks may not be relevant. Nevertheless, this rough classification of the estimated blocks may still be regarded as a useful exploratory tool.

4.6 Het-SBM Fit to Multi-subject Functional Connectivity fMRI Data

The Het-SBM estimates a multi-subject network decomposition with 24 blocks. Figure 4.14 shows the spatial location of each node and its corresponding block assignment. The blocks are further organised into panels (Panel A-F). A striking feature of this decomposition is that the Het-SBM reasonably well captures the spatial information, despite the fact that this information was not used during the estimation. For example, the nodes in Block 5 (see Panel D) belong to the spatially separated regions in the frontal and parietal lobes.

Panel A	Panel B	Panel C
Block 24 – ICN 1	Block 9 – ICN 6	Block 1 – ICN 11
Block 3 – ICN 2	Block 17 – ICN 6	Block 2 – ICN 12
Block 12 – ICN 2	Block 19 – ICN 7	Block 6 – ICN 12
Block 18 – ICN 2	Block 21 – ICN 7	Block 16 – ICN 12
Block 14 – ICN 5		
Block 20 – ICN 5		

Panel D	Panel E	Panel F
Block 5 – ICN 13	Block 7 – ICN 14	Block 4 – ICN 17
Block 10 – ICN 13	Block 23 – ICN 15	Block 15 – ICN 18
Block 11 – ICN 13	Block 13 – ICN 16	Block 8 – ICN 20
	Block 22 – ICN 16	

Table 4.2: **Correspondence between the Intrinsic Connectivity Networks (ICNs) published in Laird et al. (2011) and the 24 blocks estimated by the Het-SBM**.

To work out the approximate functional labels of the individual blocks, we contrast their nodes against the Intrinsic Connectivity Networks maps reported in Laird et al. (2011) (see Table 4.2). While this comparison is approximate, it is a

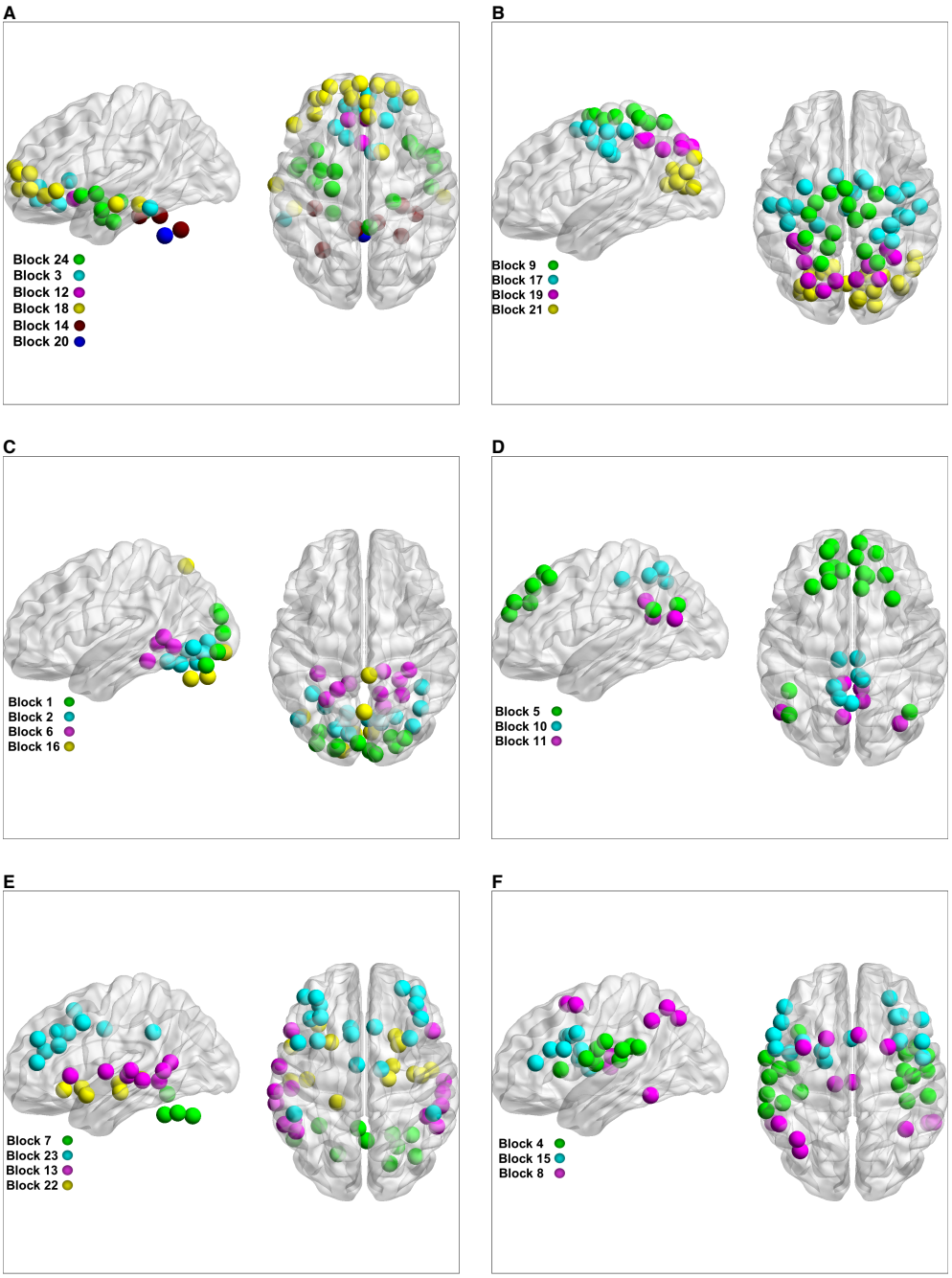


Figure 4.14: **Anatomical locations of individual nodes in the blocks, shown in terms of panels (Panel A-F)**. The plots were generated using the Matlab's BrainNet toolbox (Xia et al., 2013, `http://www.nitrc.org`, last accessed on the 9th of January 2015).

convenient exploratory tool that helps us assign relative functional labels to the estimated blocks.

Panel A: Broadly speaking, we found that the nodes in Block 24, 3, 12 and 18 are mostly associated with networks linked to emotional and autonomic processes. In particular, about 73% of nodes in Block 24 are linked to the visual recognition of emotions like fear, anger, happiness or humour, while more than 90% of nodes in Block 3 and 12 are associated with networks involved in olfaction and gustation. About 40% of nodes in Block 18 are involved in similar autonomic process and 24% in cognitive processes like reasoning, attention and memory. About 72% of nodes in Block 14 and the node in Block 20 (Right Cerebellar Tonsil) are linked to autonomic processes related to breathing.

Panel B: The majority of the nodes in Blocks 9 (36%), Block 17 (52%), Block 19 (100%) and Block 21 (32%) are associated with motor and visual processing.

Panel C: All the blocks in Panel C - Block 1 with 80%, Block 2 with 62 %, Block 6 with 40% and Block 16 with 50 % nodes - are linked to higher visual perception.

Panel D: All the blocks in Panel D - Block 5 with 39% , Block 10 with 29% and Block 11 with 86% nodes - are linked to the Default Mode Network.

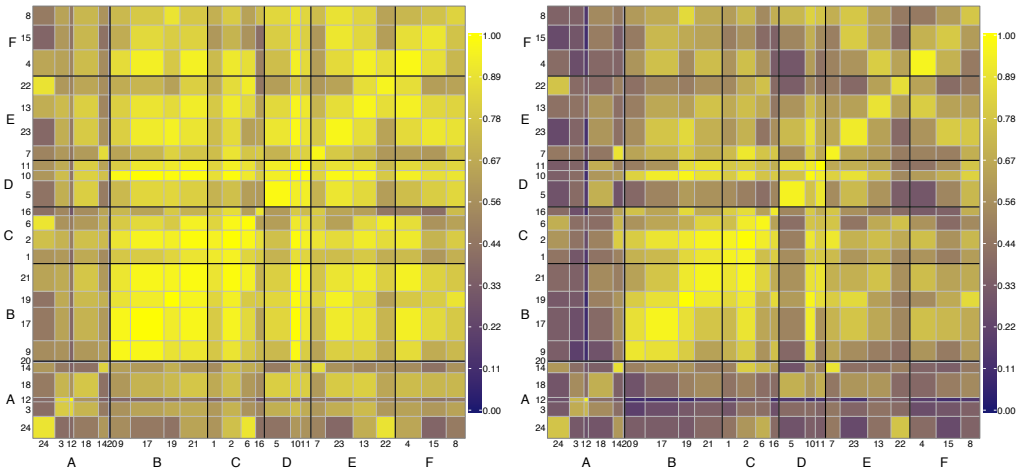
Panel E: About 80% of the nodes in Block 7 are located in the cerebellum, 37% of the nodes in Block 23 are linked to cognitive processes, while 63% of the nodes in Block 13 and 39% of the nodes in Block 22 are linked to audition and speech.

Panel F: This panel contains blocks related to speech and language (Block 4 with 67% and Block 15 with 35%). It is somewhat more challenging to functionally quantify Block 8 as its nodes are participating in a variety of networks. In particular, about 15 % of its nodes are involved in autonomic processes, 23 % in visuospatial processing and reasoning, 15 % in higher visual processing, 15 % in language and 31% are related to artefacts.

Inference

In Figure 4.15 (a) and (b), we show the estimated intercept values for the Control and Patient groups. The values are expressed as probabilities and denote the within-block or the between-block rates of connections. For convenience, the blocks are organised in terms of panels (Panel A-F). Although, both subject groups show reasonably similar patterns of connectivity structures, the Patients seems to be under-connected relative to the Controls. However, to obtain a finer view of the extent of their differences, we make some inferences on differences between the intercepts of the Patients vs. Controls.

In Figure 4.16, we show the corresponding Wald test score image thresholded at 5% after using a Bonferroni correction. Overall, we can observe significantly less



(a) **Controls** (b) **Patients**

Figure 4.15: **Estimated within/between block connectivity rates**. Individual block labels are shown on $(x - y)$ axis and each block cell is demarcated by a grey line. The blocks were further aggregated into 6 panels, labelled Panel A-F, which are visually separated by black lines. Largely similar patterns of connectivity are found in each population, though patients appear to have generally lower connection densities.

connections in the Patients compared to the Controls. In order to view these results in more detail, we discuss each panel (Panel A-F) separately.

Panel A: The most significant differences are found in the connection rates between Panel A (emotional and autonomic processes) and Panel B (motor and visual processes), especially within the connections of Blocks 3 & 17 and Blocks 18 & 17. Further significant results are found in its connection to Panel F (speech and language), particularly between Blocks 18 & 4.

Panel B: The most significant differences within Panel B (motor and visual processes) are seen in the connections of Block 21 with Blocks 9 &17. In particular, Blocks 9, 17 & 21 are significantly less connected with the blocks in Panel D (Default Mode network), Panel E (audition and speech) and Panel F (speech and language).

Panel C: There are almost no significant differences in the connectivity rates between the Patients and Controls in Panel C (higher visual perception). Also, the connection rates between this circuit and the rest of the network are weakly significant.

Panel D: The most significant differences are found in the way Panel D (Default

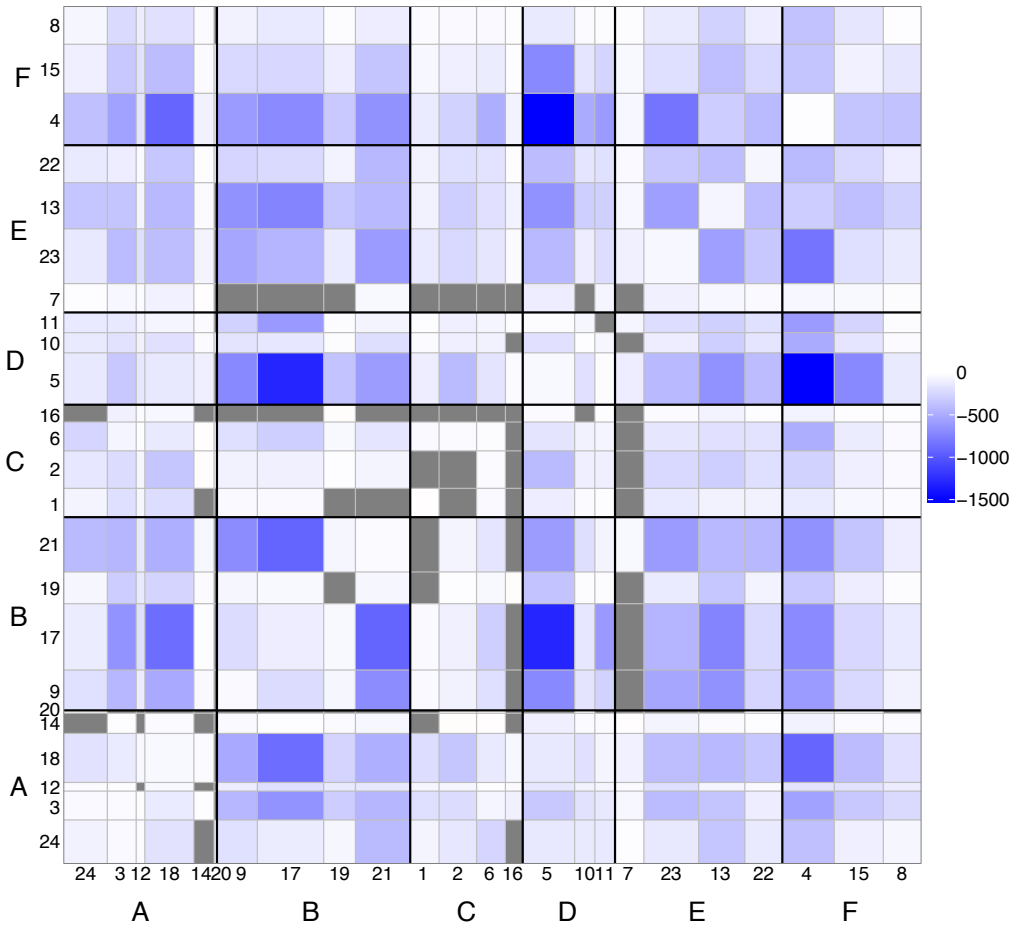


Figure 4.16: **Bonferroni thresholded (5%) LR score image of the intercept differences between Patients vs. Controls, interpretable as the Patient-Control difference in (logit-scale) connectivity rates at mean age, Nart IQ and motion (33.32 years, 113.06 Nart IQ and 7.40 mm)**. Note that the LR scores have been multiplied by the sign of $\boldsymbol{L}_{ql}\hat{\boldsymbol{\beta}}_{ql}$ to indicate direction of the effect. An increasing effect is shown in red and a decreasing effect is shown in blue.

Mode network), relates to Panel B (motor and visual processes) and Panel F (speech and language). In particular, the bulk of weaker connections is associated with the connectivity profile of Block 5.

Panel E: Overall, connection rate-differences are weakly significant with the exception of Block 23 (cognitive processes) & Block 4 (speech and language).

Panel F: Most of the blocks show weakly significant results.

Next, we tested for the effect of age on the connectivity rates on both subject

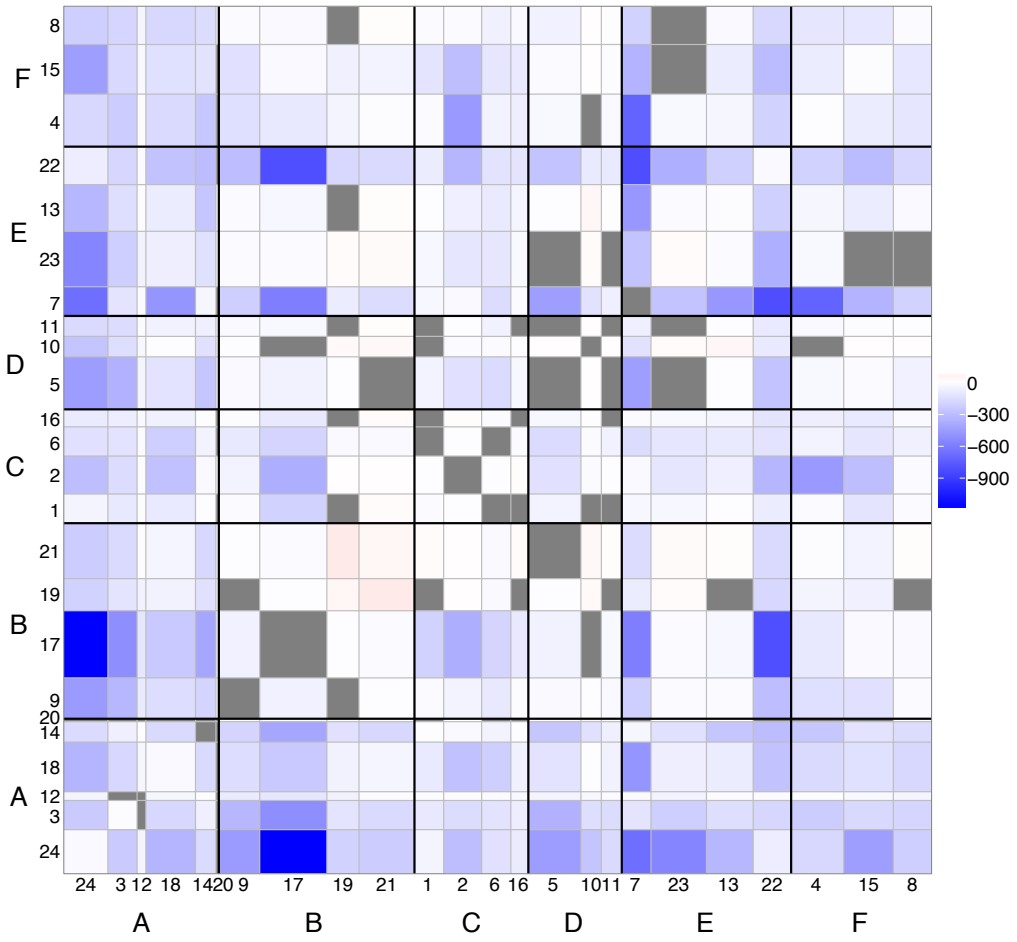


Figure 4.17: **Bonferroni thresholded (5%) LR score image of the common Age effect**. Note that the LR scores have been multiplied by the sign of $\boldsymbol{L}_{ql}\hat{\boldsymbol{\beta}}_{ql}$ to indicate the direction of the effect. An increasing effect is shown in red and a decreasing effect is shown in blue.

groups. In Figure 4.17, most of the results indicate that the connection rates decrease with increasing age. In particular, decreasing connectivity rates are strongly evident between Block 24 (emotional processes) and Block 17 (motor and visual processing), Block 17 (motor and visual processing) and Block 22 (audition and speech), Block 17 (motor and visual processing), and Block 5 (Default Mode Network) and Block 5 (Default Mode Network) and Block 4 (speech and language). Note that there is some weak evidence of an increasing connectivity with age within and between Blocks 19 & 21.

4.7 Discussion

In this chapter, we proposed three models that can be used to estimate a common network decomposition in multi-subject data. Each model accounts for different levels of variability between the subjects. The Binomial Stochastic Blockmodel (Bin-SBM) assumes no variability between subjects. The Homogeneous Stochastic Blockmodel (Hom-SBM) assumes that a covariate proportionally affects the connectivity across all blocks. The Heterogeneous Stochastic Blockmodel (Het-SBM) assumes that a covariate affects connectivity in the blocks as well as their mutual interactions.

Our results in Simulation I showed that the inference procedure with a Wald test may fail to control the FPR in some scenarios. Nevertheless, this behaviour is not necessary due to the test procedure itself, but more likely to the fact that, in these scenarios, the models had the tendency to fail to recover the true clustering. Indeed, in such cases, the connection patterns of (wrongly) estimated clusterings may not be representative of those of the true clustering. As the simulations are based on the assumption that the underlying connectivity patterns are the true null effect and not those of the estimated clusters, the test has the tendency to detect this difference, yielding an apparent liberal control of the FPR.

In connection to the cluster estimation, our simulation analysis did not investigate the impact of the regression parameters biases on the accuracy of cluster estimates. In particular, Simulation II showed that the Ordinary MLEs were severely biased while the Firth MLEs were more accurate. However, for the sake of the objective comparison between the Ordinary and Firth MLEs, we kept the cluster structure relatively simple and easy to estimate, but it would be very important to investigate the impact of the Ordinary MLE bias on the estimation of the cluster structure. For this, we would need additional simulations which would combine challenging cluster structures and connectivity rates. On these lines, it would be also interesting to investigate the differences in the clustering abilities between the three multi-subject models by allowing heterogenous covariate effects across blocks. We note that both of these points are left as a future work.

In terms of the accuracy of the inference procedures, we identified several sources of small sample behaviour, including the relative block sizes, number of subjects and sparse/saturated block connection rates. In particular, the Wald, Likelihood Ratio and permutation tests seem to be equally affected by small sample effect. So far, our analysis showed that, although the Ordinary MLEs are biased in small samples, the Wald test values based on their estimates tend to be very conservative, while their likelihood ratio test values tend to either conservative or liberal for small samples. The Wald tests based on Firth MLEs seemed to be liberal in small samples. In contrast, the likelihood ratio test based on Firth MLEs seemed were

somewhat better behaved. This is consistent with the results obtained by Heinze and Schemper (2002).

In the real data analysis, the Het-SBM identified anatomically and functionally plausible blocks, as well as differences in connectivity between groups (Controls vs. Patients) and their variations with age. Although our real data application was intended as an illustration of the Het-SBM, the results seem to be consistent with the previous literature which hypothesised that the schizophrenia is a disease of disconnectivity in the brain which heavily impacts regions associated with the Default Mode network (Van Den Heuvel and Pol, 2010). In that regard, the Het-SBM fit and its group inference pointed to even more details regarding the integrative properties of the Default Mode network which, taken globally, seems to be even less connected with other networks in Patients than in Controls. For example, nodes linked to the Default Mode Network seems to be less connected to the nodes associated with motor & visual processes, and with speech & language in Patients vs. Controls. However, further analyses would be needed to obtain deeper insights from this functional connectivity data and to validate these results with other datasets.

In the context of functional connectivity analyses, the adjacency matrices are typically constructed by the means of a hard thresholding. This implies that the researchers often select a threshold which preserves a pre-specified fraction of edges (e.g., 10% or 15% of the total number of connections). However, as this type of data imposes restrictions on each subject's total number of edges, we may encounter some artefacts in the later stages of the analysis. For example, let us consider two blocks of nodes (Block A and Block B) and a subject that happens to have a relatively higher correlation in Block A than in Block B, but the same correlation as all other subjects in Block B. To maintain a fixed connection rate, that subject must have a relatively higher (correlation) binarisation threshold, but as a result will be seen to have relatively lower connectivity in Block B. For such cases, the Hom-SBM model would be inappropriate as it can only model a homogeneous increase or decrease of connections for all the blocks. Therefore, for this kind of thresholding, the Het-SBM model would be more appropriate as it can model separate effects across blocks, allowing for decreasing effects in some blocks and increasing effects in others.

The Het-SBM and the Hom-SBM estimate a cluster structure on the residuals from a logistic regression model or, in other words, on the variability between subjects that has not been accounted for by the covariates in the model. Consequently, the residuals are not changed with the additional nuisance covariates and, therefore, the cluster structure remains approximately the same. To investigate this point, we also estimated a cluster structure on the real data using density (i.e. the proportion of observed edges over all possible edges, computed for each subject) as an additional nuisance covariate. Despite this additional covariate in the model, the

Het-SBM estimated a fit with 24 blocks whose structure was very similar to the fit presented in the paper (ARI score of 0.81).

In a wider context, the models proposed in this chapter are complementary to the models described in Mariadassou et al. (2010) which utilise Generalised Linear Models with edge-based covariates; and in Zanghi et al. (2010) which assume Normally distributed node-based covariates as additional terms in the likelihood. Both modelling approaches offer some interesting applications to fMRI data analysis, particularly if researchers wish to account for spatial information in the data, either expressed as a distance between the nodes or as a node-based vector of coordinates. Using the framework presented in this work, one can easily consider multi-subject extensions of these two models.

CHAPTER 5

Generalised Linear Mixed Stochastic Blockmodelling and Inference in Multi-subject Networks

In Chapter 4, we investigated the joint cluster analysis of independent multi-subject network data. For such data, we proposed two non-trivial extensions of the SBM, namely the Homogenous SBM (Hom-SBM) and the Heterogeneous SBM (Het-SBM). Each of these two models accounts for different levels of variability between the subjects with a logistic regression model, which embodies either a 'homogeneous' (common) or a 'heterogeneous' (interacting) relationship between the block structure and subject specific covariates like age, gender or IQ. In addition to their clustering abilities, both models also provide inference tools to detect group differences in connectivity rates.

However, although these models are useful for the analysis of independent multi-subject networks, they may not be appropriate in the cases where some form of dependence exists in the data. This may be generally attributed to two sources. First, if covariates do not fully explain inter-subject differences in connectivity rates, there may still be some dependence within the individual elements of a block structure. For example, the prevalence of edges in one subject's block (e.g., the block (q, l)) may be consistently over-estimated and, therefore, this lack of fit may be randomly distributed over subjects. Second, data may comprise more than one network per each subject inducing a repeated measures correlation. The latter data typically occurs in studies with multiple scans per-subject, who are either scanned at different time points or after being exposed to different experimental conditions, such as different drug effects (e.g., a cross-over design). Other examples of such data may include combined outcomes of different imaging modalities or data which combines different classes of connectivity (e.g., structural and functional) where

the goal would be to answer questions regarding the similarities of their networks' organisations.

In order to model within-subject correlated data, in this chapter, we investigate an extension of the SBM which uses a logistic mixed regression model with per-subject random intercepts. Here, we focus only on a 'heterogeneous' association between the subjects' covariates and the block structure, meaning that each block element has its own regression model. This model is the natural generalisation of the Het-SBM of the previous chapter, and should be able to represent a wide range of cluster structures.

In the reminder of this chapter, we define the Heterogenous Mixed SBM (Het-Mixed-SBM) and discuss estimation strategies based on the Ordinary Maximum Likelihood approach and points related to its practical implementation. We evaluate the proposed model over a range of different Monte Carlo simulations and we illustrate it on the data described in Section 4.5. Finally, we finish with discussion and future work.

5.1 Heterogeneous Mixed Stochastic Blockmodel (Het-Mixed-SBM)

To motivate what becomes a quite complex model, we first consider a simple logistic regression model with a random intercept per subject illustrated on a single block of binary data extracted from a known cluster structure. We describe the likelihood of the mixed effect model pertaining to this block and use it as a basis upon which the more complex Het-Mixed-SBM is developed. In particular, the Het-Mixed-SBM considers additional block-specific fixed effects ('heterogeneous' fixed effects) as well as subject-specific random effects for each block element ('heterogeneous' random effects). For this model, we discuss its optimisation strategy and its ICL criterion, and we finish by detailing the proposed estimation algorithm.

5.1.1 Logistic Regression with a Random Intercept

In this example, we assume that the cluster structure is known and we only consider data for a single block element for which we specify a logistic regression model with a random intercept per subject. Thus, given a subject k ($k = 1, \dots, K$) and its repeated measurement t ($t = 1, \dots, T$), we write the edge data for a particular block (q, l) to be a vector $\boldsymbol{x}_{kt}$ of n^* data points. Since the block assignment is known, the variable related to the block assignment ($\boldsymbol{Z}$) is redundant and, for each data point $s = 1, \dots, n^*$, the model can be stated as

$$X_{skt}|R_k = r_k \sim \text{Bernoulli}(\pi_{kt}),$$

$$\log\left(\frac{\pi_{kt}}{1-\pi_{kt}}\right) = \boldsymbol{d}_{kt}^{\top}\boldsymbol{\beta} + r_k, \quad \pi_{kt} = \frac{e^{\boldsymbol{d}_{kt}^{\top}\boldsymbol{\beta}+r_k}}{1+e^{\boldsymbol{d}_{kt}^{\top}\boldsymbol{\beta}+r_k}},$$
$$R_k \overset{iid}{\sim} \text{Normal}(0,\sigma^2),$$

where $\boldsymbol{d}_{kt}$ is a $P\times 1$ vector of covariates associated with the k-th subject and its t-th measurement and R_k are independent identically distributed (iid) subject-specific random intercepts with variance σ^2. As noted in McCulloch (1997), the density of this simple model can be stated as

$$\begin{aligned} f(\boldsymbol{x};\boldsymbol{\beta},\sigma^2) &= \prod_{k=1}^{K}\int_{-\infty}^{+\infty}\Big[\prod_{s=1}^{n^*}\prod_{t=1}^{T} f(x_{skt}|r_k;\boldsymbol{\beta})\Big] f(r_k;\sigma^2)\,dr_k \\ &= \prod_{k=1}^{K}\frac{1}{\sqrt{2\Pi\sigma^2}}\int_{-\infty}^{+\infty}\Big[\prod_{s=1}^{n^*}\prod_{t=1}^{T} e^{x_{skt}\boldsymbol{d}_{kt}^{\top}\boldsymbol{\beta}+x_{skt}r_k+\log(1-\pi_{kt})}\Big] e^{-\frac{r_k^2}{2\sigma^2}}\,dr_k \end{aligned}$$

and its log likelihood is given as

$$\log f(\boldsymbol{x};\boldsymbol{\beta},\sigma^2) = -\frac{K}{2}\log(2\Pi\sigma^2) + \sum_{k=1}^{K}\sum_{s=1}^{n^*}\sum_{t=1}^{T} x_{skt}\boldsymbol{d}_{kt}^{\top}\boldsymbol{\beta} + \sum_{k=1}^{K}\log\int_{-\infty}^{+\infty} e^{h_k(\boldsymbol{\beta},\sigma^2)}\,dr_k,$$

where,

$$h_k(\boldsymbol{\beta},\sigma^2) = \sum_{s=1}^{n^*}\sum_{t=1}^{T}\big[x_{skt}r_k + \log(1-\pi_{kt})\big] - \frac{r_k^2}{2\sigma^2}.$$

We now elaborate this model to account for the entire block structure with an unknown membership.

5.1.2 Model Definition

In this section, we describe the Het-Mixed-SBM. Let us consider the Bernoulli random variable X_{ijkt} such that the indices $i,j = 1,\ldots,n$, $k = 1,\ldots,K$, and $t = 1,\ldots,T$ denote the nodes, subjects, and per-subject measurements, respectively. For each of K subjects and T measurements, we use $\boldsymbol{X}_{kt} = ((X_{ijkt}))_{1\le i\neq j\le n}$ to denote their $n\times n$ random and symmetric matrix and, for the simplicity of notation, we use $\boldsymbol{X}$ to denote the set of all such random variables.

Similarly to Section 4.1 (see Eq. (4.2) and Eq. (4.2)), in this model, we assume that the cluster structure is common across all subjects and their repeated measurements. Thus, for a fixed Q, the latent random variable $\boldsymbol{Z}_i$ indicates the cluster membership of node V_i and it is assumed to independently follow a Categor-

ical distribution with Q possible outcomes,

$$\boldsymbol{Z}_i \sim \text{ Categorical}(Q, \boldsymbol{\alpha}), \tag{5.1}$$

with individual probabilities of success $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_Q)$ such that $\sum_{q=1}^{Q} \alpha_q = 1$. Writing $\boldsymbol{Z} = ((Z_{iq}))_{1 \leq i \leq n, 1 \leq q \leq Q}$ for the $n \times Q$ matrix, its probability mass function is

$$f(\boldsymbol{z}; \boldsymbol{\alpha}) = \prod_{q=1}^{Q} \prod_{i=1}^{n} \alpha_q^{z_{iq}}. \tag{5.2}$$

Finally, the Het-Mixed-SBM can be summarised as follows

$$X_{ijkt} | Z_{iq} = 1, Z_{jl} = 1, R_{qlk} = r_{qlk} \sim \text{Bernoulli}(\pi_{qlkt}), \tag{5.3}$$

$$\log\left(\frac{\pi_{qlkt}}{1 - \pi_{qlkt}}\right) = \boldsymbol{d}_{kt}^{\top} \boldsymbol{\beta}_{ql} + r_{qlk}, \quad \pi_{qlkt} = \frac{e^{\boldsymbol{d}_{kt}^{\top} \boldsymbol{\beta}_{ql} + r_{qlk}}}{1 + e^{\boldsymbol{d}_{kt}^{\top} \boldsymbol{\beta}_{ql} + r_{qlk}}}, \tag{5.4}$$

$$R_{qlk} \overset{iid}{\sim} \text{Normal}(0, \sigma_{ql}^2), \tag{5.5}$$

where R_{qlk} represents a subject specific random intercept for the block (q, l), $\boldsymbol{d}_{kt}$ is a $P \times 1$ vector of covariates associated with the k-th subject and its t-th measurement, and $\boldsymbol{\beta}_{ql}$ is a $P \times 1$ vector of regression coefficients for the block (q, l). We note that σ_{ql}^2 is the variance of each block or block-to-block regression and it is collectively written as $\boldsymbol{\sigma}^2 = ((\sigma_{ql}^2))_{1 \leq q, l \leq Q}$. In particular, the probability mass function of $\boldsymbol{x}$ given $\boldsymbol{z}$ can be written as

$$f(\boldsymbol{x} | \boldsymbol{z}; \boldsymbol{\beta}, \boldsymbol{\sigma}^2) = \prod_{k=1}^{K} \prod_{q=1}^{Q} \prod_{l=1}^{Q} \int_{-\infty}^{+\infty} \Bigg(\Bigg[\sum_{i=1}^{n} \sum_{j>i}^{n} \sum_{t=1}^{T} [f(x_{ijkt} | z_{iq}, z_{jl}, r_{qlk}; \beta_{ql})]^{z_{iq} z_{jl}} \Bigg] \times f(r_{qlk}; \sigma_{ql}^2) \Bigg) dr_{qlk}. \tag{5.6}$$

5.1.3 Estimation and Model Selection

Similarly to the models considered in Chapter 4, we use the variational approach (see Section 2.4.1) to estimate the models parameters $(\boldsymbol{\tau}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\sigma}^2)$. Using the notation γ_{ijql} introduced in Section 4.1.3 (Eq. (4.33)), the variational bound is given as

$$\begin{aligned} \mathcal{J}(f^*(\boldsymbol{z}; \boldsymbol{\tau}); \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\sigma}^2) = & \frac{1}{2} \sum_{k=1}^{K} \sum_{i=1}^{n} \sum_{j \neq i}^{n} \sum_{t=1}^{T} \sum_{q=1}^{Q} \sum_{l \geq q}^{Q} \gamma_{ijql} x_{ijkt} \boldsymbol{d}_{kt}^{\top} \boldsymbol{\beta}_{ql} \\ & - \frac{K}{2} \sum_{q=1}^{Q} \sum_{l \geq q}^{Q} \log(2 \Pi \sigma_{ql}^2) \end{aligned}$$

$$+\sum_{k=1}^{K}\sum_{q=1}^{Q}\sum_{l\geq q}^{Q}\log\int_{-\infty}^{+\infty}e^{h_{qlk}}\,dr_{qlk}$$
$$+\sum_{i=1}^{n}\sum_{q=1}^{Q}\tau_{iq}\log(\alpha_q)-\sum_{i=1}^{n}\sum_{q=1}^{Q}\tau_{iq}\log(\tau_{iq}), \tag{5.7}$$

$$h_{qlk}=\frac{1}{2}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\sum_{t=1}^{T}\gamma_{ijql}x_{ijkt}r_{qlk}+\frac{1}{2}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\sum_{t=1}^{T}\gamma_{ijql}\log(1-\pi_{qlkt})-\frac{r_{qlk}^2}{2\sigma_{ql}^2}. \tag{5.8}$$

For a fixed Q, we want to maximise the variational bound, defined by Eq. (5.7) with respect to the variational parameter $\boldsymbol{\tau}$, as well as the parameters $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ and $\boldsymbol{\sigma}^2$. The optimal variational parameter $\boldsymbol{\tau}$ satisfy the fixed point relations

$$\hat{\tau}_{iq}\propto\hat{\alpha}_q\Big(\prod_{k=1}^{K}\prod_{l=1}^{Q}\exp\left(\frac{\hat{I}_{iqlk}}{\hat{I}_{qlk1}}\right)\Big)\prod_{k=1}^{K}\prod_{j\neq i}^{n}\prod_{t=1}^{T}\prod_{l=1}^{Q}\exp\left(\hat{\tau}_{jl}x_{ijkt}\boldsymbol{d}_{kt}^{\top}\hat{\beta}_{ql}\right), \tag{5.9}$$

where, we use the notation I_{qlk1} and I_{iqlk} to denote the following integral expressions

$$\hat{I}_{qlk1}=\int_{-\infty}^{+\infty}e^{\hat{h}_{qlk}}\,dr_{qlk},$$
$$\hat{I}_{iqlk}=\frac{\partial\hat{I}_{qlk1}}{\partial\tau_{iq}}=\sum_{j\neq i}^{n}\hat{\tau}_{jl}\sum_{t=1}^{T}x_{ijkt}\int_{-\infty}^{+\infty}e^{\hat{h}_{qlk}}r_{qlk}\,dr_{qlk}$$
$$+\sum_{i\neq j}^{n}\hat{\tau}_{jl}\sum_{t=1}^{T}\int_{-\infty}^{+\infty}e^{\hat{h}_{qlk}}\log(1-\hat{\pi}_{qlkt})\,dr_{qlk}. \tag{5.10}$$

In particular, the optimisation of $\boldsymbol{\tau}$ follows the same procedure of fixed-point iteration used for the models of Chapter 4 (see Sections 4.1.2 and 4.1.3), and, therefore, $\hat{h}_{qlk}$ is evaluated on the previous values of $\boldsymbol{\tau}$. As before, the estimates of $\boldsymbol{\alpha}$ are given as

$$\hat{\alpha}_q=\frac{1}{n}\sum_{i=1}^{n}\hat{\tau}_{iq}. \tag{5.11}$$

We next turn our attention to the optimisation of the variational bound for $\boldsymbol{\beta}$ and $\boldsymbol{\sigma}^2$. For the clarity of the subsequent discussions and easy referencing, we provide a list of integrals $I_{qlk1}-\boldsymbol{I}_{qlk6}$ that will be used in the estimating equations of $\boldsymbol{\beta}$ and $\boldsymbol{\sigma}^2$. Thus, we have

$$I_{qlk1}=\int_{-\infty}^{+\infty}e^{h_{qlk}}\,dr_{qlk}, \tag{5.12}$$

$$I_{qlk2} = \frac{\partial I_{qlk1}}{\partial \sigma_{ql}^2} = \frac{1}{2\sigma_{ql}^4} \int_{-\infty}^{+\infty} e^{h_{qlk}} r_{qlk}^2 \, dr_{qlk}, \tag{5.13}$$

$$\boldsymbol{I}_{qlk3} = \frac{\partial I_{qlk1}}{\partial \boldsymbol{\beta}_{ql}} = \left(-\frac{1}{2} \sum_{i=1}^{n} \sum_{j \neq i}^{n} \gamma_{ijql} \right) \sum_{t=1}^{T} \boldsymbol{d}_{kt} \int_{-\infty}^{+\infty} e^{h_{qlk}} \pi_{qlkt} \, dr_{qlk}, \tag{5.14}$$

$$I_{qlk4} = \frac{\partial I_{qlk2}}{\partial \sigma_{ql}^2} = \frac{1}{4\sigma_{ql}^8} \int_{-\infty}^{+\infty} e^{h_{qlk}} r_{qlk}^4 \, dr_{qlk} - \frac{1}{\sigma_{ql}^6} \int_{-\infty}^{+\infty} e^{h_{qlk}} r_{qlk}^2 \, dr_{qlk}, \tag{5.15}$$

$$\begin{aligned} \boldsymbol{I}_{qlk5} = \frac{\partial \boldsymbol{I}_{qlk3}^{\top}}{\partial \boldsymbol{\beta}_{ql}} = & \left(-\frac{1}{2} \sum_{i=1}^{n} \sum_{j \neq i}^{n} \gamma_{ijql} \right)^2 \sum_{t=1}^{T} \sum_{t'=1}^{T} \boldsymbol{d}_{kt} \boldsymbol{d}_{kt'}^{\top} \int_{-\infty}^{+\infty} e^{h_{qlk}} \pi_{qlkt} \, \pi_{qlkt'} \, dr_{qlk} \\ & + \left(-\frac{1}{2} \sum_{i=1}^{n} \sum_{j \neq i}^{n} \gamma_{ijql} \right) \sum_{t=1}^{T} \boldsymbol{d}_{kt} \boldsymbol{d}_{kt}^{\top} \int_{-\infty}^{+\infty} e^{h_{qlk}} \pi_{qlkt} (1 - \pi_{qlkt}) \, dr_{qlk}, \end{aligned} \tag{5.16}$$

$$\boldsymbol{I}_{qlk6} = \frac{\partial \boldsymbol{I}_{qlk3}}{\partial \sigma_{ql}^2} = \frac{1}{2\sigma_{ql}^4} \left(-\frac{1}{2} \sum_{i=1}^{n} \sum_{j \neq i}^{n} \gamma_{ijql} \right) \sum_{t=1}^{T} \boldsymbol{d}_{kt} \int_{-\infty}^{+\infty} e^{h_{qlk}} r_{qlk}^2 \pi_{qlkt} \, dr_{qlk}. \tag{5.17}$$

Taking the partial derivatives of the variational bound $\mathcal{J}(f^*(\boldsymbol{z};\boldsymbol{\tau});\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\sigma}^2)$ (Eq. (5.7)) with respect to $\boldsymbol{\beta}_{ql}$ and σ_{ql}^2, for an individual block (q,l), we can form a score vector, $\boldsymbol{U}(\boldsymbol{\beta}_{ql}, \sigma_{ql}^2)$ as a $(P+1) \times 1$ vector of first derivatives $\boldsymbol{U}(\boldsymbol{\beta}_{ql}, \sigma_{ql}^2) = (\boldsymbol{U}(\boldsymbol{\beta}_{ql})^{\top}, U(\sigma_{ql}^2))^{\top}$ whose equations are given as

$$\boldsymbol{U}(\boldsymbol{\beta}_{ql}) = \frac{1}{2} \sum_{k=1}^{K} \sum_{i=1}^{n} \sum_{j \neq i}^{n} \sum_{t=1}^{T} \gamma_{ijql} x_{ijkt} \boldsymbol{d}_{kt} + \sum_{k=1}^{K} \frac{\boldsymbol{I}_{qlk3}}{I_{qlk1}} \tag{5.18}$$

$$U(\sigma_{ql}^2) = -\frac{K}{2\sigma_{ql}^2} + \sum_{k=1}^{K} \frac{I_{qlk2}}{I_{qlk1}}. \tag{5.19}$$

Similarly, the negative second order partial derivatives of the variational bound yield the observed Fisher Information matrix $\boldsymbol{\mathcal{I}}(\boldsymbol{\beta}, \boldsymbol{\sigma}^2)$, which appears in this model as a block diagonal matrix whose overall dimension is $(P+1)Q(Q+1)/2 \times (P+1)Q(Q+1)/2$. In particular, each diagonal block element is noted as $\boldsymbol{\mathcal{I}}(\boldsymbol{\beta}_{ql}, \sigma_{ql}^2)$ and is a $(P+1) \times (P+1)$ matrix that can be represented as

$$\boldsymbol{\mathcal{I}}(\boldsymbol{\beta}_{ql}, \sigma_{ql}^2) = \left[\begin{array}{c|c} \boldsymbol{\mathcal{I}}_1(\boldsymbol{\beta}_{ql}) & \boldsymbol{\mathcal{I}}_2(\sigma_{ql}^2, \boldsymbol{\beta}_{ql}) \\ \hline \boldsymbol{\mathcal{I}}_2^{\top}(\sigma_{ql}^2, \boldsymbol{\beta}_{ql}) & \mathcal{I}_3(\sigma_{ql}^2) \end{array} \right], \tag{5.20}$$

such that

$$\boldsymbol{\mathcal{I}}_1(\boldsymbol{\beta}_{ql}) \equiv -\frac{\partial^2 \mathcal{J}(f^*(\boldsymbol{z};\boldsymbol{\tau});\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\sigma}^2)}{\partial\boldsymbol{\beta}_{ql}\partial\boldsymbol{\beta}_{ql}^\top} = -\sum_{k=1}^{K}\left[\frac{\boldsymbol{I}_{qlk5}}{I_{qlk1}} - \frac{\boldsymbol{I}_{qlk3}\boldsymbol{I}_{qlk3}^\top}{I_{qlk1}^2}\right], \tag{5.21}$$

$$\boldsymbol{\mathcal{I}}_2(\sigma_{ql}^2,\boldsymbol{\beta}_{ql}) \equiv -\frac{\partial^2 \mathcal{J}(f^*(\boldsymbol{z};\boldsymbol{\tau});\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\sigma}^2)}{\partial\sigma_{ql}^2\partial\boldsymbol{\beta}_{ql}} = -\sum_{k=1}^{K}\left[\frac{\boldsymbol{I}_{qlk6}}{I_{qlk1}} - \frac{\boldsymbol{I}_{qlk3}I_{qlk2}}{I_{qlk1}^2}\right], \tag{5.22}$$

$$\mathcal{I}_3(\sigma_{ql}^2) \equiv -\frac{\partial^2 \mathcal{J}(f^*(\boldsymbol{z};\boldsymbol{\tau});\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\sigma}^2)}{\partial\sigma_{ql}^2\partial\sigma_{ql}^2} = -\frac{K}{2\sigma_{ql}^4} - \sum_{k=1}^{K}\left[\frac{I_{qlk4}}{I_{qlk1}} - \frac{I_{qlk2}^2}{I_{qlk1}^2}\right]. \tag{5.23}$$

The estimates of $\boldsymbol{\beta}_{ql}$ and σ_{ql}^2 can be found with the Newton-Raphson formula, so that, for the (m)-th step, we have

$$\begin{pmatrix}\boldsymbol{\beta}_{ql}\\ \sigma_{ql}^2\end{pmatrix}^{(m)} = \begin{pmatrix}\boldsymbol{\beta}_{ql}\\ \sigma_{ql}^2\end{pmatrix}^{(m-1)} + \mathcal{I}(\boldsymbol{\beta}_{ql}^{(m-1)}, {\sigma_{ql}^2}^{(m-1)})^{-1}\boldsymbol{U}(\boldsymbol{\beta}_{ql}^{(m-1)}, {\sigma_{ql}^2}^{(m-1)}). \tag{5.24}$$

As discussed in Chapter 4 (see Eq. 4.45), the ICL criterion is derived from $\log f(\boldsymbol{x},\boldsymbol{z}|\mathrm{m}_Q) = \log f(\boldsymbol{x}|\boldsymbol{z},\mathrm{m}_Q) + \log f(\boldsymbol{z}|\mathrm{m}_Q)$. The part of the ICL criterion which depends on the $\log f(\boldsymbol{z}|\mathrm{m}_Q)$ is the same as in Section 2.4.1. For the $\log f(\boldsymbol{x}|\boldsymbol{z},\mathrm{m}_Q)$, we use the BIC criterion. For this, we note that the total number of parameters in $\boldsymbol{\beta}$ is $\frac{Q(Q+1)}{2}P$ and in $\boldsymbol{\sigma}^2$ is $\frac{Q(Q+1)}{2}$, and the total number of data points in $\boldsymbol{x}$ is $\frac{n(n-1)}{2}KT$. Finally, the ICL criterion of the Het-Mixed-SBM is defined as

$$\begin{aligned}\mathrm{ICL}(\mathrm{m}_Q) = &\log f(\boldsymbol{x},\hat{\boldsymbol{z}}|\mathrm{m}_Q,\hat{\boldsymbol{\alpha}},\hat{\boldsymbol{\beta}},\hat{\boldsymbol{\sigma}}^2) - \frac{1}{2}\left(\frac{Q(Q+1)}{2}(P+1)\right)\log\left[\frac{n(n-1)}{2}KT\right]\\ &- \frac{Q-1}{2}\log[n].\end{aligned} \tag{5.25}$$

Note that we do not use any Firth-type regularisation here, due to additional complexity related to the random effect.

5.1.4 Estimation Algorithm

We propose an estimation algorithm that follows the same structure as the algorithms described in Sections 4.1.2 and 4.1.3. Starting with some initial values for $\boldsymbol{\tau}^0$, we iteratively update the model parameters according to the two steps

1. $(\boldsymbol{\alpha}^{(h+1)},\boldsymbol{\beta}^{(h+1)},{\boldsymbol{\sigma}^2}^{(h+1)}) = \underset{(\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\sigma}^2)}{\operatorname{argmax}}\left[\mathcal{J}(f^*(\boldsymbol{z};\boldsymbol{\tau}^{(h)});\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\sigma}^2)\right]$,
2. $\boldsymbol{\tau}^{(h+1)} = \underset{\boldsymbol{\tau}}{\operatorname{argmax}}\left[\mathcal{J}(f^*(\boldsymbol{z};\boldsymbol{\tau});\boldsymbol{\alpha}^{(h+1)},\boldsymbol{\beta}^{(h+1)},{\boldsymbol{\sigma}^2}^{(h+1)}))\right]$,

until the convergence is obtained. The convergence is measured by the relative changes of the parameter estimates and the improvement of the variational bound.

In the first step, $\boldsymbol{\alpha}$ is updated according to Eq. (5.11) while the parameters $(\boldsymbol{\beta}, \boldsymbol{\sigma}^2)$ are updated according to the Newton-Raphson algorithm (Eq. (5.24)). In the second step, $\boldsymbol{\tau}$ is updated according to Eq. (5.9).

The Newton-Raphson algorithm uses a naïve estimate of $\hat{\pi}_{ql}$ as starting values for the intercept β_{ql} (i.e. $\hat{\beta}_{ql} = \log(\hat{\pi}_{ql}/(1-\hat{\pi}_{ql}))$. In particular, for a given block element (i.e. block (q,l)), $\hat{\pi}_{ql}$ is the ratio of the sum of its observed edges across the K subjects and T measurements and the total number of possible edges. The starting values for $\boldsymbol{\sigma}^2$ are based on the strategy outlined in Demidenko (2004) whose objective is to estimate the random effects r_{qlk} and their sample variance, from which we finally obtain the initial estimates of $\boldsymbol{\sigma}^2$. The full details of this procedure can be found in Appendix C.1.

At each step of the Newton-Raphson algorithm, we require a numerical approximation of the six integrals (Eq. (5.12)-(5.17)). In practice, such integrals can be reasonably well estimated with the adaptive Gauss-Hermite quadrature approximation (Lesaffre and Spiessens, 2001; Liu and Pierce, 1994) whose implementation in `R` is available via the function `integrate` (R Core Team, 2015; Piessens et al., 1983). Although this computational strategy offers relatively quick and accurate approximations, in few examples, we encountered some instabilities and numerical issues. First, some integrals were evaluated as zeros which caused the variational bound to diverge to $-\infty$. The main reason for this was that the maximal value attained by the function h_{qlk} was very small (e.g., -1000), making the integrand numerically equal to zero over the whole domain of integration. To fix this, we used a reparametrisation such that the exponent h_{qlk} is counterbalanced by a quantity c_{qlk} which corresponds to the maximum of h_{qlk}, taken with respect to r_{qlk} while the values of ${\boldsymbol{\beta}_{ql}}^{(m)}$ and ${\sigma^2_{ql}}^{(m)}$ are assumed to be fixed. Thus, we have

$$c_{qlk} = - \max_{r_{qlk} \in [-1000, +1000]} \Big(h_{qlk}({\boldsymbol{\beta}_{ql}}^{(m)}, {\sigma^2_{ql}}^{(m)}) \Big), \tag{5.26}$$

and then each integral in the list $I_{qlk1} - \boldsymbol{I}_{qlk6}$ (see Eq. (5.12) - (5.17)) can be reparametrised in the following way: $I_{qlk1} = e^{c_{qlk}} \int_{-\infty}^{+\infty} e^{h_{qlk} - c_{qlk}} dr_{qlk}$. The details of this reparametrisation can be found in Appendix C.2. Second, we noticed several practical examples in which the function `integrate` would fail to provide satisfactory estimates of the integrals. This typically occurs in scenarios when the integrand is zero over most of its integration range and the quadrature approximation fails to detect the range of values for which the contribution to the integral is non-zero. We solved this issue by systematically reducing the range of integration. More precisely, this was achieved by first detecting the value of r_{qlk} corresponding to the maximum of h_{qlk} which is then used as the centre of the reduced interval of integration. Then, starting by this value as unique point of integration, we iteratively increased the

interval of integration by a value 1 in each direction with the stopping criterion that the values of h_{qlk} at the endpoints of the new interval were inferior to 10^{-10} times the maximum value of h_{qlk}. Noting that h_{qlk} is a concave function, we were sure that all the values outside the reduced interval of integration were also inferior to 10^{-10} times the maximum value of h_{qlk} and hence negligible. We used the same step-halving procedure described in Section 4.1.2 to facilitate the convergence of the Newton-Raphson algorithm.

5.1.5 Inference

Similarly to the Het-SBM, we can use a Wald test or a likelihood ratio test to make inference on a combination of the fixed effects parameters and, as before, these inferences are conditional on the block assignments. Since the Het-Mixed-SBM has block-specific parameters, the inference procedures for a linear combination of the fixed effects ($\boldsymbol{\beta}_{ql}$) have the same form as the Het-SBM described in Section 4.2 and, thus, we do not repeat them here. Note simply that the relevant Fisher's Information matrix for inference on $\boldsymbol{\beta}_{ql}$ is given in Eq. (5.21).

5.2 Methodology of Simulation

In this simulation, we consider different number of subjects ($K \in \{10, 20, 40\}$) with two network sizes ($n \in \{50, 100\}$ nodes). In addition to this, we also consider two scenarios of repeated measurements (or visits, $T \in \{1, 3\}$) and we consider three different variance values ($\boldsymbol{\sigma}^2 \in \{0.5, 1, 2\}$) which are assumed to be constant across the entire block structure. We also consider a single fixed effect, that of age, induced as 0 which is assumed to be homogeneous over the entire block structure. Due to the increased computational burden, the case with 80 subjects was also considered, but only on the networks with 50 and 100 nodes, $\boldsymbol{\sigma}^2 = 1$ and $T \in \{1, 3\}$. Binary network edges are generated according to four different types of connectivity structures, labelled PI3 & PI6-8 (see Figure 5.1). Similarly to the simulations settings described in Section 4.3, we consider three proportion designs (Balanced, Mildly Unbalanced and Unbalanced) which can be found in Table 4.1. In particular, we fix the total number of blocks to be three ($Q = 3$) and order the sizes of each block so that $n_1 \geq n_2 \geq n_3$ and we follow the same strategy as in Section 4.3 to identify the model parameters. The comparison between the estimated clustering and the true partition is carried out using the Adjusted Rand Index (ARI) discussed in Section 2.5 (Eq. (2.40)). For each combination of simulation parameters, we generated $S = 1000$ networks and fit each with the Het-Mixed-SBM, where we use only one restart for the algorithm and fix $Q = 3$.

The quality of parameter estimates is assessed with the root mean square

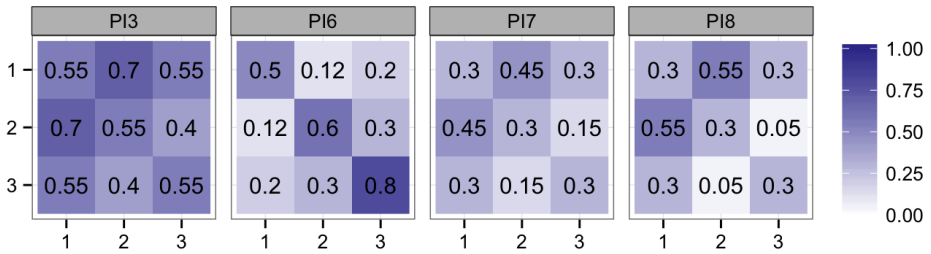


Figure 5.1: **Designs of connectivity structures**. Each of four different cases portraits various levels of homogeneity in cluster structures.

error (RMSE) and bias (Bias),

$$\text{RMSE}(\hat{\boldsymbol{\beta}}_{ql}^{(p)}) = \left(\frac{1}{S}\sum_{s=1}^{S}(\hat{\boldsymbol{\beta}}_{qls}^{(p)} - \boldsymbol{\beta}_{ql}^{(p)})^2\right)^{1/2}, \tag{5.27}$$

$$\text{RMSE}(\hat{\sigma}_{ql}^2) = \left(\frac{1}{S}\sum_{s=1}^{S}(\hat{\sigma}_{qls}^2 - \sigma_{ql}^2)^2\right)^{1/2}, \tag{5.28}$$

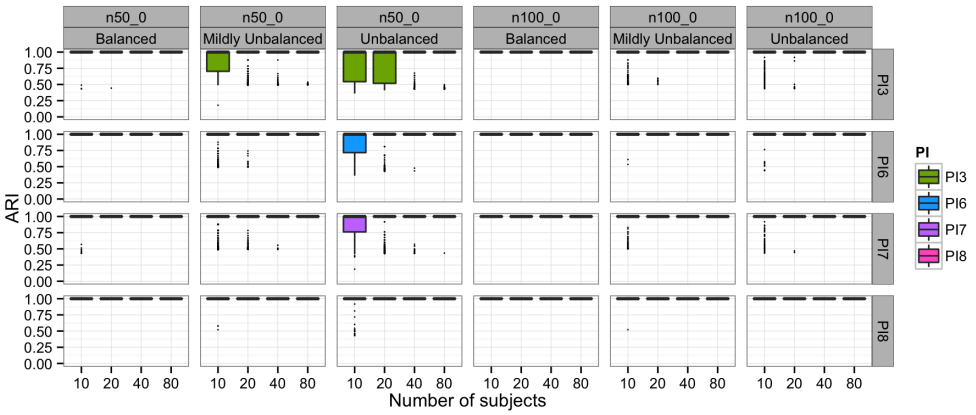
$$\text{Bias}(\hat{\boldsymbol{\beta}}_{ql}^{(p)}) = \frac{1}{S}\sum_{s=1}^{S}(\hat{\boldsymbol{\beta}}_{qls}^{(p)} - \boldsymbol{\beta}_{ql}^{(p)}), \tag{5.29}$$

$$\text{Bias}(\hat{\sigma}_{ql}^2) = \frac{1}{S}\sum_{s=1}^{S}(\hat{\sigma}_{qls}^2 - \sigma_{ql}^2), \tag{5.30}$$

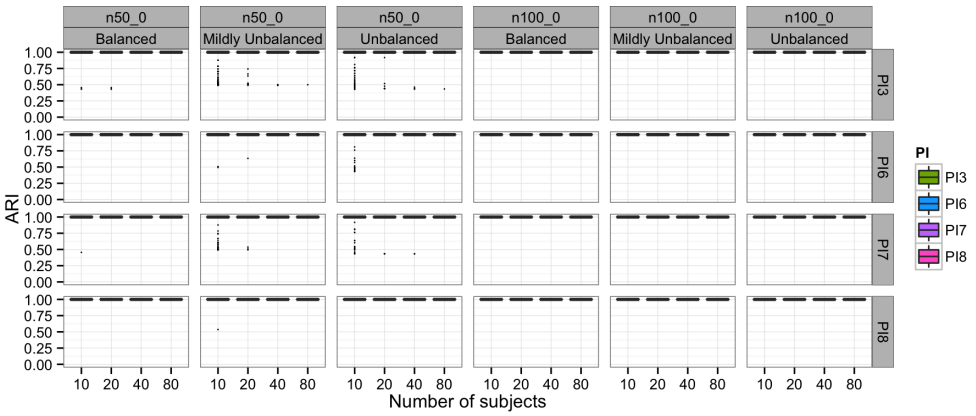
where $\boldsymbol{\beta}_{ql}^{(p)}$ is the p-th element of vector $\boldsymbol{\beta}_{ql}$.

5.3 Simulation Results

In Figure 5.2 (a) and (b), we show the box plots of ARI scores for the two scenarios of repeated measurements ($T \in \{1, 3\}$) and variance $\boldsymbol{\sigma}^2 = \mathbf{1}$ over $S = 1000$ network realisations. We note that the Het-Mixed-SBM struggles to correctly estimate the cluster structure for PI3 and that this is particularly pronounced in the cases with Mildly Unbalanced and Unbalanced designs, networks with 50 nodes and samples with 10 & 20 subjects (see Figure 5.2 (a)). However, as shown in Figure 5.2 (b), with a larger number of visits ($T = 3$), the accuracy of cluster estimates is improved. Similarly, for PI6 and PI7, the Unbalanced proportion design, networks with 50 nodes and 10 subjects, the Het-Mixed-SBM struggles to correctly estimate the cluster structure. However, the model improves the accuracy of the cluster estimates as the number of subjects increases and it tends to perform even better in the samples with more visits (Figure 5.2 (b)). In Figure 5.3, we show the box plots



(a) **One visit and $\boldsymbol{\sigma}^2 = \mathbf{1}$**



(b) **Three visits and $\boldsymbol{\sigma}^2 = \mathbf{1}$**

Figure 5.2: **ARI scores over $S = 1000$ network realisations with increasing number of subjects along the x-axis**. The Het-Mixed-SBM fits are evaluated with respect to (i) varying proportion designs (Balanced, Mildly Unbalanced and Unbalanced, on each column), (ii) varying network sizes ($n \in \{50, 100\}$) with no age effect (on each of the first three columns), and (iii) varying connectivity structures (PI3 & PI6-8, on each row and plotting colour).

of ARI scores for different variance settings ($\boldsymbol{\sigma}^2 \in \{0.5, 1, 2\}$), Unbalanced proportion designs, one visit ($T = 1$) and varying subjects $K \in \{10, 20, 40\}$. In the cases with a small number of nodes and subjects, there seems to be some evidence that the estimates of cluster structure are less accurate when the variance is larger (e.g., in the cases of PI7 & PI8 and their respective samples with 10 and 20 subjects). However, in other examples (e.g., PI3, $n = 50$ and 20 subjects), this influence is less apparent.

Next, in Figure 5.4 (a) and (b), we show the RMSE scores for the estimates of $\hat{\boldsymbol{\alpha}}$ for the cases of one and three visits and variance $\boldsymbol{\sigma}^2 = \mathbf{1}$. For both visit

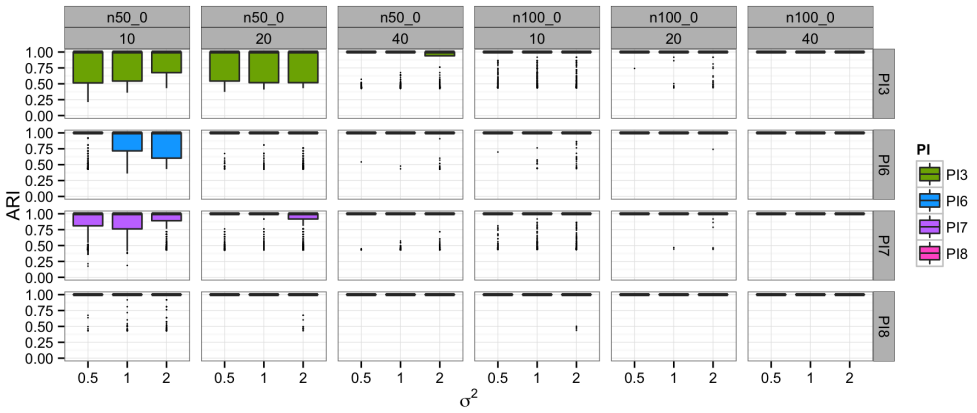
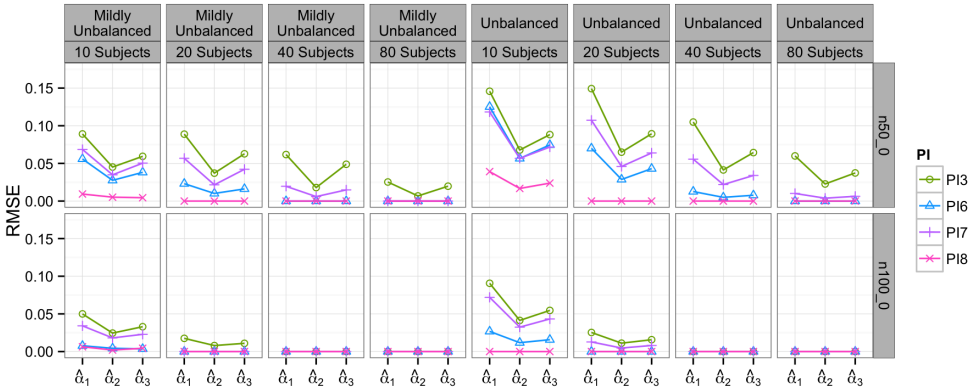


Figure 5.3: **ARI scores over $S = 1000$ network realisations with $\boldsymbol{\sigma}^2$ along the x-axis**. The Het-Mixed-SBM fits are evaluated with respect to the Unbalanced proportion design, (i) varying network sizes ($n \in \{50, 100\}$, on each of the first three columns) with no age effect, (ii) varying numbers of subjects ($K \in \{10, 20, 40\}$, on each of the first three columns), and (iii) varying connectivity structures (PI3 & PI6-8, on each row and plotting colour).

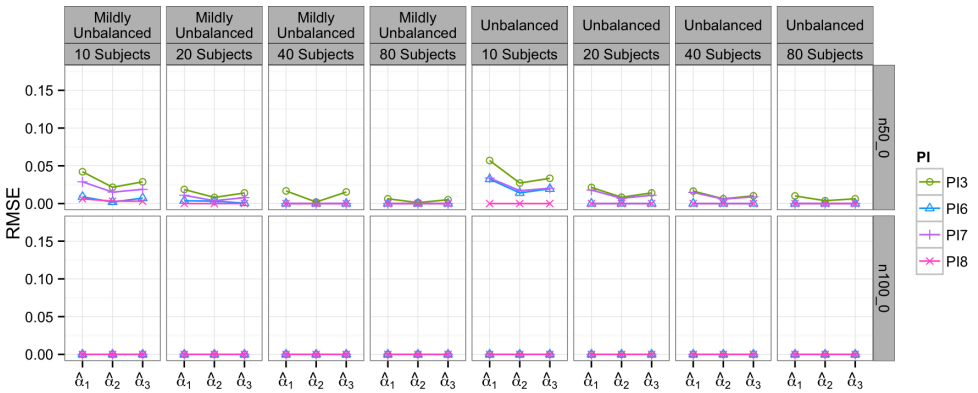
counts, the RMSE scores seems to decrease with an increasing number of subjects and an increasing number of nodes. This behaviour is consistent with the behaviour reported in Section 4.4.1 (see Figure 4.4) and seems to be also linked with the overall accuracy of the estimated cluster structure. For example, the cluster structure estimates exhibit some degree of variability over 1000 realisations for PI3 and this is captured in the RMSE scores that tend to be the largest. In Figure 5.5 (a) and (b), we show the RMSE of the intercept estimates ($\hat{\boldsymbol{\beta}}$) in the cases of one and three visits, and $\boldsymbol{\sigma}^2 = \mathbf{1}$. The RMSE is generally smaller for PI8 than for the other connectivity structures. This can be explained by the presence of a bias which occurs when the Het-Mixed-SBM struggles to correctly estimate the cluster structure. As PI8 was the least affected by this (see Figure 5.3), it is not surprising that its RMSEs are the lowest. It is also interesting to note that the RMSE is typically decreasing with an increasing number of subjects. This can be simply explained by the decrease of variance of $\hat{\boldsymbol{\beta}}$ with an increasing number of subjects. Note however that the variance does not markedly change with either number of visits or number of nodes.

In Figure 5.6, we show the bias of $\hat{\boldsymbol{\sigma}}^2$ in the cases of one and three visits, and $\boldsymbol{\sigma}^2 = \mathbf{1}$. We clearly see that the estimates tend to have an appreciable negative bias in small samples, which decreases with an increasing number of subjects.

In Figure 5.7, we show the FPR obtained using a Wald test on both the intercepts and slopes, and in the cases of one and three visits, and $\boldsymbol{\sigma}^2 = \mathbf{1}$. In small samples, the Wald test is liberal, but becomes more accurate when the number of subjects increases. Our simulations seem to suggest that a sample with 80 subjects



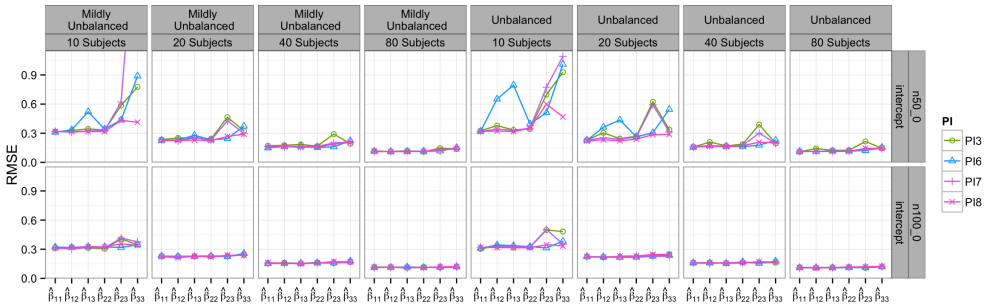
(a) **One visit and $\sigma^2 = 1$**



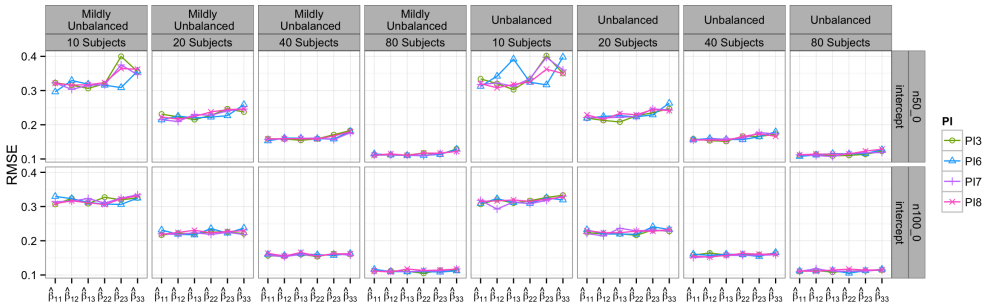
(b) **Three visits and $\sigma^2 = 1$**

Figure 5.4: **RMSE of $\hat{\alpha}$ whose individual block elements are given along the x-axis**. The RMSE scores are evaluated with respect to (i) varying proportion designs (Mildly Unbalanced and Unbalanced, on each of the first four columns), (ii) varying numbers of subjects ($K \in \{10, 20, 40, 80\}$, on each of the first four columns), (iii) varying network sizes ($n \in \{50, 100\}$) with no age effect (on each row), and (iv) varying connectivity structures (PI3 & PI6-8, plotting symbols and colour).

is sufficient to allow a relatively accurate control of the FPR.

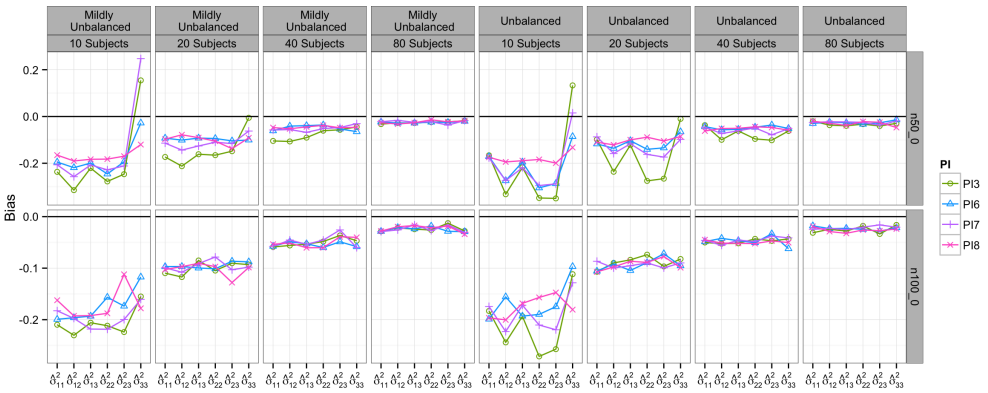


(a) **One visit and $\sigma^2 = 1$**

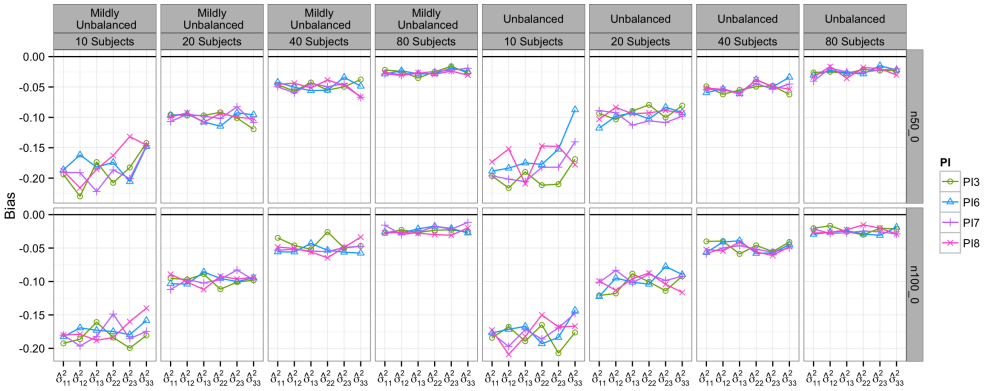


(b) **Three visits and $\sigma^2 = 1$**

Figure 5.5: **RMSE of $\hat{\beta}$-intercepts whose individual block elements are given along the x-axis**. The RMSE scores are evaluated with respect to (i) varying proportion designs (Mildly Unbalanced and Unbalanced, on each column), (ii) varying numbers of subjects ($K \in \{10, 20, 40, 80\}$, on each of the first four columns), (iii) varying network sizes ($n \in \{50, 100\}$) with no age effect (on each row), and (iv) varying connectivity structures (PI3 & PI6-8, plotting symbols and colour). Note that, for clarity, the RMSE of $\hat{\beta}_{33}$ in the first column and row of (i) is not shown and its value is 2.74.

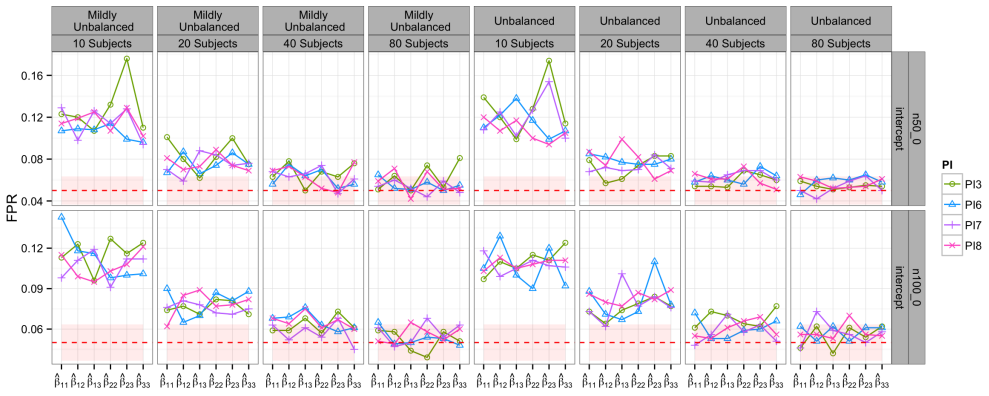


(a) **One visit and $\sigma^2 = 1$**

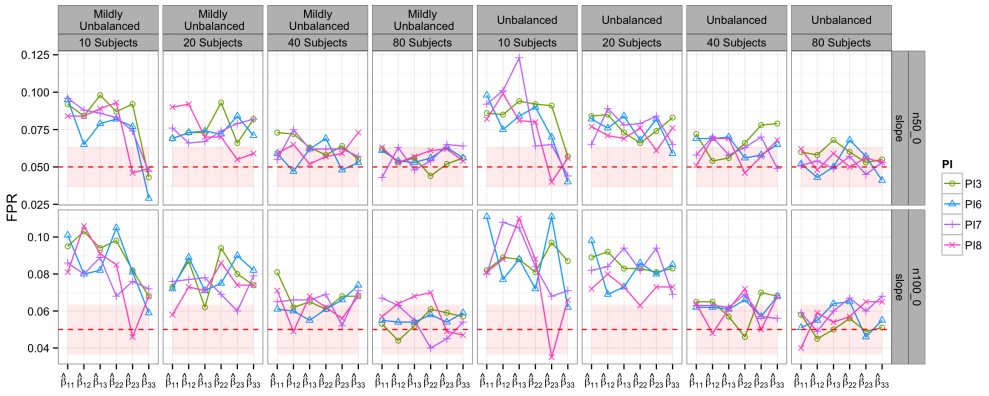


(b) **Three visits and $\sigma^2 = 1$**

Figure 5.6: **Bias of $\hat{\sigma}^2$ whose individual block elements are given along the x-axis**. The bias scores are evaluated with respect to (i) varying proportion designs (Mildly Unbalanced and Unbalanced, on each column), (ii) varying numbers of subjects ($K \in \{10, 20, 40, 80\}$, on each of the first four columns), (iii) varying network sizes ($n \in \{50, 100\}$) with no age effect (on each row), and (iv) varying connectivity structures (PI3 & PI6-8, plotting symbols and colour).



(a) **Intercept for three visits and $\sigma^2 = 1$**



(b) **Slope for three visits and $\sigma^2 = 1$**

Figure 5.7: **False Positive Rates (FPR) at 5% significance level for the Wald test on each element of $\hat{\beta}$, whose individual block elements are given along the x-axis**. The FPR scores are evaluated with respect to (i) varying proportion designs (Mildly Unbalanced and Unbalanced, on each column), (ii) varying numbers of subjects ($K \in \{10, 20, 40, 80\}$, on each of the first four columns), (iii) varying network sizes ($n \in \{50, 100\}$) with no age effect (on each row), and (iv) varying connectivity structures (PI3 & PI6-8, plotting symbols and colour).

5.4 Het-Mixed-SBM Fit to Multi-subject Functional Connectivity fMRI Data

We fit the Het-Mixed-SBM to the same data considered in the previous chapter, Section 4.5. Specifically, we compare resting state fMRI connectivity in 13 healthy controls (Controls) to 12 patients with schizophrenia (Patients). The Het-Mixed-SBM estimates a multi-subject network decomposition with 22 blocks. In Figure 5.8, we show the spatial location of each node and its corresponding block assignment. Similarly, to Section 4.6, the blocks are further organised into panels (Panel A-F) and matched against the Intrinsic Connectivity Networks (ICNs) assignments (see Section 4.5.1). The individual panels in Figure 5.8 indicate that the estimated cluster structure accounted reasonably well for the spatial information in the brain. To work out the approximate functional labels of the individual blocks, we contrast

Panel A	Panel B	Panel C
Block 12 – ICN 1	Block 4 – ICN 6	Block 5 – ICN 11
Block 1 – ICN 2	Block 9 – ICN 6	Block 14 – ICN 12
Block 2 – ICN 2	Block 11 – ICN 7	Block 16 – ICN 12
Block 13 – ICN 2	Block 6 – ICN 10	
Block 15 – ICN 5		
Block 19 – ICN 5		

Panel D	Panel E	Panel F
Block 3 – ICN 13	Block 18 – ICN 14	Block 8 – ICN 17
Block 21 – ICN 13	Block 10 – ICN 15	Block 17 – ICN 18
	Block 7 – ICN 16	Block 20 – ICN 20
	Block 22 – ICN 16	

Table 5.1: **Correspondence between the Intrinsic Connectivity Networks (ICNs) published in Laird et al. (2011) and the 22 blocks estimated by the Het-Mixed-SBM**.

their nodes against the Intrinsic Connectivity Networks maps reported in Laird et al. (2011) (see Table 5.1). Below, we describe the functional characteristics of each block.

Panel A: The nodes in Block 12, 1, 2 and 13 are predominately associated with the networks ICN 1 & 2 that are linked to emotional and autonomic processes. About 67 % of the nodes in Block 12 are linked to the visual recognition of emotions like fear, anger, happiness or humour, while all the nodes in Block 1 & 2, and 42% of the nodes in Block 13 are associated with the networks implicated in olfaction and gustation. About 57% of the nodes in Block 15

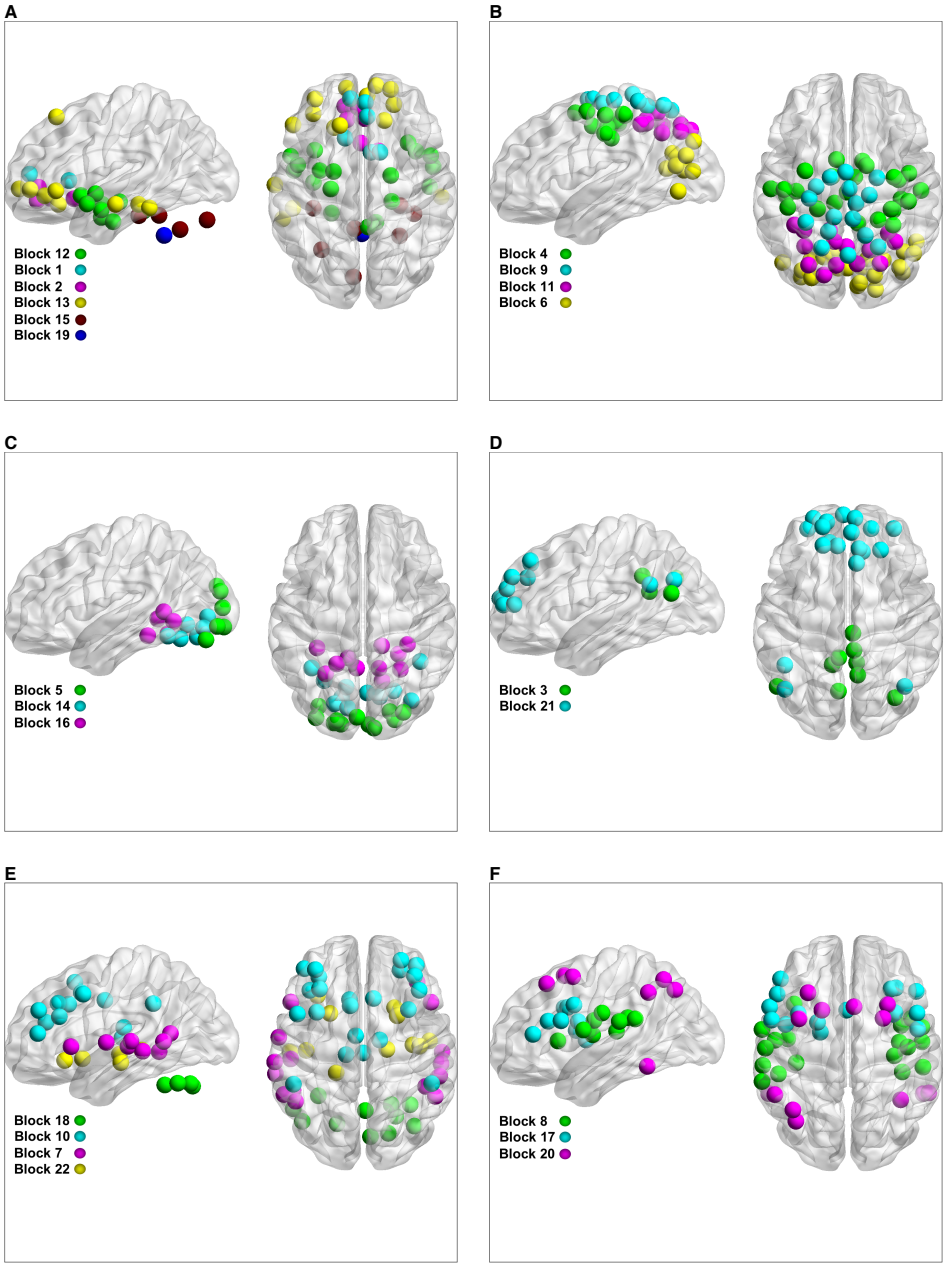


Figure 5.8: **Anatomical locations of individual nodes in the blocks, shown in terms of panels (Panel A-F)**.The plots were generated using the Matlab's BrainNet toolbox (Xia et al., 2013) (`http://www.nitrc.org`, last accessed on the 12th of September 2015).

are located in the midbrain and are linked to autonomic processes that reflect responses to internal stimuli (e.g., headache and stomach aches). Finally, Block 19 has only one node (Right Cerebellar Tonsil) and it is linked to autonomic processes related to breathing. Thus, we will broadly label this panel

as 'autonomic and visual' processes.

Panel B: About 44 % of the nodes in Block 4 and 31 % of the nodes in Block 9 are associated with ICN 6 which is located in the superior frontal gyrus and the middle frontal gyrus. These areas are linked to cognitive processes that include working memory, reasoning and problem solving. About 86% of the nodes in Block 11 are linked to ICN 7 which is located in the middle frontal gyrus and superior parietal lobules. This network is linked to cognitive processes that include visuospatial processing and reasoning. Finally, about 34% of the nodes in Block 6 are associated with ICN 10 which is located in the middle and inferior temporal gyri. This network is activated by viewing complex and emotional stimuli, as well as visual tracking of moving objects. As most of the areas in this panel are involved in functions related to the motor and visuospatial integration, coordination, and execution, we will broadly label this panel as 'motor and visual' processing.

Panel C: About 82 % of the nodes in Block 5 are linked to ICN 11 while 72% of nodes in Block 14 and 36% of nodes in Block 16 are linked to ICN 12. These two networks are located in the lateral and medial posterior occipital cortices and are linked to higher visual processes concerned with recognition of objects (ICN 11) as well as responses to simple visual stimuli (ICN 12). Thus, we will broadly label this panel as 'higher visual processes'.

Panel D: About 78% of the nodes in Block 3 and 37% of the nodes in Block 21 are linked to ICN 13 which represent the 'Default Mode Network'. In particular, the Default Mode Network is associated with the theory of mind which allows one to assign thought, desires and attentions to others and to explain or predict their actions.

Panel E: The majority of the nodes (91%) in Block 18 are linked to ICN 14 (located in the cerebellum) that is associated with a distributed range of sensorimotor, autonomic and cognitive functions. Approximatively 36% of the nodes in Block 10 are linked to ICN 15 which is associated with several cognitive processes, including reasoning, attention, inhibition, and memory. Next, about 63% of the nodes in Block 7 and 36% of the nodes in Block 22 correspond to ICN 16 which is mostly linked to functions related to audition and speech. Thus, we will broadly label this panel as 'audition and speech'.

Panel F: Finally, 67% of the nodes in Block 8 correspond to ICN 17 which includes primary sensorimotor cortices for mouth. It is associated with actions corresponding to speech and chewing/swallowing. A relatively small proportion of the nodes in Block 18 (38%) and in Block 20 (38%) are linked to ICN 18 &

20. In particular, ICN 18 is linked to language tasks while ICN 20 is linked to artefacts. Thus, we will label this panel as 'speech and language'.

It is also interesting to note that both of the cluster structures estimated by the Het-SBM (see Section 4.6) and the Het-Mixed-SBM show a very high level of mutual similarity (ARI score of 0.81). As indicated by Figure 5.9, relatively large portions of the cluster structure are simultaneously captured by both fits. In particular, both fits estimated the same mono-node (or a single node) block which contains the Right Cerebellar Tonsil.

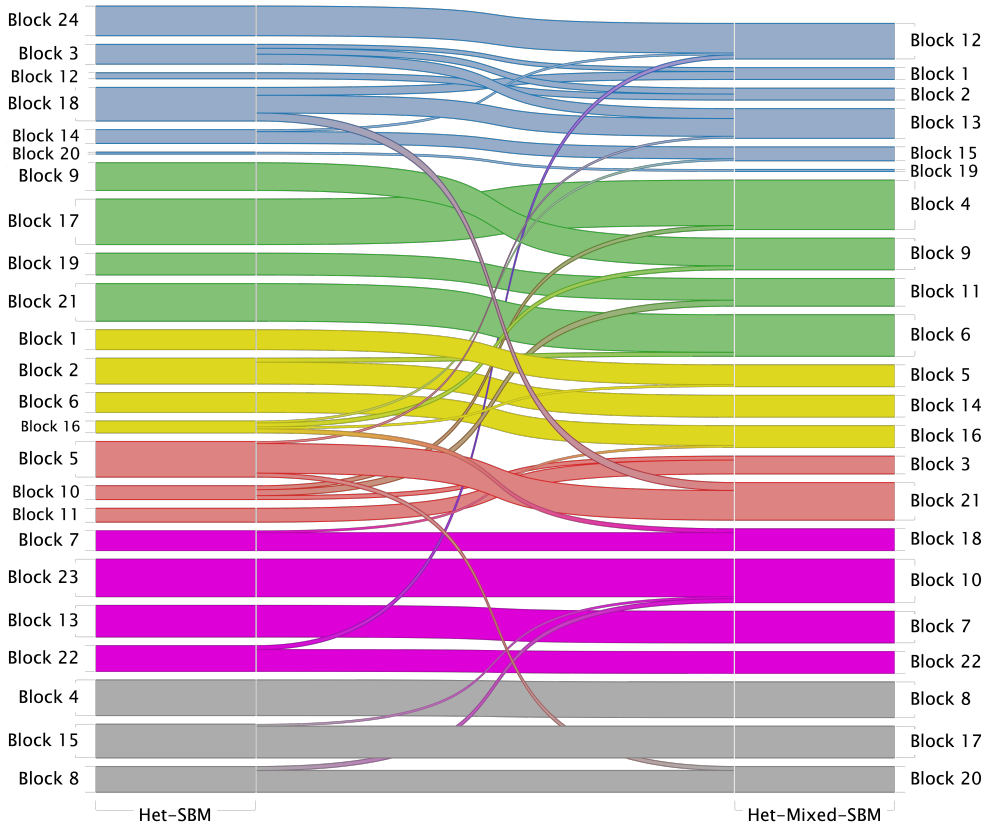
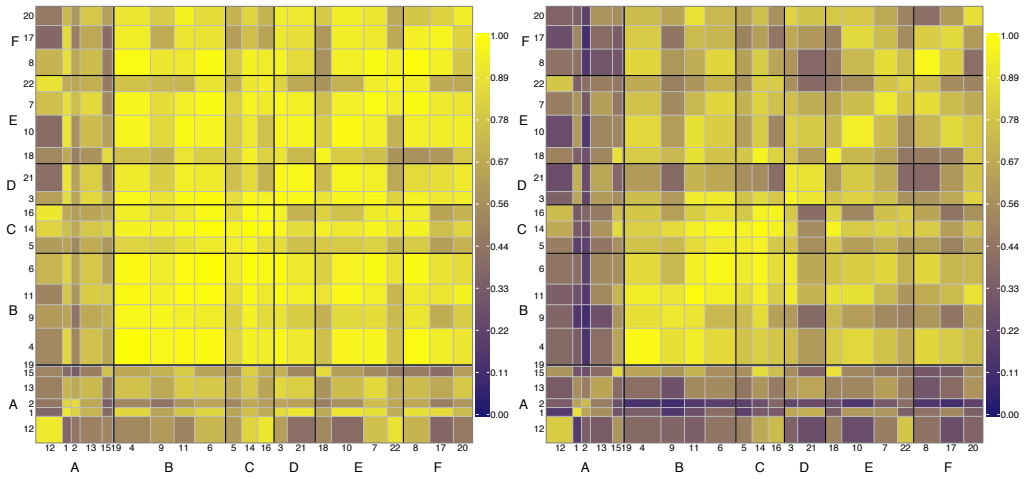


Figure 5.9: **Correspondence between the fits of the Het-SBM and Het-Mixed-SBM**. The strands of the alluvial diagram show each block of the Het-SBM fit (on the left) and how these merge and split to form the blocks of the Het-Mixed-SBM fit (on the right). The strands of the alluvial diagram are colour coded in terms of their respective panels (Panel A-F). Large portions of the cluster structure are similar between the two fits and both fits estimated a mono-node block (Block 20 in the Het-SBM & Block 19 in the Het-Mixed-SBM).

5.4.1 Inference

In Figure 5.10 (a) and (b), we show the estimated intercept values for the Control and Patient groups. The values are expressed as probabilities and denote the within-block or the between-block connection rates and the blocks are further organised in terms of panels A-F. A visual inspection of Figure 5.10 seems to suggest that the cluster structure in the Patient group (b) exhibits much weaker connection rates than the Control group (a). However, to gauge the extent of their differences, we make some inferences on differences between the intercepts of the Patients vs. Controls.



(a) Controls (b) Patients

Figure 5.10: **Estimated within/between block connectivity rates for the Het-Mixed-SBM fit**. Individual block labels are shown on $(x - y)$ axis and each block cell is demarcated by a grey line. The blocks are further aggregated into 6 panels, labelled A-F, which are visually separated by black lines. Largely similar patterns of connectivity are found in each population, though patients appear to have generally lower connection densities.

We note that our simulation indicates a danger of inflated false positive risk for a fixed effect for fewer than 60 subjects. Thus, with only 25 subjects in this analysis, we offer only tentative interpretation of the p-value computed. In Figure 5.11, we show the Wald test score image thresholded using a two-sided False Discovery Rate (FDR) method at 5% significance level. Overall, we can observe that the connection rates are significantly decreased in Patients compared to Controls and this is mainly captured in the relationships between the blocks rather than within-blocks. For example, a decreased connectivity is noted between Blocks 1 & 2 (emotional and autonomic processes) and all blocks in Panel B (motor and visual processing), the majority of the blocks in Panel C (higher visual processes),

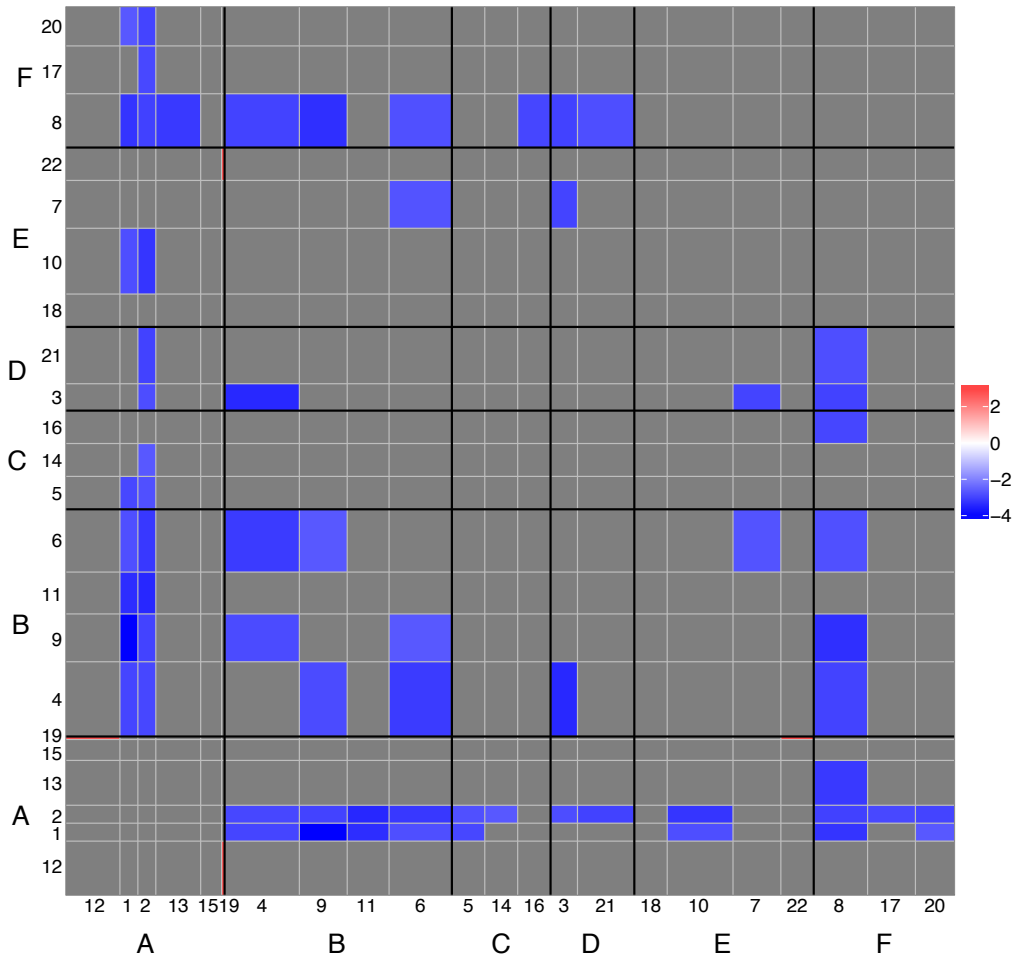


Figure 5.11: **Two-sided FDR thresholded (5%) Wald score image of the intercept differences between Patients vs. Controls, interpretable as the Patient-Control difference in (logit-scale) connectivity rates at mean age, IQ and motion (33.32 years, 113.06 NART and 7.40 mm)**. An increasing effect is shown in red and a decreasing effect is shown in blue.

all blocks in Panel D (Default Mode Network), some blocks in Panel E (audition and speech) and almost all blocks in Panel F (speech and language). Similarly, a decreased connectivity is observed in the blocks comprising Panel B, as well as in their links to the Default Mode Network (Block 3 & 4), Block 7 in Panel E (audition and speech), and Block 8 in Panel F (speech and language). It is also interesting to note that the nodes in the Default Mode Network do not show significant differences between the two populations, however, a significant decrease of connectivity is noted between the Default Mode Network and the nodes in Block 8 (actions corresponding to speech). However, some significant results also suggest an increased connectivity

in Patients relative to Controls, for example, mono-node Block 19 (Right Cerebellar Tonsil) and its connections to Block 12 (emotional and autonomic processes) and Block 22 (audition and speech).

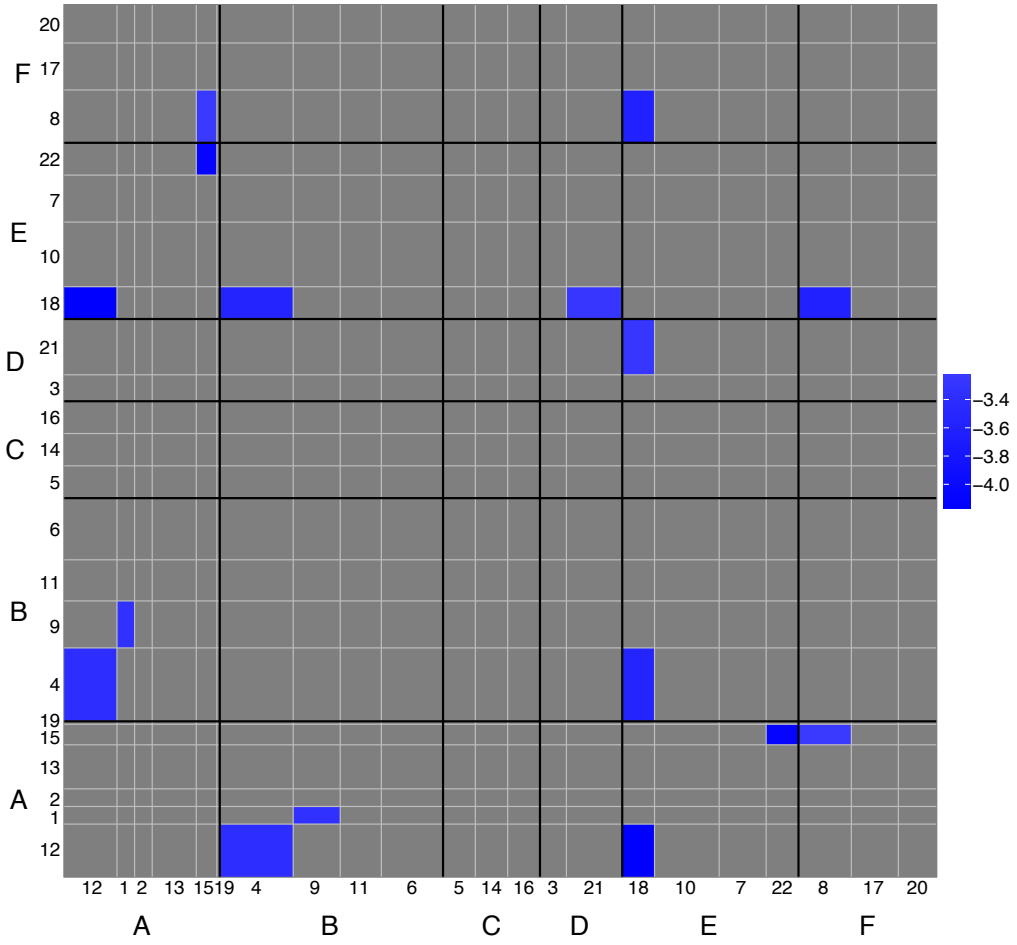


Figure 5.12: **Two-sided FDR thresholded (5%) Wald score image of the common age effect**. A decreasing effect is shown in blue.

In contrast to the Het-SBM inference results (see Figure 4.16) where almost the entire cluster structure tested significantly different in Patients vs. Controls, in the case of the Het-Mixed-SBM a much smaller number of significant results is present. It is also interesting to note that no results were significant after a Bonferroni correction to control the FWE. This is either attributable to small-sample conservativeness, a decreased detection sensitivity with between-subject variance now explicitly modelled, or both.

In Figure 5.12, we show the two-sided FDR thresholded (5%) Wald score image of the common age effect. Overall significant results suggest a decrease of

connectivity with an increase of age. This is especially evident for connections from Block 12 to Block 4 & 18, Block 1 to Block 9, Block 15 to Block 22 & 8. In particular, connections from Block 18 (sensorimotor, autonomic and cognitive functions) to Block 4 (working memory, reasoning and problem solving), 21 (Default Mode Network) and 8 (speech and language) show a decrease of connectivity with age.

5.5 Discussion

In this chapter, we proposed a novel multi-subject model, termed the Het-Mixed-SBM, which can be used for the analysis of multi-subject binary network data, allowing for random subject effects. As noted previously in the introductory section, this type of data may arise as a result of repeated networks per subject or simply dependences between the edges within-subjects not otherwise accounted for by covariates. In the real data analysis, the Het-Mixed-SBM estimated a very similar cluster structure as the Het-SBM (see Section 4.6). However, the differences between the two models are mainly reflected in their inference results, where the Het-Mixed-SBM detected a much smaller number of significant results than the Het-SBM. Although our simulation results showed that the Wald test for the Het-Mixed-SBM tend be liberal in small samples, the results of our analysis seem to suggest that much of the detected effect in the Het-SBM real data inference may be attributed to the dependencies in the data which were not accounted for by the model. Thus, it seems that the Het-Mixed-SBM may be a more appropriate model for this type of data. Nevertheless, more work is needed to investigate the behaviour of other tests, including the likelihood ratio and permutation tests.

One limitation of this work is that our simulations did not investigate the quality of the Het-Mixed-SBM parameter estimates in very challenging cases, for example, (i) when connection rates approach 0 or 1, and (ii) when a covariate has a heterogeneous impact onto a cluster structure, such as different effect sizes or directions per each block. In the context of example (i), it would be also interesting to compare the estimates between the Ordinary and Firth MLEs. However, it is important to note that the exact form of Firth adjustment used in the Het-SBM would not be appropriate for the Het-Mixed-SBM because it would have introduced a severe shrinkage to the variance estimate $\hat{\boldsymbol{\sigma}}^2$ (Firth, 1993). This shrinkage would likely have persevered even if the Fisher Information matrix was simplified by setting the fixed-random off-diagonal blocks to zero. As pointed out by Firth (1993), the appropriate regularisation term for generalised linear mixed models is based on the expected Fisher Information matrix, but this would necessitate an even more computationally involved procedures than the current Het-Mixed-SBM implementation. To our knowledge there is very little in the literature on the logistic mixed effects models and Firth type estimates. Recently, Claassen (2014) considered the practical aspects and the advantages of using Firth adjusted estimates. Therein, using a very simple logistic mixed effects model, comprised of one intercept and one random effect, she demonstrated that the Firth adjusted variance components are substantially less biased than those estimated by the Ordinary MLE approach and that this leads to more accurate inferences. However, more analysis is needed to formally assess how these findings would scale to our applications, where we have

a more complex logistic mixed effects model than the model considered in Claassen (2014), as well as already computationally involved estimation procedures. Indeed, the current implementation of the Het-Mixed-SBM has particularly slow computational times due to the estimation of $\boldsymbol{\tau}$ and its numerous evaluations of the integrals I_{iqlk} (see Eq. (5.10)) whose total number scales with the number of nodes in the network.

Finally, as part of our future work, we will also consider a useful simplification of our model that may speed up computation while accounting for subject-level randomness. Specifically, a model that combines 'homogeneous random effects' (per-subject random intercepts common to the entire network) with 'heterogeneous fixed effects' regression for each block element. We expect this model to have similar clustering potential as the Het-Mixed-SBM with much reduced computational burden.

CHAPTER 6

Conclusions and Future Work

This work has been motivated by the immense interest neuroscientists have shown in modelling brain networks. However, much of the methods for this work has consisted of network summary statistics that are not rooted in statistical models. To address these issues and to facilitate a methodological framework in which neuroimaging researchers can carry out multi-subject network analyses, this thesis proposed various extensions of a SBM that may be suitable for such neuroscientific applications. One particularly important aspect of the proposed models is that they allow for inference procedures on the block level parameters which may be particularly useful in studies that aim to detect topological abnormalities between the brains in a state of health and pathology. In the reminder of this chapter, we first give a brief summary of this thesis and then discuss relevant points which place this work in a wider context.

In Chapter 3, we used the neuronal network of *C. elegans* to compare cluster estimates based on the SBM, and the Fast Louvain and Spectral algorithms. We used the *C. elegans* network data for this task since its nodes (neurons) and edges have additional biological information which could be used to validate resulting cluster estimates. Our results showed that the 'cores-in-modules' decomposition of the SBM (comprising 9 blocks) was more compatible with prior biological information about the *C. elegans* than the modular decompositions found by the Fast Louvain and Spectral algorithms (comprising 5 and 4 modules respectively). These results also seem to suggest that a more plausible topological organisation of the *C. elegans* is that of a strongly connected core and a weakly connected periphery than a purely modular topological organisation. The SBM fit also provided a reasonable approximation of the network's observed degree distribution but it underestimated the network's empirical clustering coefficient.

In Chapter 4, we focused on the extensions of the SBMs which can be used for joint multi-subject network analysis. In particular, the two non-trivial multi-subject SBMs, the Hom-SBM and Het-SBM, capitalise on a logistic regression model which allows for an interaction between subject specific covariates and the block structure. The Hom-SBM assumes a 'homogeneous' regression model, where covariates exert a common influence over all block structure elements, while the Het-SBM is 'heterogeneous', having a regression for each block element. Parameter estimation for these models, can be based on the Ordinary and Firth MLEs and, in most scenarios, there are no striking differences between the two approaches. However, in the situations where the block connectivity rates approach 0 or 1, the Firth MLEs are preferable to the Ordinary MLEs as they tend to be the least biased. In terms of the inference, we investigated a series of issues including the accuracy of the Wald, likelihood ratio and permutation tests. Our analysis showed that the Wald test tended to be conservative for the Ordinary MLEs and liberal for the Firth MLEs, and that the likelihood ratio test is more preferable for the Firth MLEs, than the Wald test. In the real data analysis, the Het-SBM fit estimated a cluster structure with 24 blocks and found that the patients diagnosed with schizophrenia showed overall weaker rates of connections than the healthy controls. While our results showed a decreased connectivity in the Default Mode Network for the Patient group, a stronger difference seems to be noted in the way Default Mode Network connects to the nodes linked to motor & visual processes, and speech & language. In particular, these connections rates seems to be much weaker in the Patient than in the Control group. However, more analysis is needed to rigorously investigate this on other datasets.

In Chapter 5, we considered extensions of the SBMs that explicitly account for intra-subject correlation. When there is only one network per subject, this model can be useful if there is an excess of subject-level variation (i.e. not fully accounted for by covariates). When there are multiple networks per subject, e.g., as part of a 'repeated measures' design, this model is essential. The proposed Het-Mixed-SBM combines the classical SBM with a logistic mixed effect model with 'heterogeneous' block effects and per subject random intercepts. In the Monte Carlo simulations, the Het-Mixed-SBM showed a relatively satisfactory estimation of the cluster structure and its performance was very similar to the Het-SBM. However, the Het-Mixed-SBM showed poor small sample behaviour and, thus, we require generally larger datasets to attain accurate inferences.

Nevertheless, the Het-Mixed-SBM may bring new insights into the network organisations as it can estimate a common cluster structure while controlling for within-subject sources of variations. In the real data analysis, we illustrated this model on data with one network per subject. However, it would be also interesting

to apply this model to datasets with a larger number of subjects, where in addition each subject has several measurements. For example, the connectivity data shared by the Human Connectome Project (HCP) is particularly appealing as it offers high quality multi-subject data involving up to 1200 subjects and repeated measurements which pertain to different imaging modalities, including fMRI and DTI scans. In addition, there is extensive information on each subject, including their age, gender, intelligence quotient (IQ) and various family history and environmental factors. We expect that the application of the Het-Mixed-SBM will offer some interesting insights into the networks' organisation across different connectivity types (structural and functional) and disease states. Also, as indicated by our simulation results in Chapter 5, for such a large sample of subjects, we could expect accurate inferences.

In connection to the Het-Mixed-SBM, we would also like to develop models with additional random effects that would be nested within subjects. Indeed, with only a random intercept per subject, the model assumes a unique level of dependence between all the nodes within each block and each subject, even if they pertain to different networks. However, when we have multiple networks per subject, it seems more realistic to assume that, within each subject, we have two different levels of dependence, one within a network at measurement t, and the other between the T multiple of networks. Thus, to account for this situation more realistically, we could include a random intercept per network in addition to the random intercept per subject. Note that, even if this model appears to be more flexible, it might not be rich enough to model more complicated levels of dependence that can occur, for example, in longitudinal studies. Indeed, in such studies, the dependence between two networks from the same subject may also depend on the elapsed time between their measurements. In such cases, it would be also useful to include a random effect of time per subject.

The class of SBMs and their various multi-subject extensions considered in this thesis are expected to be very useful tools in the analysis of brain networks as well as other types of applications that are broadly related to brain parcellations. For example, the study of Johansen-Berg et al. (2004) established ideas for parcellation based on structural connectivity, where the individual voxels are grouped into clusters with homogeneous connectivity profiles and similar ideas were explored in the domain of functional connectivity analysis. We acknowledge that our multi-subject SBMs are based on 'binary connectivity data' and, as such, may not be appropriate for researchers who prefer to work directly with correlations (Smith et al., 2009) or weighted network data. Nevertheless, given the general outline of this work, adapting the multi-subject SBMs for such kind of data would be an important direction to pursue.

All multi-subject SBMs considered in this thesis, use parameter estimation

procedures that are based on a variational approximation. However, although the variational approximation tends to show promising results in practice, it is still important to highlight some theoretical concerns which are yet to be resolved. First, we still do not have a full understanding of the asymptotic behaviour of the MLEs based on the marginal likelihood and, second, as the variational bound approximates the behaviour of such MLEs, it would be desirable to have a set of sufficient conditions which would guarantee similar properties between the variational and marginal MLEs. With regards to the first point, some interesting findings were presented in the article of Ambroise and Matias (2012) who showed that, under the assumption of a modular or a disassortative cluster structure (termed 'affiliation structure') which are also assumed to be fully specified by two parameters, the estimators based on a method of moments as well as the estimators based on the optimisation of three-variate composite likelihoods are consistent and $\sqrt{n}$-convergent. The motivation for the use of three-variate composite likelihoods comes from the fact that the mixture components are identifiable for three-variate Bernoulli variables (X_{ij}, X_{ik}, X_{jk}), but not for the univariate Bernoulli variables (X_{ij}). In particular, both of these approaches offer some interesting applications that may be particularly relevant in the analysis of networks with a large number of nodes. In the contexts of brain connectivity analysis, such datasets typically arise in the animal studies that focus on the connections between neuronal elements on a small scale and these networks tend to have about 6000 nodes. In such cases, the researchers so far would have very limited methodological options that would be mostly based on modular algorithms. Therefore, it would be interesting to see how these two approaches compare to the results from the modular algorithms and if these two approaches identify a modular structure as well.

The work presented in this thesis relied upon the assumption that each node in a network can only belong to one block, that is, that the estimated blocks are disjoint. However, such assumptions may reveal only the coarsest views of the network's functional specialisation because the nodes rarely facilitate only one function and instead they tend to be involved in several functions. Therefore, it seems to be more plausible to expect that an overlapping cluster structure might be more representative of the network's overall functions. Motivated by this, the article of Latouche et al. (2011) proposed an overlapping SBM in the context of a single network. The key assumption, which separates this model from the classical SBM, is that the nodes are allocated into Q blocks according to the Bernoulli distribution rather than the Categorical distribution. In particular, this ensures that some nodes are allocated into several blocks, but, also, that some nodes are excluded from the overall cluster structure. The block connection rates are modelled through a sequence of interactions terms which account for various types of cluster structures,

and the estimation procedure is based upon two levels of variational approximation. As a part of our future work, we will consider the ways in which an edge-based covariates can be incorporated in this model and used as a source of additional clustering information.

APPENDIX A

Background

A.1 Lower Bound of the Stochastic Blockmodel

We note that $\mathbb{E}(Z_{iq}Z_{jl}) = \tau_{iq}\tau_{jl}$ and $\mathbb{E}(Z_{iq}) = \tau_{iq}$. The lower bound $\mathcal{J}(f^*(\boldsymbol{z};\boldsymbol{\tau}))$ of the SBM can be explicitly derived as

$$
\begin{aligned}
\mathcal{J}(f^*(\boldsymbol{z};\boldsymbol{\tau})) &= \mathbb{E}\Big(\log[f(\boldsymbol{x}, \boldsymbol{Z};\boldsymbol{\pi},\boldsymbol{\alpha})]\Big) - \mathbb{E}\Big(\log[f^*(\boldsymbol{Z};\boldsymbol{\tau})]\Big) \\
&= \mathbb{E}\Big(\frac{1}{2}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\sum_{q=1}^{Q}\sum_{l=1}^{Q} Z_{iq}Z_{jl}\log\left[\pi_{ql}^{x_{ij}}(1-\pi_{ql})^{1-x_{ij}}\right] + \sum_{i=1}^{n}\sum_{q=1}^{Q} Z_{iq}\log\alpha_q\Big) \\
&\quad - \mathbb{E}\Big(\sum_{i=1}^{n}\sum_{q=1}^{Q} Z_{iq}\log\tau_{iq}\Big) \\
&= \frac{1}{2}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\sum_{q=1}^{Q}\sum_{l=1}^{Q} \mathbb{E}(Z_{iq}Z_{jl})\log\left[\pi_{ql}^{x_{ij}}(1-\pi_{ql})^{1-x_{ij}}\right] \\
&\quad + \sum_{i=1}^{n}\sum_{q=1}^{Q}\mathbb{E}(Z_{iq})\log\alpha_q - \sum_{i=1}^{n}\sum_{q=1}^{Q}\mathbb{E}(Z_{iq})\log\tau_{iq} \\
&= \frac{1}{2}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\sum_{q=1}^{Q}\sum_{l=1}^{Q} \tau_{iq}\tau_{jl}\log\left[\pi_{ql}^{x_{ij}}(1-\pi_{ql})^{1-x_{ij}}\right] \\
&\quad + \sum_{i=1}^{n}\sum_{q=1}^{Q}\tau_{iq}\log\alpha_q - \sum_{i=1}^{n}\sum_{q=1}^{Q}\tau_{iq}\log\tau_{iq}. \qquad \text{(A.1)}
\end{aligned}
$$

A.2 Factorisation of the Integrated Classification Likelihood

If the prior of the model parameters can be factorised, so that $p(\boldsymbol{\pi}, \boldsymbol{\alpha}|\mathrm{m}_Q) = p(\boldsymbol{\pi}|\mathrm{m}_Q)p(\boldsymbol{\alpha}|\mathrm{m}_Q)$, then integrated classification $f(\boldsymbol{x}, \boldsymbol{z}|\mathrm{m}_Q) = f(\boldsymbol{x}|\boldsymbol{z}, \mathrm{m}_Q)f(\boldsymbol{z}|\mathrm{m}_Q)$, as

$$\begin{aligned}
f(\boldsymbol{x}, \boldsymbol{z}|\mathrm{m}_Q) &= \int_{\boldsymbol{\Pi}, \boldsymbol{A}} f(\boldsymbol{x}, \boldsymbol{z}|\mathrm{m}_Q, \boldsymbol{\pi}, \boldsymbol{\alpha}) p(\boldsymbol{\pi}, \boldsymbol{\alpha}|\mathrm{m}_Q) d\boldsymbol{\pi} d\boldsymbol{\alpha} \\
&= \int_{\boldsymbol{\Pi}, \boldsymbol{A}} f(\boldsymbol{x}|\boldsymbol{z}, \mathrm{m}_Q, \boldsymbol{\pi}, \boldsymbol{\alpha}) f(\boldsymbol{z}|\mathrm{m}_Q, \boldsymbol{\pi}, \boldsymbol{\alpha}) p(\boldsymbol{\pi}|\mathrm{m}_Q) p(\boldsymbol{\alpha}|\mathrm{m}_Q) d\boldsymbol{\pi} d\boldsymbol{\alpha} \\
&= \int_{\boldsymbol{\Pi}} f(\boldsymbol{x}|\boldsymbol{z}, \mathrm{m}_Q, \boldsymbol{\pi}) p(\boldsymbol{\pi}|\mathrm{m}_Q) d\boldsymbol{\pi} \int_{\boldsymbol{A}} f(\boldsymbol{z}|\mathrm{m}_Q, \boldsymbol{\alpha}) p(\boldsymbol{\alpha}|\mathrm{m}_Q) d\boldsymbol{\alpha} \\
&= f(\boldsymbol{x}|\boldsymbol{z}, \mathrm{m}_Q) f(\boldsymbol{z}|\mathrm{m}_Q),
\end{aligned}$$

where $\boldsymbol{\Pi}, \boldsymbol{A}$ are parameter spaces for $\boldsymbol{\pi}$ and $\boldsymbol{\alpha}$ for a particular Q.

A.3 Likelihood of Mixing Proportions

Assuming that the prior of $\boldsymbol{\alpha}$ is Dirichlet $\mathrm{D}(\delta, \ldots, \delta)$, we consider

$$\begin{aligned}
\log f(\boldsymbol{z}|\mathrm{m}_Q) &= \log \int_{\boldsymbol{A}} f(\boldsymbol{z}|\mathrm{m}_Q, \boldsymbol{\alpha}) p(\boldsymbol{\alpha}|\mathrm{m}_Q) d\boldsymbol{\alpha}, \\
&= \log \int_{\boldsymbol{A}} \prod_q^Q \alpha_q^{n_q} \frac{\Gamma(\sum_q^Q \delta)}{\prod_q^Q \Gamma(\delta)} \prod_q^Q \alpha_q^{\delta - 1} d\boldsymbol{\alpha}, \\
&= \log \left[\frac{\Gamma(Q\delta)}{\Gamma(\delta)^Q} \right] + \log \int_{\boldsymbol{A}} \prod_q^Q \alpha_q^{(n_q + \delta) - 1} d\boldsymbol{\alpha}, \\
&= \log \left[\frac{\Gamma(Q\delta)}{Q\Gamma(\delta)} \right] + \log \left[\frac{\prod_q^Q \Gamma(n_q + \delta)}{\Gamma(\sum_q^Q [n_q + \delta])} \int_{\boldsymbol{A}} \prod_q^Q \frac{\Gamma(\sum_q^Q [n_q + \delta])}{\prod_q^Q \Gamma(n_q + \delta)} \alpha_q^{(n_q + \delta) - 1} d\boldsymbol{\alpha} \right], \\
&= \log [\frac{\Gamma(Q\delta)}{Q\Gamma(\delta)}] + \log \left[\frac{\prod_q^Q \Gamma(n_q + \delta)}{\Gamma(\sum_q^Q [n_q + \delta])} \right], \\
&\text{substituting for } \delta = \frac{1}{2}, \\
&= \log \Gamma(\frac{Q}{2}) - \log[Q\Gamma(\frac{1}{2})] + \sum_{q=1}^{Q} \log[\Gamma(n_q + \frac{1}{2})] - \log[\Gamma(n + \frac{Q}{2})].
\end{aligned} \tag{A.2}$$

A.4 Stirling's Approximation

Given large values of n_q, Stirling's formula approximates Gamma function,

$$\Gamma(n_q+1) \approx \sqrt{(2\pi)} n_q^{n_q+\frac{1}{2}} e^{-n_q}. \tag{A.3}$$

Starting from

$$\begin{aligned}\log f(\boldsymbol{z}|\mathrm{m}_Q) = &\log\Gamma\left(\frac{Q}{2}\right) - \log[Q\Gamma\left(\frac{1}{2}\right)] + \sum_{q=1}^{Q}\log[\Gamma\left(n_q+\frac{1}{2}\right)] \\ &- \log[\Gamma\left(n+\frac{Q}{2}\right)],\end{aligned} \tag{A.4}$$

we apply Eq. (A.3), and we drop order one terms (note that Q is assumed to be of smaller order than n) and we also make use of the fact that $\log\left[n_q-\frac{1}{2}\right]$ is asymptotically equivalent to $\log[n_q]$ and the same is true for $\log\left[n+\frac{Q}{2}-1\right]$ which is asymptotically equivalent to $\log[n]$,

$$\begin{aligned}\log f(\boldsymbol{z}|\mathrm{m}_Q) &\approx \sum_{q=1}^{Q}\left[\log[\sqrt{2\pi}] + n_q\log\left[n_q-\frac{1}{2}\right] - \left[n_q-\frac{1}{2}\right]\right] \\ &\quad - \left[\log[\sqrt{2\pi}] + \left(n+\frac{Q-1}{2}\right)\log\left[n+\frac{Q}{2}-1\right] - \left(n+\frac{Q}{2}-1\right)\right] \\ &\approx \sum_{q}^{Q} n_q\log[n_q] - n - [n\log(n) + \frac{Q-1}{2}\log[n] - n] \\ &\approx \sum_{q}^{Q} n_q\log[n_q] - n\log[n] - \frac{Q-1}{2}\log[n] \\ &\approx \sum_{q}^{Q} n_q\log[n_q] - \sum_{q}^{Q} n_q\log[n] - \frac{Q-1}{2}\log[n] \\ &\approx \sum_{q}^{Q} n_q\log\left[\frac{n_q}{n}\right] - \frac{Q-1}{2}\log[n] \\ &\approx \max_{\boldsymbol{\alpha}}\log f(\boldsymbol{z}|\mathrm{m}_Q) - \frac{Q-1}{2}\log[n].\end{aligned} \tag{A.5}$$

Notice that MLE of $\log f(\boldsymbol{z}|\mathrm{m}_Q)$ is $\hat{\alpha}_q = \frac{n_q}{n}$ as

$$\frac{\partial}{\partial\alpha_q}\left[\sum_{q}^{Q}\sum_{i}^{n} z_{iq}\log[\alpha_q]\right] = c\frac{\partial}{\partial\alpha_q}\left[\sum_{q}^{Q}\alpha_q\right]$$

$$
\begin{aligned}
&\frac{\partial}{\partial \alpha_q}\left[\sum_q^Q n_q \log[\alpha_q]\right] = c \\
&\frac{n_q}{\alpha_q} = c \\
&\sum_q^Q n_q = c\sum_q^Q \alpha_q \\
&n = c, \qquad \text{(A.6)}
\end{aligned}
$$

and, therefore, $\hat{\alpha}_q = \frac{n_q}{n}$.

A.5 Clustering Coefficient in the Stochastic Blockmodel

As noted in Section 2.2 clustering coefficient is defined as the proportion of the prevalence of fully connected triplets of nodes among the set of triplets that have at least two connections. Mathematically this can be noted as

$$C = \frac{3\#\triangle}{\#\vee} = \frac{\sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n x_{ij}x_{jk}x_{ik}}{2\sum_{i=1}^n \sum_{j=1}^n \sum_{k<j}^n x_{ij}x_{ik}},$$

and this can be simply expressed as the probability of seeing three nodes connected given that the two of these nodes are already connected,

$$
\begin{aligned}
\mathrm{P}(X_{ij}X_{jk}X_{ik} = 1 | X_{ij}X_{ik} = 1) &= \frac{\mathrm{P}(X_{ij}X_{jk}X_{ik} = 1)}{\mathrm{P}(X_{ij}X_{ik} = 1)}, \\
&= \frac{\sum_{q,l,s}^Q \mathrm{P}(X_{ij}X_{jk}X_{ik} = 1, Z_{iq} = 1, Z_{jl} = 1, Z_{ks} = 1)}{\sum_{q,l,s}^Q \mathrm{P}(X_{ij}X_{ik} = 1, Z_{iq} = 1, Z_{jl} = 1, Z_{ks} = 1)}, \\
&= \frac{\sum_{q,l,s}^Q \mathrm{P}(X_{ij} = 1, X_{jk} = 1, X_{ik} = 1, Z_{iq} = 1, Z_{jl} = 1, Z_{ks} = 1)}{\sum_{q,l,s}^Q \mathrm{P}(X_{ij} = 1, X_{ik} = 1, Z_{iq} = 1, Z_{jl} = 1, Z_{ks} = 1)}, \\
&= \frac{\sum_{q,l,s}^Q \hat{\alpha}_q\hat{\alpha}_l\hat{\alpha}_s\hat{\pi}_{ql}\hat{\pi}_{qs}\hat{\pi}_{ls}}{\sum_{q,l,s}^Q \hat{\alpha}_q\hat{\alpha}_l\hat{\alpha}_s\hat{\pi}_{ql}\hat{\pi}_{qs}} = \hat{C},
\end{aligned}
$$

A.6 Errata for the Stochastic Blockmodel of Zanghi, Volant and Ambroise

This model retains the complete structure of the SBM described in Section 2.4.1, but introduces the nodal features as a $n \times P$ matrix of random variables $\boldsymbol{Y}$ such that the features pertaining to a node V_i are distributed according to Multivariate Normal density,

$$\boldsymbol{Y}_i | Z_{iq} = 1 \sim \mathrm{MVN}(\boldsymbol{\mu}_q, \Sigma_q), \qquad \text{(A.7)}$$

where $\boldsymbol{\mu}_q = (\mu_{q1}, \ldots, \mu_{qP})$ is the vector of population means related to block (q, q) and the variance Σ_q, is $P \times P$ diagonal matrix related to block (q, q). The complete data likelihood of this model can be written as

$$\log f(\boldsymbol{x}, \boldsymbol{z}, \boldsymbol{y}; \boldsymbol{\pi}, \boldsymbol{\alpha}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \log f(\boldsymbol{x}|\boldsymbol{z}; \boldsymbol{\pi}) + \log f(\boldsymbol{y}|\boldsymbol{z}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) + \log f(\boldsymbol{z}; \boldsymbol{\alpha}). \quad \text{(A.8)}$$

It is worth mentioning that the details related to the variance component in the original work are difficult to interpret due to numerous typographical and notational inconsistencies. In that regard, we could think of two intended formulations for which we give corrected equations. First, $\Sigma_q = \sigma^2 \mathbb{I}$ in which case the updating equations are

$$\hat{\tau}_{iq} \propto \frac{\hat{\alpha}_q}{\hat{\sigma}^2} \prod_{j \neq i}^{n} \prod_{l=1}^{Q} \left[\hat{\pi}_{ql}^{x_{ij}} (1 - \hat{\pi}_{ql})^{1 - x_{ij}} \right]^{\hat{\tau}_{jl}} \prod_{p=1}^{P} \exp\left(-\frac{(\boldsymbol{y}_i^{(p)} - \hat{\boldsymbol{\mu}}_q^{(p)})^2}{2\hat{\sigma}^2}\right) \quad \text{(A.9)}$$

$$\hat{\boldsymbol{\mu}}_q = \frac{\sum_{i=1}^{n} \hat{\tau}_{iq} \boldsymbol{y}_i}{\sum_{i}^{n} \hat{\tau}_{iq}} \quad \text{(A.10)}$$

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n} \sum_{q=1}^{Q} \hat{\tau}_{iq} (\boldsymbol{y}_i - \hat{\boldsymbol{\mu}}_q)^\top (\boldsymbol{y}_i - \hat{\boldsymbol{\mu}}_q)}{P \sum_{i=1}^{n} \sum_{q=1}^{Q} \hat{\tau}_{iq}} \quad \text{(A.11)}$$

and the ICL criterion is given as

$$\begin{aligned} &ICL(\mathrm{m}_Q) = \log f(\boldsymbol{x}, \boldsymbol{y}, \hat{\boldsymbol{z}}|\mathrm{m}_Q, \hat{\boldsymbol{\pi}}, \hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}}) \\ &- \frac{1}{2} \frac{Q(Q+1)}{2} \log\left[\frac{n(n-1)}{2}\right] - \frac{Q-1}{2} \log[n] - \frac{1}{2}(QP + 1) \log[nP]. \end{aligned} \quad \text{(A.12)}$$

Second, $\Sigma_q = \sigma_q^2 \mathbb{I}$ and in this case $\hat{\boldsymbol{\mu}}_q$ is the same as in the first scenario and

$$\hat{\tau}_{iq} \propto \frac{\hat{\alpha}_q}{\hat{\sigma}_q^2} \prod_{j \neq i}^{n} \prod_{l=1}^{Q} \left[\hat{\pi}_{ql}^{x_{ij}} (1 - \hat{\pi}_{ql})^{1 - x_{ij}} \right]^{\hat{\tau}_{jl}} \prod_{p=1}^{P} \exp\left(-\frac{(\boldsymbol{y}_i^{(p)} - \hat{\boldsymbol{\mu}}_q^{(p)})^2}{2\hat{\sigma}_q^2}\right) \quad \text{(A.13)}$$

$$\hat{\sigma}_q^2 = \frac{\sum_{i=1}^{n} \hat{\tau}_{iq} (\boldsymbol{y}_i - \hat{\boldsymbol{\mu}}_q)^T (\boldsymbol{y}_i - \hat{\boldsymbol{\mu}}_q)}{P \sum_{i=1}^{n} \hat{\tau}_{iq}} \quad \text{(A.14)}$$

and the ICL criterion is given as

$$\begin{aligned} &ICL(\mathrm{m}_Q) = \log f(\boldsymbol{x}, \boldsymbol{y}, \hat{\boldsymbol{z}}|\mathrm{m}_Q, \hat{\boldsymbol{\pi}}, \hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}}) \\ &- \frac{1}{2} \frac{Q(Q+1)}{2} \log\left[\frac{n(n-1)}{2}\right] - \frac{Q-1}{2} \log[n] - \frac{1}{2}(Q(P+1)) \log[nP]. \end{aligned} \quad \text{(A.15)}$$

Note that, for both cases, the point estimating equations for $\boldsymbol{\alpha}$ and $\boldsymbol{\pi}$ are the same as in the SBM Eq. (2.23) and (2.24). It is also worth highlighting that, in the original work, the ICL criterion is inconsistent with the formulation of the model and it

seems to reward for the complexity of a model rather than penalise it. This is also corroborated by many inconsistencies, including the fact that there is nP number of data points in $\boldsymbol{y}$ rather than $n(n-1)/2$ as stated in the original publication.

APPENDIX B

Generalised Linear Stochastic Blockmodelling and Inference in Multi-subject Networks

B.1 Additional Results for the Simulation II

In this section, we provide full details of the simulation results based on PI9 connectivity matrix. In Figures B.1 and B.2, we report the bias of the intercept and slope. In most cases, the Ordinary and Firth MLEs tend to be similar, but in the cases with smaller block sizes and a smaller number of subjects the Ordinary MLE tends to be severely biased while the Firth MLE is almost unbiased. This is especially striking in the cases with 30 nodes, three subjects and all three proportion designs. The quality of the Ordinary MLEs seem to improve with increasing blocks sizes and number of subjects. Next, in Figures B.3 and B.4, we show the FPR rates of the intercepts for the Wald and likelihood ratio test based on the Ordinary and Firth MLEs. The Wald test based on the Ordinary MLEs is mostly conservative with the exception of the cases with 30 nodes, three subjects and Balanced proportion design (see also Figure B.5 & B.8 for its variance bias and power). In contrast, the Wald test based on the Firth MLEs is mostly liberal in small samples (e.g., networks with 30 and 50 nodes) but in other cases shows similar behaviour as the Ordinary MLEs. However, the likelihood ratio test based on the Ordinary MLEs (see Figure B.4) seems to be either liberal or conservative in small samples while the Firth MLEs tend to be liberal. These results seems to suggest that for the Firth MLEs the likelihood ratio test is preferable to the Wald test. The Ordinary and Firth MLEs are very similar for the slope estimates in the Wald test, but the Firth MLEs are conservative in the likelihood ratio test while the Ordinary MLEs are liberal (see Figures B.6 and B.7).

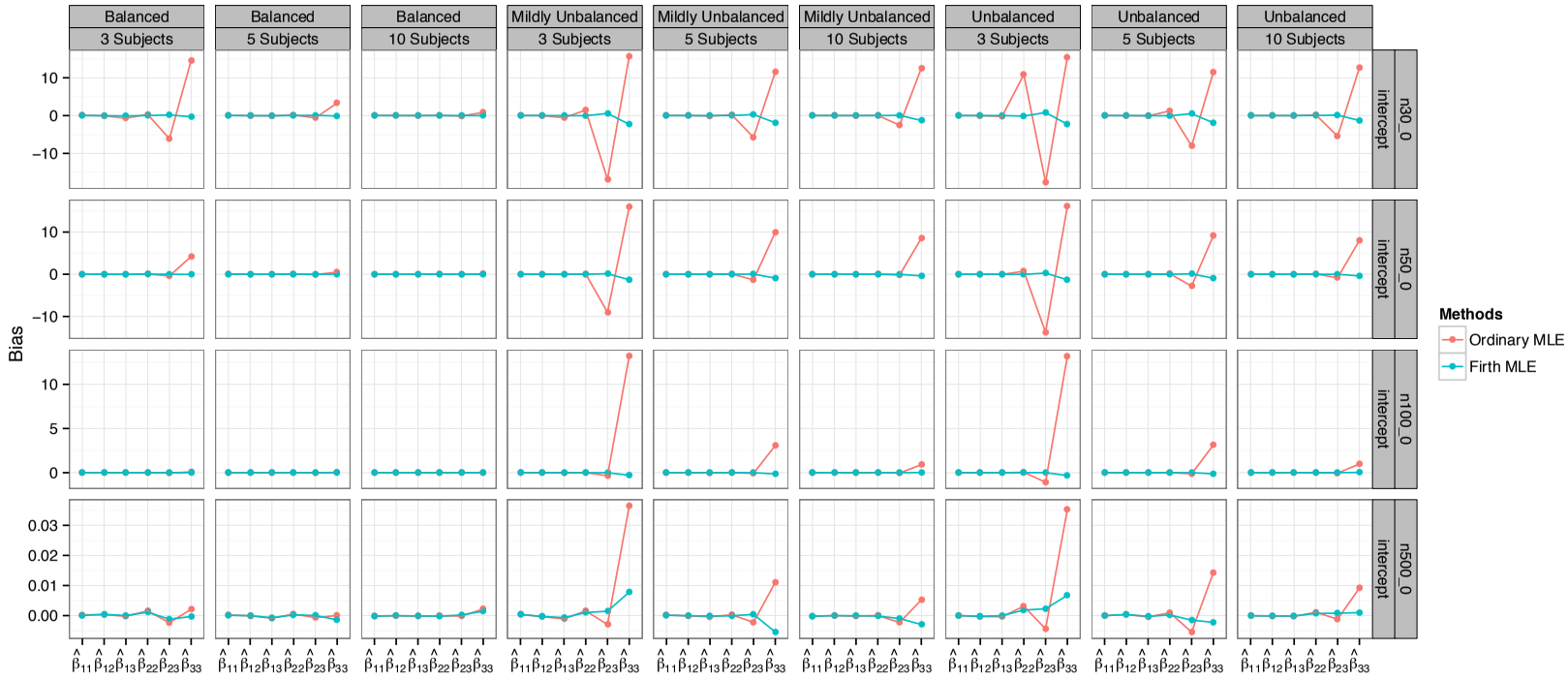


Figure B.1: **Bias of $\hat{\beta}_{ql}$-intercepts in the Het-SBM for the Ordinary and Firth ML estimates**. Between/within block intercept values are displayed along the x-axis and correspond to the data simulated from PI9 connectivity matrix. These estimates are evaluated with respect to varying (i) proportion designs (Balanced, Mildly Unbalanced and Unbalanced by columns), (ii) varying number of subjects (3, 5 and 10 subjects by columns), and (iii) network sizes ($n \in \{30, 50, 100, 500\}$ nodes by rows). In many scenarios, the Ordinary ML estimates present a large bias that is corrected by the the Firth ML estimates.

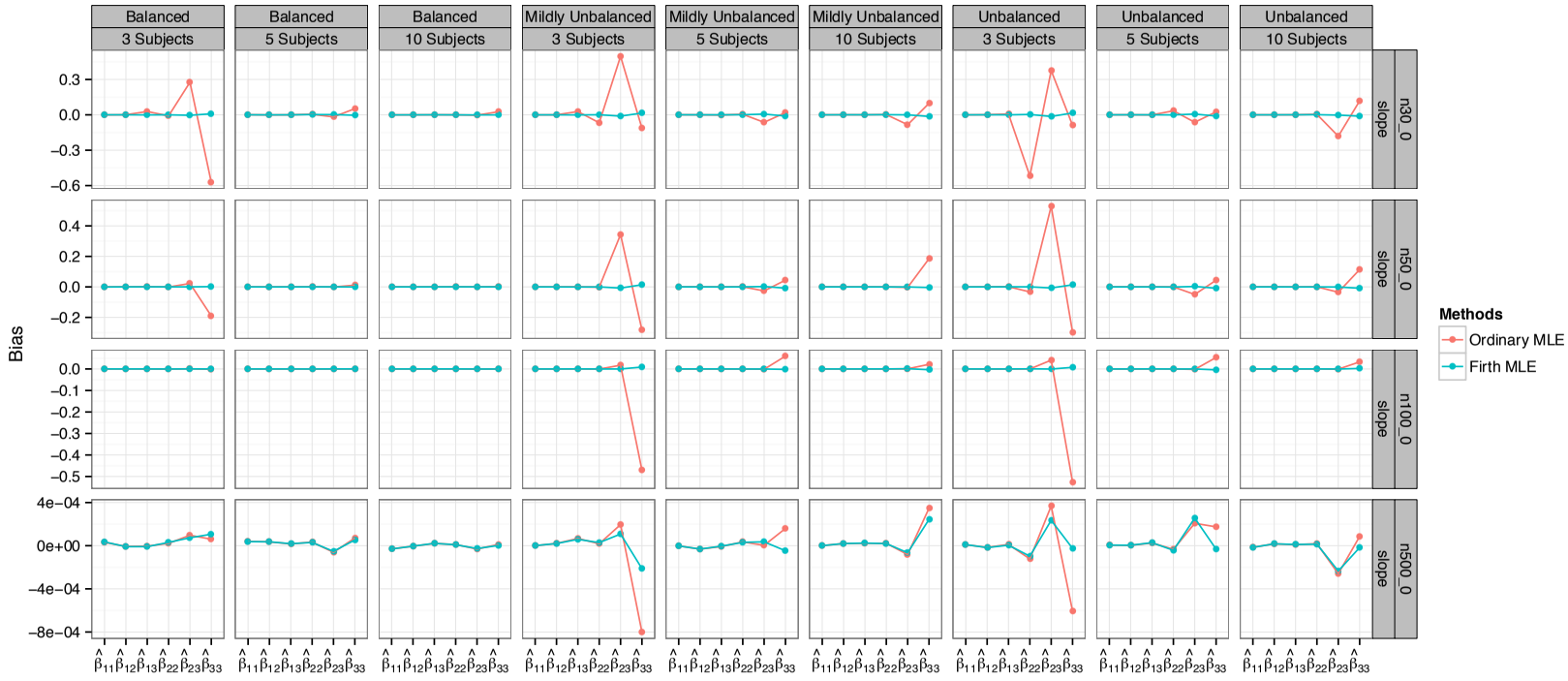


Figure B.2: **Bias of $\hat{\beta}_{ql}$-slopes in the Het-SBM for the Ordinary and Firth ML estimates**. Between/within block slope values are displayed along the x-axis and correspond to the data simulated from PI9 connectivity matrix. These estimates are evaluated with respect to varying (i) proportion designs (Balanced, Mildly Unbalanced and Unbalanced by columns), (ii) varying number of subjects (3, 5 and 10 subjects by columns), and (iii) network sizes ($n \in \{30, 50, 100, 500\}$ nodes by rows). In many scenarios, the Ordinary ML estimates present a large bias that is corrected by the the Firth ML estimates.

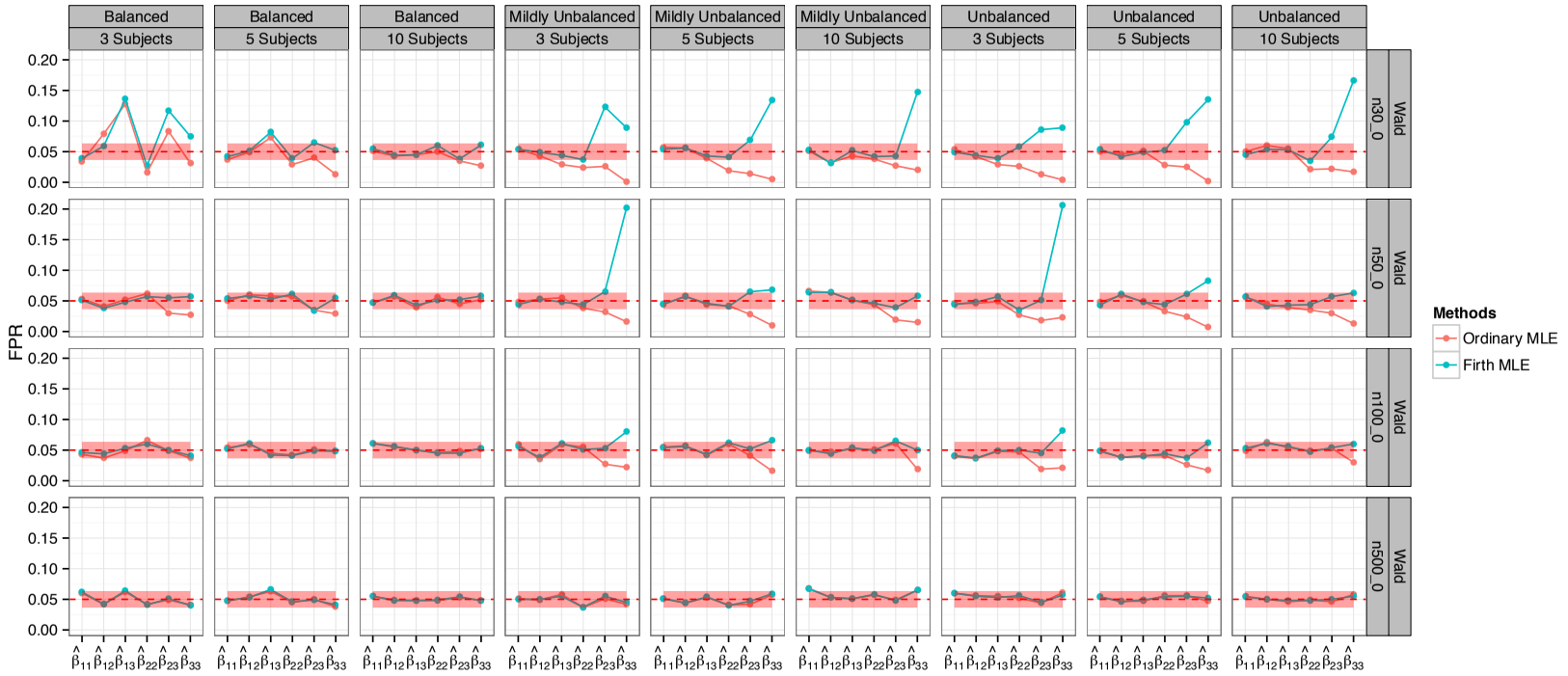


Figure B.3: **False Positive Rates (FPR) at 5% significance level of $\hat{\beta}_{ql}$-intercepts in the Het-SBM for the Ordinary and Firth ML estimates obtained with Wald tests**. Between/within block intercept values are displayed along the x-axis and correspond to the data simulated from PI9 connectivity matrix. The FPR are evaluated with respect to varying (i) proportion designs (Balanced, Mildly Unbalanced and Unbalanced by columns), (ii) varying number of subjects (3, 5 and 10 subjects by columns), and (iii) network sizes ($n \in \{30, 50, 100, 500\}$ nodes by rows). In many scenarios, the control of the FPR is accurate for both the Ordinary and Firth ML estimates. However, in scenarios with a small number of nodes and subjects, the control of the FPR tends to be conservative for the Ordinary estimates and liberal for the Firth estimates

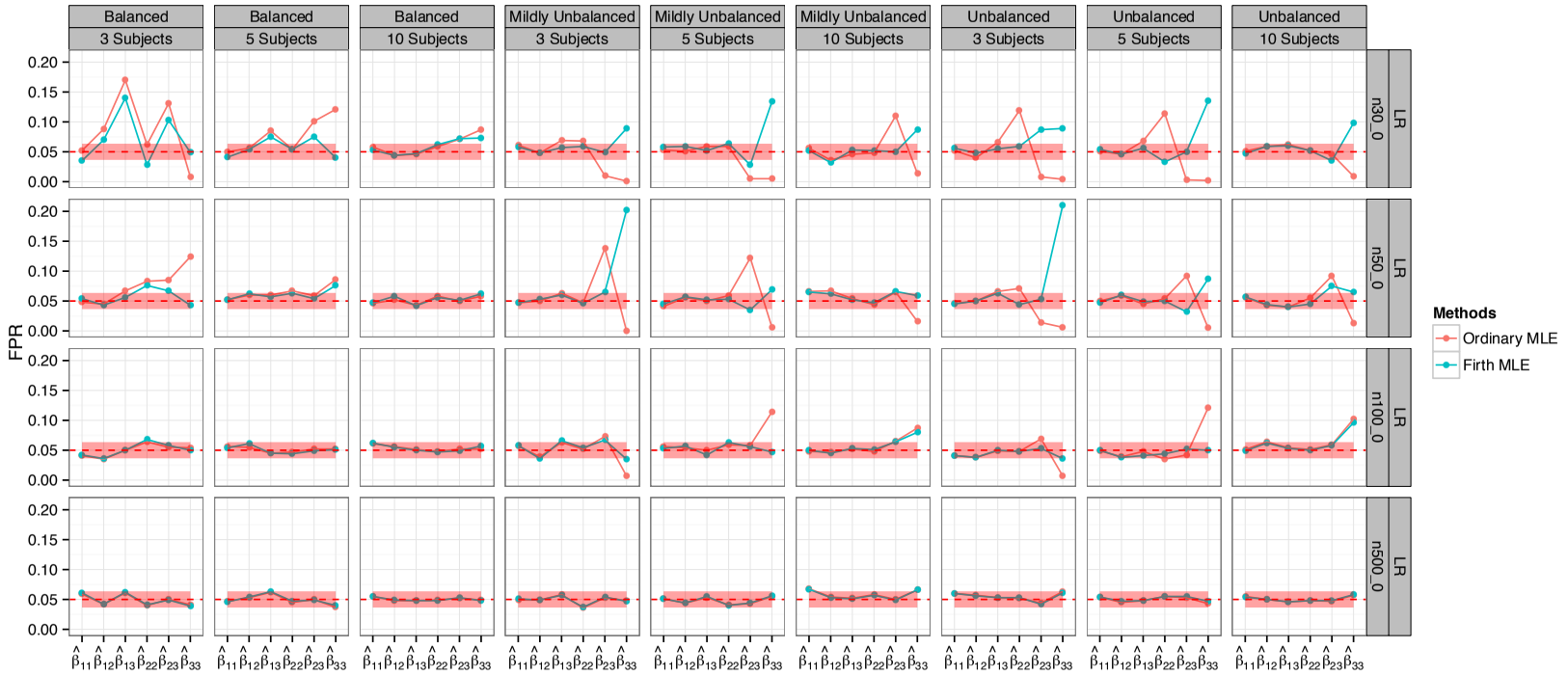


Figure B.4: **False Positive Rates (FPR) at 5% significance level of $\hat{\beta}_{ql}$-intercepts in the Het-SBM for the Ordinary and Firth ML estimates obtained with likelihood ratio (LR) test**. Between/within block intercept values are displayed along the x-axis and correspond to the data simulated from PI9 connectivity matrix. The FPR are evaluated with respect to varying (i) proportion designs (Balanced, Mildly Unbalanced and Unbalanced by columns), (ii) varying number of subjects (3, 5 and 10 subjects by columns), and (iii) network sizes ($n \in \{30, 50, 100, 500\}$ nodes by rows). In many scenarios, the control of the FPR is accurate for both the Ordinary and Firth ML estimates. However, in scenarios with a small number of nodes and subjects, the control of the FPR tends to be conservative or liberal for the Ordinary estimates and liberal for the Firth estimates

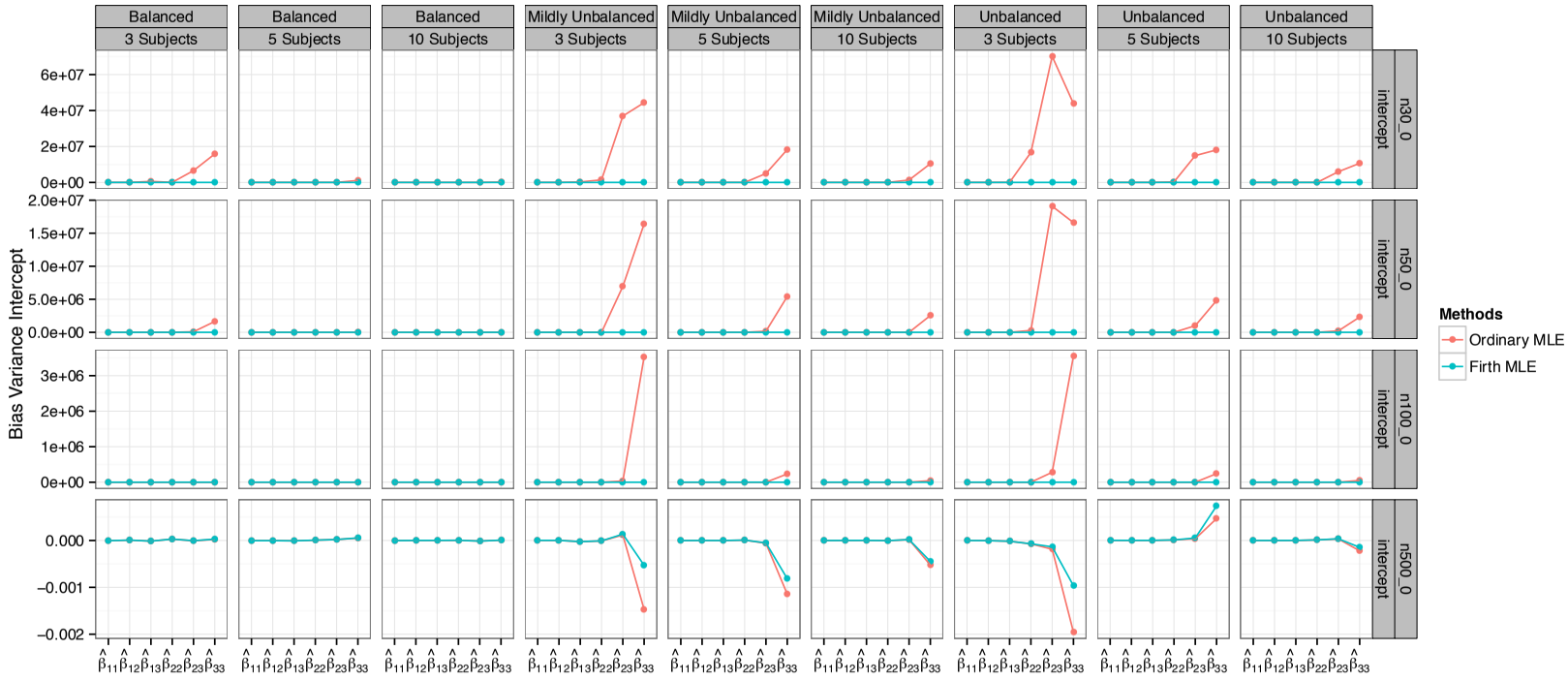


Figure B.5: **Bias of $\widehat{\text{var}}(\hat{\beta}_{ql}$-intercepts) in the Het-SBM for the Ordinary and Firth ML estimates**. Between/within block intercept values are displayed along the x-axis and correspond to the data simulated from PI9 connectivity matrix. These estimates are evaluated with respect to varying (i) proportion designs (Balanced, Mildly Unbalanced and Unbalanced by columns), (ii) varying number of subjects (3, 5 and 10 subjects by columns), and (iii) network sizes ($n \in \{30, 50, 100, 500\}$ nodes by rows). In many scenarios, the Ordinary ML estimates present a large bias that is corrected by the the Firth ML estimates.

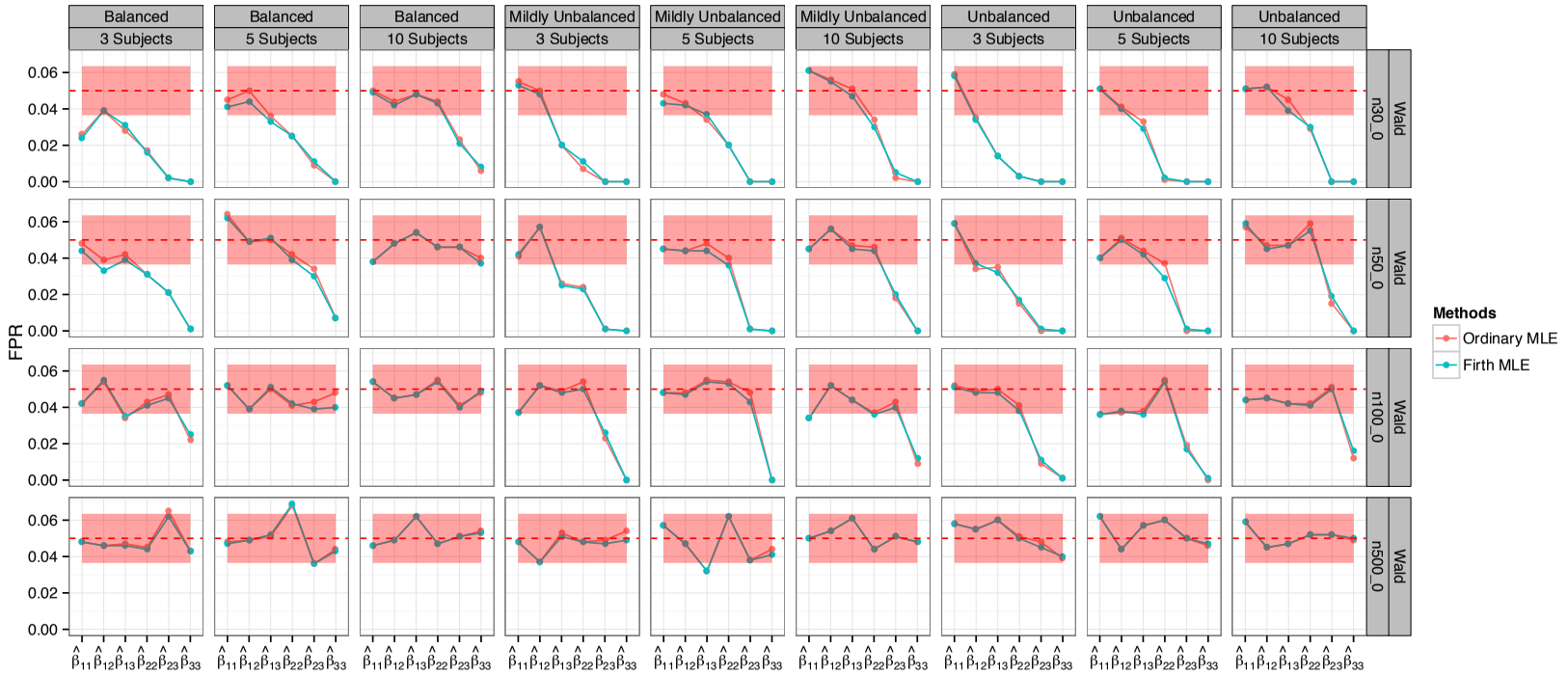


Figure B.6: **False Positive Rates (FPR) at 5% significance level of $\hat{\beta}_{ql}$-slopes in the Het-SBM for the Ordinary and Firth ML estimates obtained with Wald tests**. Between/within block slope values are displayed along the x-axis and correspond to the data simulated from PI9 connectivity matrix. The FPR are evaluated with respect to varying (i) proportion designs (Balanced, Mildly Unbalanced and Unbalanced by columns), (ii) varying number of subjects (3, 5 and 10 subjects by columns), and (iii) network sizes ($n \in \{30, 50, 100, 500\}$ nodes by rows). In many scenarios, the control of the FPR is accurate for both the Ordinary and Firth ML estimates. However, in scenarios with a small number of nodes and subjects, the control of the FPR tends to be conservative for the Ordinary and Firth estimates

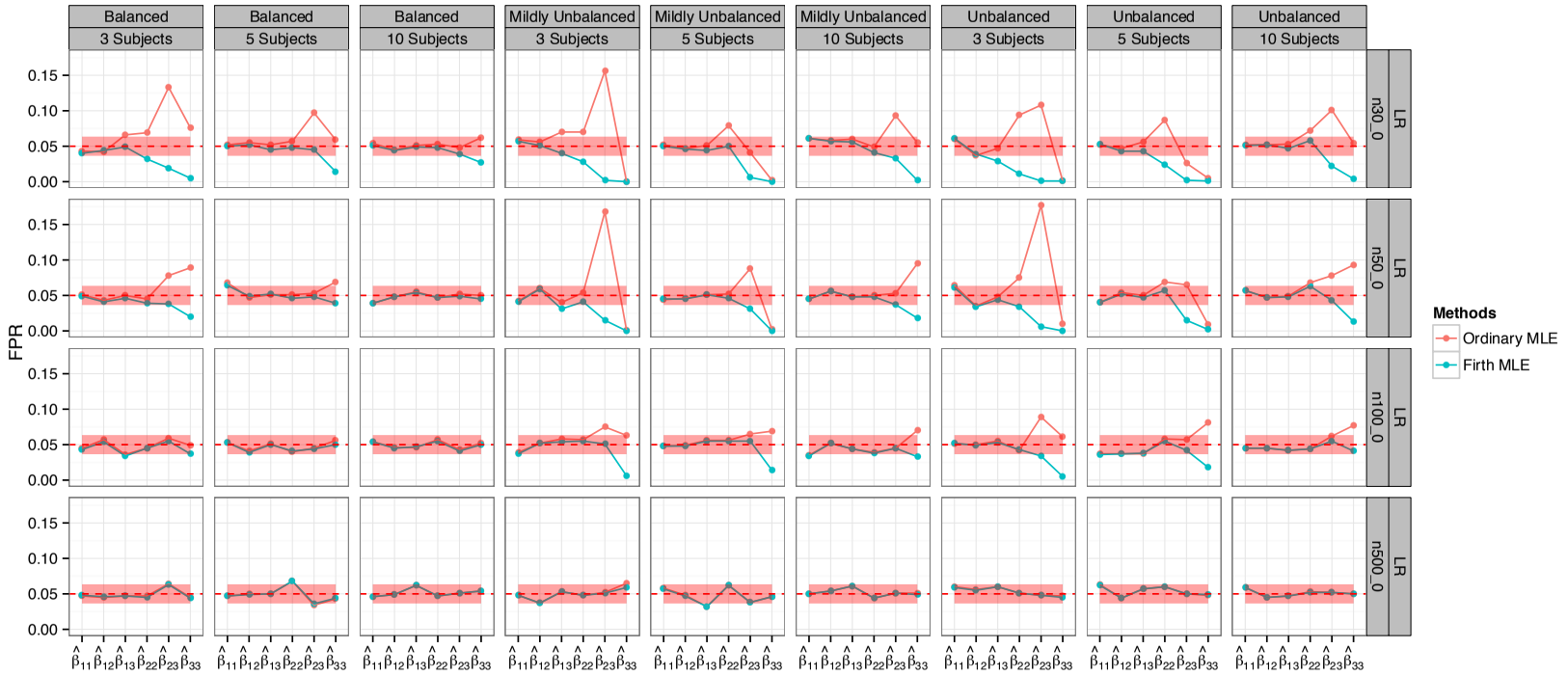


Figure B.7: **False Positive Rates (FPR) at 5% significance level of $\hat{\beta}_{ql}$-slopes in the Het-SBM for the Ordinary and Firth ML estimates obtained with the likelihood ratio tests**. Between/within block slope values are displayed along the x-axis and correspond to the data simulated from PI9 connectivity matrix. The FPR are evaluated with respect to varying (i) proportion designs (Balanced, Mildly Unbalanced and Unbalanced by columns), (ii) varying number of subjects (3, 5 and 10 subjects by columns), and (iii) network sizes ($n \in \{30, 50, 100, 500\}$ nodes by rows). In many scenarios, the control of the FPR is accurate for both the Ordinary and Firth ML estimates. However, in scenarios with a small number of nodes and subjects, the control of the FPR tends to be conservative or liberal for the Ordinary and conservative for the Firth estimates

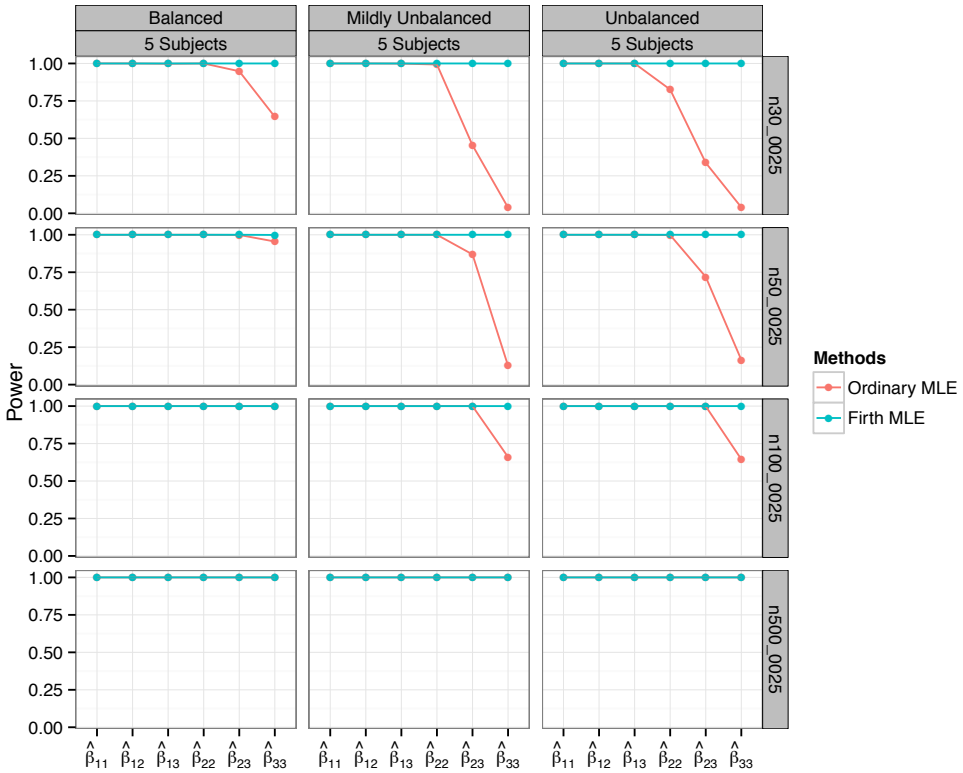


Figure B.8: **Power at 5% significance level of $\hat{\beta}_{ql}$-intercepts in the Het-SBM for the Ordinary and Firth ML estimates obtained with Wald tests**. Between/within block slope values are displayed along the x-axis and correspond to the data simulated from PI9 connectivity matrix. The power values are evaluated with respect to varying (i) proportion designs (Mildly Unbalanced by columns), (ii) number of subjects (by columns), and (iii) network sizes ($n \in \{30, 50, 100, 500\}$ nodes by rows). In the cases with the smallest samples, the power of the Wald test on the Ordinary estimates is significantly lower than for the Firth estimates.

APPENDIX C

Generalised Linear Mixed Stochastic Blockmodelling and Inference in Multi-subject Networks

C.1 Starting Values for Variance of the Random Effects

In this section, we describe a procedure for obtaining the initial values of the $\boldsymbol{\sigma}^2$. The goal is to first obtain the estimate of random effects or deviation of a subject specific intercept from the population intercept and then estimate their variances and then finally construct the initial value estimates of $\boldsymbol{\sigma}^2$. For the k-th subject and the (q,l)-th block, the goal is to maximise h_{qlk} with respect to r_{qlk}. Dropping the terms which depend on σ^2_{ql}, we have

$$h_{qlk}(r_{qlk}) = \frac{1}{2}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\sum_{t=1}^{T}\gamma_{ijql}x_{ijkt}r_{qlk} - \frac{1}{2}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\sum_{t=1}^{T}\gamma_{ijql}\log(1+e^{\boldsymbol{d}_{kt}^{\top}\boldsymbol{\beta}_{ql}+r_{qlk}}) \quad \text{(C.1)}$$

Taking the derivatives with respect to r_{qlk}, gives $U(r_{qlk})$

$$\begin{aligned} U(r_{qlk}) = &\frac{1}{2}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\sum_{t=1}^{T}\frac{\gamma_{ijql}}{1+e^{\boldsymbol{d}_{kt}^{\top}\boldsymbol{\beta}_{ql}+r_{qlk}}} - \frac{T}{2}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\gamma_{ijql} \\ &+ \frac{1}{2}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\sum_{t=1}^{T}\gamma_{ijql}x_{ijkt}, \end{aligned} \quad \text{(C.2)}$$

and setting $U(r_{qlk}) = 0$, we get

$$\frac{1}{2}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\sum_{t=1}^{T}\frac{\gamma_{ijql}}{1+e^{\boldsymbol{d}_{kt}^{\top}\boldsymbol{\beta}_{ql}+r_{qlk}}} = \frac{T}{2}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\gamma_{ijql} - \frac{1}{2}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\sum_{t=1}^{T}\gamma_{ijql}x_{ijkt} \quad \text{(C.3)}$$

Using substitution $\omega_{qlk} = e^{r_{qlk}}$, we obtain the

$$\frac{1}{2}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\sum_{t=1}^{T}\frac{\gamma_{ijql}}{1+\omega_{qlk}e^{\boldsymbol{d}_{kt}^{\top}\boldsymbol{\beta}_{ql}}} = \frac{T}{2}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\gamma_{ijql} - \frac{1}{2}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\sum_{t=1}^{T}\gamma_{ijql}x_{ijkt}, \tag{C.4}$$

and taking derivatives with respect to ω_{qlk} gives

$$-\frac{1}{2}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\sum_{t=1}^{T}\frac{\gamma_{ijql}e^{\boldsymbol{d}_{kt}^{\top}\boldsymbol{\beta}_{ql}}}{(1+\omega_{qlk}e^{\boldsymbol{d}_{kt}^{\top}\boldsymbol{\beta}_{ql}})^2}. \tag{C.5}$$

Estimate of ω_{qlk} are found via Newton-Raphson updating equations

$$\omega_{qlk}^{(m+1)} = \omega_{qlk}^{(m)} + \frac{\frac{1}{2}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\sum_{t=1}^{T}\gamma_{ijql}x_{ijkt} - \frac{T}{2}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\gamma_{ijql} + \frac{1}{2}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\sum_{t=1}^{T}\frac{\gamma_{ijql}}{1+\omega_{ql}^{(m)}e^{\boldsymbol{d}_{kt}^{\top}\boldsymbol{\beta}_{ql}}}}{\frac{1}{2}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\sum_{t=1}^{T}\frac{\gamma_{ijql}e^{\boldsymbol{d}_{kt}^{\top}\boldsymbol{\beta}_{ql}}}{(1+\omega_{qlk}^{(m)}e^{\boldsymbol{d}_{kt}^{\top}\boldsymbol{\beta}_{ql}})^2}}. \tag{C.6}$$

For the starting point ω_{qlk}^{0}, we take advantage of the relationship between arithmetic and harmonic mean, so that

$$\sum_{t=1}^{T}\frac{1}{1+e^{\boldsymbol{d}_{kt}^{\top}\boldsymbol{\beta}_{ql}}\omega_{qlk}^{0}} \geq \frac{T^2}{T+\omega_{qlk}^{0}\sum_{t=1}^{T}e^{\boldsymbol{d}_{kt}^{\top}\boldsymbol{\beta}_{ql}}}. \tag{C.7}$$

Substituting (C.7) in (C.4) we can estimate starting point ω_{qlk}^{0} as

$$\omega_{qlk}^{0} = \frac{\frac{T}{2}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\sum_{t=1}^{T}\gamma_{ijql}x_{ijkt}}{\sum_{t=1}^{T}e^{\boldsymbol{d}_{kt}^{\top}\boldsymbol{\beta_{ql}}}\Big(\frac{T}{2}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\gamma_{ijql} - \frac{1}{2}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\sum_{t=1}^{T}\gamma_{ijql}x_{ijkt}\Big)}. \tag{C.8}$$

For the solution of Newton-Raphson step $\omega_{qlk}^{(*)}$, we take the transformation $r_{qlk} = \log(\omega_{qlk}^{(*)})$. After obtaining estimate for every subject k, we get a sequence $\{u_{ql1}, \ldots, r_{qlk}\}$. Then the goal is to use this sequence to estimate σ_{ql}^2. For this we need $Var(r_{qlk}) = v_{qlk}$, which is just

$$v_{qlk} = \frac{1}{-\frac{\partial}{\partial r_{qlk}}U(r_{qlk})} = \frac{1}{\frac{1}{2}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\sum_{t=1}^{T}\gamma_{ijql}\frac{e^{\boldsymbol{d}_{kt}^{\top}\boldsymbol{\beta_{ql}}}}{1+e^{\boldsymbol{d}_{kt}^{\top}\boldsymbol{\beta_{ql}}}}}, \tag{C.9}$$

and finally the estimate of a starting point is

$$\hat{\sigma}_{ql}^2 = \frac{1}{K}\sum_{k=1}^{K}(r_{qlk}^2 - v_{qlk}). \tag{C.10}$$

C.2 Reparametrisation of the Integrals in the Het-Mixed-SBM

As mentioned in Chapter 5, when h_{qlk} assumes a small value (e.g., $h_{qlk} = -1000$) the integrals $I_{qlk1} - \boldsymbol{I}_{qlk6}$ vanish and, consequently, the likelihood diverges to $-\infty$. To treat this point of numerical instability, we add an offset value c_{qlk} to the exponent h_{qlk}. In particular, this offset value is chosen to be the maximum of h_{qlk}, taken with respect to r_{qlk}, while iterative values of ${\boldsymbol{\beta}_{ql}}^{(m)}$ and ${\sigma_{ql}^2}^{(m)}$ are assumed to be fixed. Thus, we have

$$c_{qlk} = -\max_{r_{qlk}\in[-1000,+10000]} \Big(h_{qlk}({\boldsymbol{\beta}_{ql}}^{(m)}, {\sigma_{ql}^2}^{(m)})\Big). \tag{C.11}$$

Using this, we can now write the integral $\int_{-\infty}^{+\infty} e^{h_{qlk}}\, dr_{qlk} = e^{-c_{qlk}} \int_{-\infty}^{+\infty} e^{h_{qlk}+c_{qlk}}\, dr_{qlk}$ and the edge based component of the variational bound is given as

$$\begin{aligned}\mathbb{E}(\log f(\boldsymbol{x}|\boldsymbol{Z};\boldsymbol{\beta},\boldsymbol{\sigma}^2)) = &\frac{1}{2}\sum_{k=1}^{K}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\sum_{t=1}^{T}\sum_{q=1}^{Q}\sum_{l\geq q}^{Q}\gamma_{ijql}x_{ijkt}\boldsymbol{d}_{kt}^{\top}\boldsymbol{\beta}_{ql}\\ &-\frac{K}{2}\sum_{q=1}^{Q}\sum_{l\geq q}^{Q}\log(2\Pi\sigma_{ql}^2) - \sum_{k=1}^{K}\sum_{q=1}^{Q}\sum_{l\geq q}^{Q}c_{qlk}\\ &+\sum_{k=1}^{K}\sum_{q=1}^{Q}\sum_{l\geq q}^{Q}\log\int_{-\infty}^{+\infty} e^{h_{qlk}+c_{qlk}}\, dr_{qlk},\end{aligned} \tag{C.12}$$

$$h_{qlk} = \frac{1}{2}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\sum_{t=1}^{T}\gamma_{ijql}x_{ijkt}r_{qlk} + \frac{1}{2}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\sum_{t=1}^{T}\gamma_{ijql}\log(1-\pi_{qlkt}) - \frac{r_{qlk}^2}{2\sigma_{ql}^2}. \tag{C.13}$$

Next, we state the list of all integrals $I_{qlk1}^c - \boldsymbol{I}_{qlk6}^c$,

$$I_{qlk1}^c = \int_{-\infty}^{+\infty} e^{h_{qlk}+c_{qlk}}\, dr_{qlk}, \tag{C.14}$$

$$I_{qlk2}^c = \frac{\partial I_{qlk1}^c}{\partial\sigma_{ql}^2} = \frac{1}{2\sigma_{ql}^4}\int_{-\infty}^{+\infty} e^{h_{qlk}+c_{qlk}} r_{qlk}^2\, dr_{qlk}, \tag{C.15}$$

$$\boldsymbol{I}^c_{qlk3} = \frac{\partial I^c_{qlk1}}{\partial \boldsymbol{\beta}_{ql}} = \left(-\frac{1}{2} \sum_{i=1}^{n} \sum_{j \neq i}^{n} \gamma_{ijql} \right) \sum_{t=1}^{T} \boldsymbol{d}_{kt} \int_{-\infty}^{+\infty} e^{h_{qlk}+c_{qlk}} \pi_{qlkt}\, dr_{qlk}, \tag{C.16}$$

$$I^c_{qlk4} = \frac{\partial I^c_{qlk2}}{\partial \sigma^2_{ql}} = \frac{1}{4\sigma^8_{ql}} \int_{-\infty}^{+\infty} e^{h_{qlk}+c_{qlk}} r^4_{qlk}\, dr_{qlk} - \frac{1}{\sigma^6_{ql}} \int_{-\infty}^{+\infty} e^{h_{qlk}+c_{qlk}} r^2_{qlk}\, dr_{qlk}, \tag{C.17}$$

$$\begin{aligned} \boldsymbol{I}^c_{qlk5} = \frac{\partial \boldsymbol{I}^{c\top}_{qlk3}}{\partial \boldsymbol{\beta}_{ql}} = & \left(-\frac{1}{2} \sum_{i=1}^{n} \sum_{j \neq i}^{n} \gamma_{ijql} \right)^2 \sum_{t=1}^{T} \sum_{t'=1}^{T} \boldsymbol{d}_{kt} \boldsymbol{d}^\top_{kt'} \int_{-\infty}^{+\infty} e^{h_{qlk}+c_{qlk}} \pi_{qlkt}\, \pi_{qlkt'}\, dr_{qlk} \\ & + \left(-\frac{1}{2} \sum_{i=1}^{n} \sum_{j \neq i}^{n} \gamma_{ijql} \right) \sum_{t=1}^{T} \boldsymbol{d}_{kt} \boldsymbol{d}^\top_{kt} \int_{-\infty}^{+\infty} e^{h_{qlk}+c_{qlk}} \pi_{qlkt}(1-\pi_{qlkt})\, dr_{qlk}, \end{aligned} \tag{C.18}$$

$$\begin{aligned} \boldsymbol{I}^c_{qlk6} = \frac{\partial \boldsymbol{I}^c_{qlk3}}{\partial \sigma^2_{ql}} &= \frac{\partial \boldsymbol{I}^c_{qlk2}}{\partial \boldsymbol{\beta}_{ql}} \\ &= \frac{1}{2\sigma^4_{ql}} \left(-\frac{1}{2} \sum_{i=1}^{n} \sum_{j \neq i}^{n} \gamma_{ijql} \right) \sum_{t=1}^{T} \boldsymbol{d}_{kt} \int_{-\infty}^{+\infty} e^{h_{qlk}+c_{qlk}} r^2_{qlk} \pi_{qlkt}\, dr_{qlk}. \end{aligned} \tag{C.19}$$

Equations for the score vector $\boldsymbol{U}(\boldsymbol{\beta}_{ql}, \sigma^2_{ql}) = (\boldsymbol{U}(\boldsymbol{\beta}_{ql})^\top, U(\sigma^2_{ql}))^\top$ are given as

$$\begin{aligned} \boldsymbol{U}(\boldsymbol{\beta}_{ql}) = & \frac{1}{2} \sum_{k=1}^{K} \sum_{i=1}^{n} \sum_{j \neq i}^{n} \sum_{t=1}^{T} \gamma_{ijql} x_{ijkt} \boldsymbol{d}_{kt} - \frac{1}{2} \sum_{k=1}^{K} \sum_{i=1}^{n} \sum_{j \neq i}^{n} \sum_{t=1}^{T} \gamma_{ijql} \psi_{qlkt} \boldsymbol{d}_{kt} \\ & + \sum_{k=1}^{K} \frac{\boldsymbol{I}^c_{qlk3}}{I^c_{qlk1}} \end{aligned} \tag{C.20}$$

$$U(\sigma^2_{ql}) = -\frac{K}{2\sigma^2_{ql}} + \sum_{k=1}^{K} \frac{I^c_{qlk2}}{I^c_{qlk1}}. \tag{C.21}$$

As mentioned in Chapter 5, the Fisher Information matrix $\boldsymbol{\mathcal{I}}(\boldsymbol{\beta}, \boldsymbol{\sigma}^2)$ is a block diagonal matrix, whose overall dimension is $(P+1)Q(Q+1)/2 \times (P+1)Q(Q+1)/2$. In particular, an individual block component noted as $\boldsymbol{\mathcal{I}}(\boldsymbol{\beta}_{ql}, \sigma^2_{ql})$ is a $(P+1)\times(P+1)$ matrix, and it can be represented as

$$\boldsymbol{\mathcal{I}}(\boldsymbol{\beta}_{ql}, \sigma^2_{ql}) = \left[\begin{array}{c|c} \boldsymbol{\mathcal{I}}_1(\boldsymbol{\beta}_{ql}) & \boldsymbol{\mathcal{I}}_2(\sigma^2_{ql}, \boldsymbol{\beta}_{ql}) \\ \hline \boldsymbol{\mathcal{I}}^\top_2(\sigma^2_{ql}, \boldsymbol{\beta}_{ql}) & \mathcal{I}_3(\sigma^2_{ql}) \end{array} \right], \tag{C.22}$$

such that

$$\boldsymbol{\mathcal{I}}_1(\boldsymbol{\beta}_{ql}) \equiv -\frac{\partial^2 \mathbb{E}(\log f(\boldsymbol{x}|\boldsymbol{Z}; \boldsymbol{\beta}, \boldsymbol{\sigma}^2))}{\partial \boldsymbol{\beta}_{ql} \partial \boldsymbol{\beta}^\top_{ql}} = -\sum_{k=1}^{K} \left[\frac{\boldsymbol{I}^c_{qlk5}}{I^c_{qlk1}} - \frac{\boldsymbol{I}^c_{qlk3} \boldsymbol{I}^{c\top}_{qlk3}}{\left(I^c_{qlk1}\right)^2} \right], \tag{C.23}$$

$$\boldsymbol{\mathcal{I}}_2(\sigma_{ql}^2, \boldsymbol{\beta}_{ql}) \equiv -\frac{\partial^2 \mathbb{E}(\log f(\boldsymbol{x}|\boldsymbol{Z}; \boldsymbol{\beta}, \boldsymbol{\sigma}^2))}{\partial \sigma_{ql}^2 \partial \boldsymbol{\beta}_{ql}} = -\sum_{k=1}^{K} \left[\frac{\boldsymbol{I}_{qlk6}^c}{I_{qlk1}^c} - \frac{\boldsymbol{I}_{qlk3}^c I_{qlk2}^c}{\left(I_{qlk1}^c\right)^2} \right], \tag{C.24}$$

$$\mathcal{I}_3(\sigma_{ql}^2) \equiv -\frac{\partial^2 \mathbb{E}(\log f(\boldsymbol{x}|\boldsymbol{Z}; \boldsymbol{\beta}, \boldsymbol{\sigma}^2))}{\partial \sigma_{ql}^2 \partial \sigma_{ql}^2} = -\frac{K}{2\sigma_{ql}^4} - \sum_{k=1}^{K} \left[\frac{I_{qlk4}^c}{I_{qlk1}^c} - \frac{\left(I_{qlk2}^c\right)^2}{\left(I_{qlk1}^c\right)^2} \right]. \tag{C.25}$$

Bibliography

Sophie Achard and Ed Bullmore. Efficiency and cost of economical brain functional networks. *PLoS Comput Biol*, 3(2):e17–e17, 2007.

Sophie Achard, Raymond Salvador, Brandon Whitcher, John Suckling, and ED Bullmore. A resilient, low-frequency, small-world human brain functional network with highly connected association cortical hubs. *The Journal of neuroscience*, 26(1): 63–72, 2006.

Adelin Albert and JA Anderson. On the existence of maximum likelihood estimates in logistic regression models. *Biometrika*, 71(1):1–10, 1984.

Réka Albert, István Albert, and Gary L Nakarado. Structural vulnerability of the north american power grid. *Physical Review E*, 69(2):025103, 2004.

Aaron F Alexander-Bloch, Petra E Vértes, Reva Stidd, François Lalonde, Liv Clasen, Judith Rapoport, Jay Giedd, Edward T Bullmore, and Nitin Gogtay. The anatomical distance of functional connections predicts brain network topology in health and schizophrenia. *Cerebral cortex*, 23(1):127–138, 2013.

ZF Altun and DH Hall. Handbook of *C. elegans* anatomy. *WormAtlas. http://www. wormatlas. org/handbook/contents. htm*, 2005.

Hall DH Altun ZF. *Nervous system, general description.* Worm atlas, 2009.

Christophe Ambroise and Catherine Matias. New consistent and asymptotically normal parameter estimates for random-graph mixture models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 74(1):3–35, 2012.

John Ashburner and Karl J Friston. *Rigid body registration.* Academic Press, London, 2007.

Danielle S Bassett and Edward T Bullmore. Human brain networks in health and disease. *Current opinion in neurology*, 22(4):340, 2009.

Danielle S Bassett, Andreas Meyer-Lindenberg, Sophie Achard, Thomas Duke, and Edward Bullmore. Adaptive reconfiguration of fractal small-world human brain functional networks. *Proceedings of the National Academy of Sciences*, 103(51): 19518–19523, 2006.

Danielle S Bassett, Edward Bullmore, Beth A Verchinski, Venkata S Mattay, Daniel R Weinberger, and Andreas Meyer-Lindenberg. Hierarchical organization of human cortical networks in health and schizophrenia. *The Journal of Neuroscience*, 28(37):9239–9248, 2008.

Douglas Bates, Martin Maechler, and Ben Bolker. *lme4: Linear mixed-effects models using S4 classes*, 2012. URL `http://CRAN.R-project.org/package=lme4`. R package version 0.999999-0.

Robert Beverly and Karen R Sollins. An internet protocol address clustering algorithm. In *SysML*, 2008.

Christophe Biernacki and Gérard Govaert. Using the classification likelihood to choose the number of clusters. *Computing Science and Statistics*, pages 451–457, 1997.

Christophe Biernacki, Gilles Celeux, Gérard Govaert, et al. Assessing a mixture model for clustering with the integrated classification likelihood. 1998.

V.D. Blondel, J.L. Guillaume, R. Lambiotte, and E. Lefebvre. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10):P10008, 2008.

Stephen P Borgatti and Martin G Everett. Models of core/periphery structures. *Social networks*, 21(4):375–395, 2000.

Sydney Brenner. The genetics of *Caenorhabditis elegans*. *Genetics*, 77(1):71–94, 1974.

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

Ed Bullmore, Jalal Fadili, Michael Breakspear, Raymond Salvador, John Suckling, and Michael Brammer. Wavelets and statistical analysis of functional magnetic resonance images of the human brain. *Statistical methods in medical research*, 12 (5):375–399, 2003.

Edward T Bullmore and Danielle S Bassett. Brain graphs: graphical models of the human brain connectome. *Annual review of clinical psychology*, 7:113–140, 2011.

Miao Cao, Jin-Hui Wang, Zheng-Jia Dai, Xiao-Yan Cao, Li-Li Jiang, Feng-Mei Fan, Xiao-Wei Song, Ming-Rui Xia, Ni Shu, Qi Dong, et al. Topological organization of the human brain functional connectome across the lifespan. *Developmental cognitive neuroscience*, 7:76–93, 2014.

B.L. Chen, D.H. Hall, and D.B. Chklovskii. Wiring optimization can relate neuronal structure and function. *Proceedings of the National Academy of Sciences of the United States of America*, 103(12):4723–4728, 2006.

Elizabeth A Claassen. *A Reduced Bias Method of Estimating Variance Components in Generalized Linear Mixed Models*. PhD thesis, University of Nebraska-Lincoln, 2014.

Douglas B Clarkson and Robert I Jennrich. Computing extended maximum likelihood estimates for linear parameter models. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 417–426, 1991.

Lilia Costa, Jim Smith, Thomas Nichols, James Cussens, Eugene P Duff, Tamar R Makin, et al. Searching multiregression dynamic models of resting-state fmri networks using integer programming. *Bayesian Analysis*, 10(2):441–478, 2015.

R Cameron Craddock, G Andrew James, Paul E Holtzheimer, Xiaoping P Hu, and Helen S Mayberg. A whole brain fmri atlas generated via spatially constrained spectral clustering. *Human brain mapping*, 33(8):1914–1928, 2012.

Abhijit Dasgupta and Adrian E Raftery. Detecting features in spatial point processes with clutter via model-based clustering. *Journal of the American Statistical Association*, 93(441):294–302, 1998.

J-J Daudin, Franck Picard, and Stéphane Robin. A mixture model for random graphs. *Statistics and computing*, 18(2):173–183, 2008.

Catherine E Davey, David B Grayden, Gary F Egan, and Leigh A Johnston. Filtering induces correlation in fmri resting state data. *Neuroimage*, 64:728–740, 2013.

Eugene Demidenko. *Mixed Models: Theory and Applications (Wiley Series in Probability and Statistics)*. Wiley-Interscience, 2004.

Peter Dicken, Philip F Kelly, Kris Olds, and Henry Wai-Chung Yeung. Chains and networks, territories and scales: towards a relational framework for analysing the global economy. *Global networks*, 1(2):89–112, 2002.

Jean Diebolt and Christian P Robert. Estimation of finite mixture distributions through bayesian sampling. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 363–375, 1994.

A.J. Dobson. *An introduction to generalized linear models*. Chapman & Hall/CRC, 2001.

Ian Ellison-Wright, David C Glahn, Angela R Laird, Sarah M Thelen, et al. The anatomy of first-episode and chronic schizophrenia: An anatomical likelihood estimation meta-analysis. *The American journal of psychiatry*, 165(8):1015, 2008.

P. Erdős. Some remarks on the theory of graphs. *Bull. Amer. Math. Soc*, 53:292–294, 1947.

P. Erdős and A. Rényi. On random graphs. *Publicationes Mathematicae Debrecen*, 6:290–297, 1959.

Brian S Everitt and David J Hand. Finite mixture distributions. *Monographs on Applied Probability and Statistics, London: Chapman and Hall, 1981*, 1, 1981.

David Firth. Bias reduction of maximum likelihood estimates. *Biometrika*, 80(1): 27–38, 1993.

Ronald A Fisher. Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population. *Biometrika*, pages 507–521, 1915.

Santo Fortunato. Community detection in graphs. *Physics Reports*, 486(3):75–174, 2010.

Chris Fraley and Adrian E Raftery. How many clusters? which clustering method? answers via model-based cluster analysis. *The computer journal*, 41(8):578–588, 1998.

Chris Fraley and Adrian E. Raftery. Model-based clustering, discriminant analysis and density estimation. *Journal of the American Statistical Association*, 97:611–631, 2002.

Chris Fraley and Adrian E. Raftery. *MCLUST Version 3 for R: Normal Mixture Modeling and Model-based Clustering*, 2006. (revised in 2012).

Karl J Friston, John Ashburner, Christopher D Frith, Jean-Baptiste Poline, John D Heather, Richard SJ Frackowiak, et al. Spatial registration and normalization of images. *Human brain mapping*, 3(3):165–189, 1995.

Karl J Friston et al. Functional and effective connectivity in neuroimaging: a synthesis. *Human brain mapping*, 2(1-2):56–78, 1994.

KJ Friston, CD Frith, PF Liddle, and RSJ Frackowiak. Functional connectivity: the principal-component analysis of large (pet) data sets. *Journal of cerebral blood flow and metabolism*, 13:5–5, 1993.

Cedric E Ginestet, Thomas E Nichols, Ed T Bullmore, and Andrew Simmons. Brain network analysis: separating cost from topology using cost-integration. *PloS one*, 6(7):e21570, 2011.

Phillip Good. *Permutation tests*. Springer, 2000.

Isobel Claire Gormley and Thomas Brendan Murphy. Mixture of experts modelling with social science applications. *Mixtures: Estimation and Applications*, pages 101–121, 2011.

Patric Hagmann, Maciej Kurant, Xavier Gigandet, Patrick Thiran, Van J Wedeen, Reto Meuli, and Jean-Philippe Thiran. Mapping human whole-brain structural networks with diffusion mri. *PloS one*, 2(7):e597, 2007.

J. Handl, J. Knowles, and D.B. Kell. Computational cluster validation in post-genomic data analysis. *Bioinformatics*, 21(15):3201–3212, 2005.

D.A. Harville. Maximum likelihood approaches to variance component estimation and to related problems. *Journal of the American Statistical Association*, 72(358): 320–338, 1977.

G. Haspel, M.J. O'Donovan, and A.C. Hart. Motoneurons dedicated to either forward or backward locomotion in the nematode *Caenorhabditis elegans*. *The Journal of Neuroscience*, 30(33):11151–11156, 2010.

Gal Haspel and Michael J O'Donovan. A perimotor framework reveals functional segmentation in the motoneuronal network controlling locomotion in caenorhabditis elegans. *The Journal of Neuroscience*, 31(41):14611–14623, 2011.

Georg Heinze and Michael Schemper. A solution to the problem of separation in logistic regression. *Statistics in medicine*, 21(16):2409–2419, 2002.

RNA Henson, C Buechel, O Josephs, and KJ Friston. The slice-timing problem in event-related fmri. *NeuroImage*, 9:125, 1999.

Claus-C Hilgetag, Gully APC Burns, Marc A O'Neill, Jack W Scannell, and Malcolm P Young. Anatomical connectivity defines the organization of clusters of cortical areas in the macaque and the cat. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 355(1393):91–110, 2000.

Karim F Hirji, Anastasios A Tsiatis, and Cyrus R Mehta. Median unbiased estimation for binary data. *The American Statistician*, 43(1):7–11, 1989.

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

Paul W Holland, Kathryn Blackmond Laskey, and Samuel Leinhardt. Stochastic blockmodels: First steps. *Social networks*, 5(2):109–137, 1983.

Sture Holm. A simple sequentially rejective multiple test procedure. *Scandinavian journal of statistics*, pages 65–70, 1979.

Petter Holme. Core-periphery organization of complex networks. *Physical Review E*, 72(4):046111, 2005.

I.A. Hope. C. elegans*: A practical approach*, volume 213. Oxford University Press, USA, 1999.

Hai-Bo Hu and Xiao-Fan Wang. Disassortative mixing in online social networks. *EPL (Europhysics Letters)*, 86(1):18003, 2009.

L. Hubert and P. Arabie. Comparing partitions. *Journal of classification*, 2(1): 193–218, 1985.

Y Iturria-Medina, EJ Canales-Rodriguez, L Melie-Garcia, PA Valdes-Hernandez, E Martinez-Montes, Y Alemán-Gómez, and JM Sánchez-Bornot. Characterizing brain anatomical connections using diffusion weighted mri and graph theory. *Neuroimage*, 36(3):645–660, 2007.

Heidi Johansen-Berg, TEJ Behrens, MD Robson, I Drobnjak, MFS Rushworth, JM Brady, SM Smith, DJ Higham, and PM Matthews. Changes in connectivity profiles define functionally distinct regions in human medial frontal cortex. *Proceedings of the National Academy of Sciences of the United States of America*, 101(36):13335–13340, 2004.

Marcus Kaiser and Sreedevi Varier. Evolution and development of brain networks: from caenorhabditis elegans to homo sapiens. *Network: Computation in Neural Systems*, 22(1-4):143–147, 2011.

Joanne SM Kim. The nervous system of c. elegans: Its in vivo visualization and practical applications. *Hypothesis*, 4(2):29, 2008.

John E Kolassa. Infinite parameter estimates in logistic regression, with application to approximate conditional inference. *Scandinavian journal of statistics*, 24(4): 523–530, 1997.

Marek Kubicki, Robert W McCarley, and Martha E Shenton. Evidence for white matter abnormalities in schizophrenia. *Current opinion in psychiatry*, 18(2):121, 2005.

Marek Kubickia, Robert McCarleya, Carl-Fredrik Westinb, Hae-Jeong Parka, Stephan Maierc, Ron Kikinisb, Ferenc A Joleszc, and Martha E Shenton. A review of diffusion tensor imaging studies in schizophrenia. *J Psychiatr Res*, 41 (1-2):15–30, 2007.

Angela R Laird, P Mickle Fox, Simon B Eickhoff, Jessica A Turner, Kimberly L Ray, D Reese McKay, David C Glahn, Christian F Beckmann, Stephen M Smith, and Peter T Fox. Behavioral interpretations of intrinsic connectivity networks. *Journal of cognitive neuroscience*, 23(12):4022–4037, 2011.

Pierre Latouche, Etienne Birmele, and Christophe Ambroise. Bayesian methods for graph clustering. SSB Research Report, 2008.

Pierre Latouche, Etienne Birmelé, Christophe Ambroise, et al. Overlapping stochastic block models with application to the french political blogosphere. *The Annals of Applied Statistics*, 5(1):309–336, 2011.

Brian G Leroux and Martin L Puterman. Maximum-penalized-likelihood estimation for independent and markov-dependent mixture models. *Biometrics*, pages 545–558, 1992.

Emmanuel Lesaffre and Adelin Albert. Partial separation in logistic discrimination. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 109–116, 1989.

Emmanuel Lesaffre and Brian D Marx. Collinearity in generalized linear regression. *Communications in Statistics-Theory and Methods*, 22(7):1933–1952, 1993.

Emmanuel Lesaffre and Bart Spiessens. On the effect of the number of quadrature points in a logistic random effects model: an example. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 50(3):325–335, 2001.

Meng Liang, Yuan Zhou, Tianzi Jiang, Zhening Liu, Lixia Tian, Haihong Liu, and Yihui Hao. Widespread functional disconnectivity in schizophrenia with resting-state functional magnetic resonance imaging. *Neuroreport*, 17(2):209–213, 2006.

Qing Liu and Donald A Pierce. A note on gausshermite quadrature. *Biometrika*, 81(3):624–629, 1994.

Yong Liu, Meng Liang, Yuan Zhou, Yong He, Yihui Hao, Ming Song, Chunshui Yu, Haihong Liu, Zhening Liu, and Tianzi Jiang. Disrupted small-world networks in schizophrenia. *Brain*, 131(4):945–961, 2008.

Francois Lorrain and Harrison C White. Structural equivalence of individuals in social networks. *The Journal of mathematical sociology*, 1(1):49–80, 1971.

R.D. Luce. Connectivity and generalized cliques in sociometric group structure. *Psychometrika*, 15(2):169–190, 1950.

T. Luczak. Sparse random graphs with a given degree sequence. In *Proceedings of the Symposium on Random Graphs, Poznan*, pages 165–182, 1989.

Mary-Ellen Lynall, Danielle S Bassett, Robert Kerwin, Peter J McKenna, Manfred Kitzbichler, Ulrich Muller, and Ed Bullmore. Functional connectivity and brain networks in schizophrenia. *The Journal of Neuroscience*, 30(28):9477–9487, 2010.

Mahendra Mariadassou, Stéphane Robin, Corinne Vacher, et al. Uncovering latent structure in valued graphs: a variational approach. *The Annals of Applied Statistics*, 4(2):715–742, 2010.

Sergei Maslov and Kim Sneppen. Specificity and stability in topology of protein networks. *Science*, 296(5569):910–913, 2002.

Sergei Maslov, Kim Sneppen, and Alexei Zaliznyak. Detection of topological patterns in complex networks: correlation profile of the internet. *Physica A: Statistical Mechanics and its Applications*, 333:529–540, 2004.

Charles E McCulloch. Maximum likelihood algorithms for generalized linear mixed models. *Journal of the American statistical Association*, 92(437):162–170, 1997.

David Meunier, Renaud Lambiotte, and Edward T Bullmore. Modular and hierarchically modular organization of brain networks. *Frontiers in neuroscience*, 4, 2010.

Sifis Micheloyannis, Ellie Pachou, Cornelis Jan Stam, Michael Breakspear, Panagiotis Bitsios, Michael Vourkas, Sophia Erimaki, and Michael Zervakis. Small world networks and disturbed functional connectivity in schizophrenia. *Schizophrenia research*, 87(1):60–66, 2006.

Saeed Moghaddam, Ahmed Helmy, Sanjay Ranka, and Manas Somaiya. Data-driven co-clustering model of internet usage in large mobile societies. In *Proceedings of the 13th ACM international conference on Modeling, analysis, and simulation of wireless and mobile systems*, pages 248–256. ACM, 2010.

M. Molloy and B. Reed. A critical point for random graphs with a given degree sequence. *Random Structures & Algorithms*, 6(2-3):161–180, 1995.

J.L. Moreno. *Who shall survive?* Beacon House New York, 1953.

L.C. Morey and A. Agresti. The measurement of classification agreement: an adjustment to the rand statistic for chance agreement. *Educational and Psychological Measurement*, 44(1):33–37, 1984.

Morten Mørup, Kristoffer Madsen, Anne-marie Dogonowski, Hartwig Siebner, and Lars K Hansen. Infinite relational modeling of functional connectivity in resting state fmri. In *Advances in Neural Information Processing Systems*, pages 1750–1758, 2010.

Sach Mukherjee and Steven M Hill. Network clustering: probing biological heterogeneity by sparse graphical models. *Bioinformatics*, 27(7):994–1000, 2011.

H E Nelson. The national adult reading test (nart): test manual (windsor, berks, uk, nfer nelson). 1982.

Mark Newman. *Networks: an introduction*. Oxford University Press, 2010.

Mark EJ Newman. The structure and function of complex networks. *SIAM review*, 45(2):167–256, 2003.

Mark EJ Newman, Duncan J Watts, and Steven H Strogatz. Random graph models of social networks. *Proceedings of the National Academy of Sciences*, 99(suppl 1): 2566–2572, 2002.

M.E.J. Newman. Modularity and community structure in networks. *Proceedings of the National Academy of Sciences*, 103(23):8577–8582, 2006.

M.E.J. Newman and M. Girvan. Finding and evaluating community structure in networks. *Physical review E*, 69(2):026113, 2004.

Thomas Nichols and Satoru Hayasaka. Controlling the familywise error rate in functional neuroimaging: a comparative review. *Statistical methods in medical research*, 12(5):419–446, 2003.

Seiji Ogawa, Tso-Ming Lee, Alan R Kay, and David W Tank. Brain magnetic resonance imaging with contrast dependent on blood oxygenation. *Proceedings of the National Academy of Sciences*, 87(24):9868–9872, 1990.

Raj Kumar Pan, Nivedita Chatterjee, and Sitabhra Sinha. Mesoscopic organization reveals the constraints governing caenorhabditis elegans nervous system. 2010.

PE Pattison and GL Robins. Handbook of probability theory with applications. chapter probabilistic network theory, 2007.

Dragana M Pavlović, Petra E Vértes, Edward T Bullmore, William R Schafer, and Thomas E Nichols. Stochastic blockmodeling of the modules and core of the caenorhabditis elegans connectome. *PloS one*, 9(7):e97584, 2014.

Donald B Percival and Andrew T Walden. Wavelet methods for time series analysis (cambridge series in statistical and probabilistic mathematics). 2000.

Robert Piessens, De Doncker-Kapenga, C Überhuber, David Kahaner, et al. Quadpack, a subroutine package for automatic integration. *status: published*, page 301p, 1983.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2015. URL `http://www.R-project.org/`.

W.M. Rand. Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical association*, 66(336):846–850, 1971.

Erzsébet Ravasz and Albert-László Barabási. Hierarchical organization in complex networks. *Physical Review E*, 67(2):026112, 2003.

Erzsébet Ravasz, Anna Lisa Somera, Dale A Mongru, Zoltán N Oltvai, and A-L Barabási. Hierarchical organization of modularity in metabolic networks. *science*, 297(5586):1551–1555, 2002.

Christian P Robert. The bayesian choice: From decision-theoretic foundations to computational implementation (springer texts in statistics) by. 2001.

Loic Royer, Matthias Reimann, Bill Andreopoulos, and Michael Schroeder. Unraveling protein networks with power graph analysis. *PLoS computational biology*, 4 (7):e1000108, 2008.

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

Thomas J Santner and Diane E Duffy. A note on a. albert and ja anderson's conditions for the existence of maximum likelihood estimates in logistic regression models. *Biometrika*, pages 755–758, 1986.

Stephen M Smith, Mark Jenkinson, Mark W Woolrich, Christian F Beckmann, Timothy EJ Behrens, Heidi Johansen-Berg, Peter R Bannister, Marilena De Luca, Ivana Drobnjak, David E Flitney, et al. Advances in functional and structural mr image analysis and implementation as fsl. *Neuroimage*, 23:S208–S219, 2004.

Stephen M Smith, Peter T Fox, Karla L Miller, David C Glahn, P Mickle Fox, Clare E Mackay, Nicola Filippini, Kate E Watkins, Roberto Toro, Angela R Laird, et al. Correspondence of the brain's functional architecture during activation and rest. *Proceedings of the National Academy of Sciences*, 106(31):13040–13045, 2009.

Stephen M Smith, Karla L Miller, Gholamreza Salimi-Khorshidi, Matthew Webster, Christian F Beckmann, Thomas E Nichols, Joseph D Ramsey, and Mark W Woolrich. Network modelling methods for fmri. *Neuroimage*, 54(2):875–891, 2011.

T.A.B. Snijders and K. Nowicki. Estimation and prediction for stochastic blockmodels for graphs with latent block structure. *Journal of Classification*, 14(1): 75–100, 1997.

Olaf Sporns. *Networks of the Brain*. MIT Press, 2010.

Olaf Sporns, Dante R Chialvo, Marcus Kaiser, and Claus C Hilgetag. Organization, development and function of complex brain networks. *Trends in cognitive sciences*, 8(9):418–425, 2004.

Russell J Steele and Adrian E Raftery. Performance of bayesian model selection criteria for gaussian mixture models. *Dept. Stat., Univ. Washington, Washington, DC, Tech. Rep*, 559, 2009.

J Suckling, C Long, C Triantafyllou, M Brammer, and E Bullmore. Variable precision registration via wavelets: optimal spatial scales for inter-subject registration of functional mri. *Neuroimage*, 31(1):197–208, 2006.

Jeremy Tantrum, Alejandro Murua, and Werner Stuetzle. Assessment and pruning of hierarchical model based clustering. In *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 197–205. ACM, 2003.

Bertrand Thirion, Guillaume Flandin, Philippe Pinel, Alexis Roche, Philippe Ciuciu, and Jean-Baptiste Poline. Dealing with the shortcomings of spatial normalization: Multi-subject parcellation of fmri datasets. *Human brain mapping*, 27(8):678–693, 2006.

Giulio Tononi, Gerald M Edelman, and Olaf Sporns. Complexity and coherency: integrating information in the brain. *Trends in cognitive sciences*, 2(12):474–484, 1998.

Emma K Towlson, Petra E Vértes, Sebastian E Ahnert, and Edward T Bullmore. Brain networks reveal the effects of antipsychotic drugs on schizophrenia patients and controls.

Emma K Towlson, Petra E Vértes, Sebastian E Ahnert, William R Schafer, and Edward T Bullmore. The rich club of the *C. elegans* neuronal connectome. *The Journal of Neuroscience*, 33(15):6380–6387, 2013.

Nathalie Tzourio-Mazoyer, Brigitte Landeau, Dimitri Papathanassiou, Fabrice Crivello, Olivier Etard, Nicolas Delcroix, Bernard Mazoyer, and Marc Joliot. Automated anatomical labeling of activations in spm using a macroscopic anatomical parcellation of the mni mri single-subject brain. *Neuroimage*, 15(1):273–289, 2002.

Robert E Ulanowicz. Quantitative methods for ecological network analysis. *Computational Biology and Chemistry*, 28(5):321–339, 2004.

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

Martijn P Van den Heuvel and Olaf Sporns. Rich-club organization of the human connectome. *The Journal of neuroscience*, 31(44):15775–15786, 2011.

Martijn P Van Den Heuvel and Olaf Sporns. Network hubs in the human brain. *Trends in cognitive sciences*, 17(12):683–696, 2013.

Bernadette CM Van Wijk, Cornelis J Stam, and Andreas Daffertshofer. Comparing brain networks of different size and connectivity density using graph theory. *PloS one*, 5(10):e13701, 2010.

S. Varier and M. Kaiser. Neural development features: Spatio-temporal development of the *Caenorhabditis elegans* neuronal network. *PLoS computational biology*, 7 (1):e1001044, 2011.

Gaël Varoquaux and R Cameron Craddock. Learning and comparing functional connectomes across subjects. *NeuroImage*, 80:405–415, 2013.

L.R. Varshney, B.L. Chen, E. Paniagua, D.H. Hall, and D.B. Chklovskii. Structural properties of the *Caenorhabditis elegans* neuronal network. *PLoS computational biology*, 7(2):e1001066, 2011.

Stephen E Von Stetina, Millet Treinin, and DM Miller. The motor circuit. *Int. Rev. Neurobiol*, 69:125–167, 2006.

Duy Quang Vu, David R Hunter, and Michael Schweinberger. Model-based clustering of large networks. *arXiv preprint arXiv:1207.0188*, 2012.

Y.J. Wang and G.Y. Wong. Stochastic blockmodels for directed graphs. *Journal of the American Statistical Association*, pages 8–19, 1987.

Hiroshi Watanabe, Toshitaka Fujisawa, and Thomas W Holstein. Cnidarians and the evolutionary origin of the nervous system. *Development, growth & differentiation*, 51(3):167–183, 2009.

Songjie Wei, Jelena Mirkovic, and Ezra Kissel. Profiling and clustering internet hosts. *DMIN*, 6:269–75, 2006.

Daniel R Weinberger, Michael F Egan, Alessandro Bertolino, Joseph H Callicott, Venkata S Mattay, Barbara K Lipska, Karen F Berman, and Terry E Goldberg. Prefrontal neurons and the genetics of schizophrenia. *Biological psychiatry*, 50 (11):825–844, 2001.

Peter H Westfall and S Stanley Young. *Resampling-based multiple testing: Examples and methods for p-value adjustment*, volume 279. John Wiley & Sons, 1993.

JG White, E Southgate, JN Thomson, and S Brenner. The structure of the nervous system of the nematode caenorhabditis elegans: the mind of a worm. *Phil. Trans. R. Soc. Lond*, 314:1–340, 1986.

Anderson M Winkler, Gerard R Ridgway, Matthew A Webster, Stephen M Smith, and Thomas E Nichols. Permutation inference for the general linear model. *NeuroImage*, 92:381–397, 2014.

William Barry Wood et al. *The nematode Caenorhabditis elegans.* Cold Spring Harbour Laboratory, 1987.

Mingrui Xia, Jinhui Wang, and Yong He. Brainnet viewer: a network visualization tool for human brain connectomics. *PLoS One*, 8(7):e68910, 2013.

BT Thomas Yeo, Fenna M Krienen, Jorge Sepulcre, Mert R Sabuncu, Danial Lashkari, Marisa Hollinshead, Joshua L Roffman, Jordan W Smoller, Lilla Zöllei, Jonathan R Polimeni, et al. The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *Journal of neurophysiology*, 106(3): 1125–1165, 2011.

Andrew Zalesky and Alex Fornito. A dti-derived measure of cortico-cortical connectivity. *Medical Imaging, IEEE Transactions on*, 28(7):1023–1036, 2009.

Andrew Zalesky, Alex Fornito, and Ed Bullmore. On the use of correlation as a measure of network connectivity. *Neuroimage*, 60(4):2096–2106, 2012.

Hugo Zanghi, Christophe Ambroise, and Vincent Miele. Fast online graph clustering via erdos-renyi mixture. *Pattern Recognition*, 41:3592–3599, 2008a.

Hugo Zanghi, Franck Picard, Vincent Miele, and Christophe Ambroise. Strategies for online inference of network mixture. Research publication, 2008b.

Hugo Zanghi, Stevenn Volant, and Christophe Ambroise. Clustering based on random graph model embedding vertex features. *Pattern Recognition Letters*, 31(9): 830–836, 2010.

Zhi-Qiang Zhang. Animal biodiversity: An update of classification and diversity in 2013. in: Zhang, z.-q.(ed.) animal biodiversity: An outline of higher-level classification and survey of taxonomic richness (addenda 2013). *Zootaxa*, 3703(1):5–11, 2013.