Integration of Multimodal MR Data

Karen Marie Sandø Ambrosen

Kongens Lyngby 2013
IMM-M.Sc.-2013-4

Technical University of Denmark
Informatics and Mathematical Modelling
Building 321, DK-2800 Kongens Lyngby, Denmark
Phone +45 45253351, Fax +45 45882673
reception@imm.dtu.dk
www.imm.dtu.dk IMM-M.Sc.-2013-4

Summary (English)

The brain is an extremely complex organ. Understanding the function and the anatomy of the brain that underlies normal and adaptive behaviours, are the foundation for understanding the pathology and the pathophysiology of neurological diseases. Both functional magnetic resonance imaging (fMRI) and diffusion weighted resonance imaging (DWI) give an insight into how the brain works by imaging the functional and the structural connectivity of the brain, respectively.
The stochastic blockmodel, the Infinite Relational Model (IRM), can be used to find the structure in relational data. It partitions the graph such that equivalent nodes are assigned to the same cluster. The optimal partition is found by maximizing the likelihood of the model. In this project the IRM is used to find the structure in fMRI and DWI data.

In this project fMRI data from 29 healthy subjects is used. During the acquisition the subjects perform finger tapping with alternating left and right hand. The pre-processing of the fMRI data is performed in SPM8, and the activated regions in the motor cortex and the striatum is found. The fMRI graphs are created by extracting the time series in the voxels in the regions of interest (ROIs), and by calculating the correlation between the time series. Three graphs for each subject are created, one for each of the states: Baseline, Left, and Right.
DWI data from 14 of the subjects is used. The processing of the DWI data is performed in FSL. Probabilistic tractography is performed to find the streamlines between the voxels in the ROIs. The DWI graphs are obtained by assuming a link between two voxels if a streamline exists.

Two different analyses are performed. An analysis only regarding fMRI data and an analysis of combined fMRI and DWI data. In each of the analyses, two hypotheses are tested: 1) The partition of the ROIs is the same in the different states and 2) The communication pattern is the same in the different states, assuming the partition is the same.

When testing the first hypothesis in both analyses, it is found that the partition of the ROIs is significantly different between the different states, but despite of the significant difference, the differences are hard to interpret by visual inspection. The different partition between the states show that the IRM is highly sensitive, since, it is sensitive to the task performed. The structure found by the IRM is consistent across states, as the number of clusters only vary a little between the states.
When testing the second hypothesis, it is found that the communication pattern between the clusters, found by the IRM, is significantly different in the different states in both analyses. This result suggest that the link probability matrix is highly sensitive to the task performed. The inspection of the cluster pairs with pair-wise most significant difference in link probability between the states, indicates that there is a coordination between cortex and striatum, as clusters in cortex and striatum are connected.
Furthermore, it is found that the IRM is reliable, as it is stable across runs.

Summary (Danish)

Hjernen er et ekstremt komplekst organ. At forstå funktionen og anatomien af hjernen, der ligger til grund for almindelig og adaptiv opførsel er grundlæggende for at forstå patologien og patofysiologien i neurologiske sygdomme. Både funktionel magnetisk resonans billeddannelse (fMRI) og diffusionsvægtet billeddannelse (DWI) giver et indblik i hvordan hjernen fungerer, ved at kunne måle hhv. de funktionelle og de strukturelle forbindelser i hjernen.
Den stokastiske blokmodel, *the Infinite Relational Model* (IRM), kan bruges til at finde strukturen i relationel data. Modellen opdeler grafen således at ækvivalente knuder samles i den samme gruppe. Den optimale opdeling findes ved at maksimere modellens likelihood. I dette projekt bruges IRM til at finde strukturen i fMRI og DWI data.

I dette projekt anvendes der fMRI data fra 29 raske personer. Under optagelsen udfører personerne *finger tapping* med skiftevis højre og venstre hånd. Behandlingen af fMRI data udføres i SPM8, og de aktiverede områder i motor cortex og striatum (ROIs) findes. fMRI graferne dannes ved at udtrække tidsserierne i hver voxel i ROIs og beregne korrelationen mellem tidsserierne. Der dannes tre grafer for hver person, en for hver tilstand: Baseline, Venstre og Højre.
Der anvendes DWI data fra 14 af personerne. Behandlingen af DWI data udføres i FSL. Der laves probabilistisk traktografi for at finde forbindelserne mellem voxels i ROIs. DWI graferne genereres ved at antage at der findes et link mellem to voxels, hvis der i traktografien findes en forbindelse.

Der laves to forskellige analyser. En analyse der kun tager højde for fMRI data,

og en analyse der kombinerer fMRI og DWI data. I hver af analyserne testes to hypoteser: 1) Opdelingen af ROIs er parvis den samme i de forskellige tilstande, og 2) Kommunikations mønstret er parvis det samme i de forskellige tilstande, antaget at opdelingen er den samme.

Når den første hypotese testes, findes i begge analyser, at opdelingen af ROIs er signifikant forskellig mellem tilstande. På trods af den signifikante forskel er det dog svært at fortolke forskellen ved visuel inspektion. At opdelingen er forskellig mellem tilstande viser, at IRM er et meget følsom, da den er følsom overfor den opgave der udføres. Strukturen der findes er konsistent over tilstande, da antallet af grupper, der findes, i de forskellige tilstande kun varierer en smule.
Når den anden hypotese testes, findes i begge analyser at kommunikations mønstret mellem grupperne er parvist signifikant forskelligt i de forskellige tilstande. Dette resultat viser, at matricen med link sandsynligheder er meget følsom overfor den opgave der udføres. Den visuelle inspektion af de grupper der er mest signifikant forskellige i deres link sandsynlighed mellem tilstande, indikerer at der er koordination mellem cortex og striatum, da grupper i cortex og striatum er forbundet.
Det findes endvidere at IRM er pålidelig, da den er stabil over kørsler.

Preface

This thesis was prepared at the department of Informatics and Mathematical Modelling at the Technical University of Denmark in fulfilment of the requirements for acquiring an M.Sc. in Medicine and Technology.

The thesis deals with integration of functional and diffusion weighted magnetic resonance imaging. The partition of and the communication patterns between the motor cortex and the striatum are investigated using stochastic block modelling.

The project was carried out under the supervision of Prof. Lars Kai Hansen from the Cognitive Systems Group, IMM, Technical University of Denmark. The co-supervisors on the project were PhD Kristoffer Hougaard Madsen from Danish Research Centre for Magnetic Resonance (DRCMR), Copenhagen University Hospital Hvidovre, and PhD-student Kasper Winther Andersen from the Cognitive Systems Group, IMM, Technical University of Denmark.

Lyngby, 04-February-2013

Karen Marie Sandø Ambrosen

Acknowledgements

I would like to thank my supervisor Lars Kai Hansen for your support and good advices. A thank to Kristoffer Hougaard Madsen for providing the data and help on pre-processing of the fMRI data. Also, a huge thank to my supervisor Kasper Winther Andersen for your always quick and thorough answers on all my many questions.

I would like to thank my sister Anna and my good friend Camilla for reading and commenting on the thesis. Also, a thank to my fellow students for all the time spent together the last five month. The working days had not been the same without your company. Finally, huge thanks to my loving husband Bjarne for all your patience, support and constant faith in me.

Contents

Chapter 1

Introduction

1.1 The Neuron

The brain is made up of billions of nerve cells, also called neurons. The neurons are organized to form complex networks that perform the functions of the nervous system. A neuron consists of a soma (neuron cell body), dendrites, and an axon (nerve fiber), as illustrated in Figure 1.1.
The soma is the central part of the neuron and includes the nucleus, which contains the genetic information, and controls the activities of the cell by regulating gene expression and protein synthesis. The specialised part of the soma that connects to the axon is called the axon Hillock.
From the soma, the dendrites branch out like the branches on a tree. The dendrites receive inhibitory and excitatory signals from other neurons via their axons. The communication between the dendrites and the axons takes place in the terminal cleft. The electrical signal propagates along the axon towards the axonal terminal. When reaching the terminal the electrical signal sets free chemical substances, which generates an electrical signal in the dendrite. In the axon Hillock all the inputs from the dendrites are summated and an action potential is build up. When a certain threshold is reached, the neuron creates electrical pulses, which propagate along the axon towards the axonal terminal.

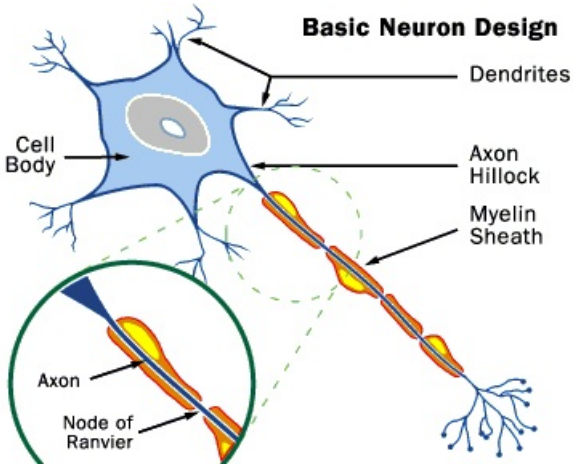


Figure 1.1: The basic design of a neuron consisting of a cell body, dendrites and a myelinated axon. The gap between two myelin sheaths are called the node of Ranvier. The specialised part of the soma that connects to the axon is called the axon Hillock. From [1]

Axons may extend from a few micrometers in the brain to more than a meter in the leg (the sciatic nerve). The axon begins at a region called axon hillock. Like the dendrites, the axon branches out like a tree, and these branches can communicate with dendrites of it's own neuron, or with dendrites of the neighbouring neurons. The end of the axon is called the axonal terminal. The terminal is the site where the axon comes in contact with other cells through the synapse, and passes information on to them.
To make the propagation of the signal along the axon faster, myelin segments are wrapped around the axon, forming the myelin sheaths. This insulation facilitates the conductance of the electrical signals. The myelin sheaths are supplied by glial cells, which in the central nerve system comprise oligodendrocytes. The sheaths are interrupted periodically and the short length where the axon is exposed is called the node of Ranvier. Only the axons in the white matter of the brain contain myelin sheaths, which makes the appearance white, hence the name. On the other hand the axons of the gray matter does not have myelin sheaths and hence the gray appearance.

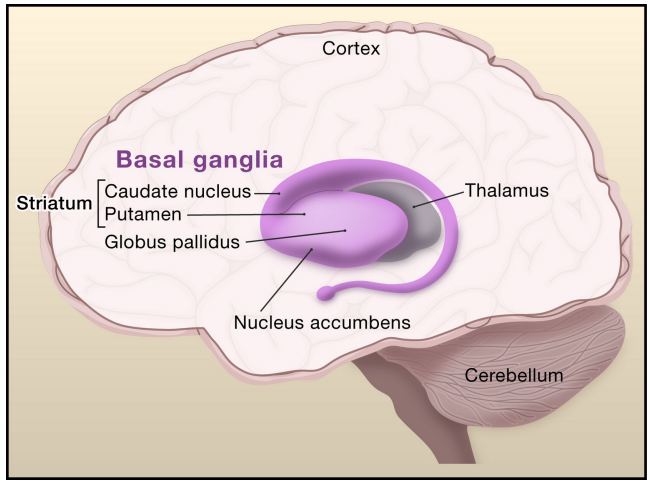


Figure 1.2: Transparent drawing of the basal ganglia. From [33]

1.2 The Basal Ganglia

The basal ganglia is a group of functionally related nuclei located bilaterally in the inferior cerebrum, diencephalon, and midbrain. The basal ganglia is composed of the striatum, the globus pallidus, the substantia nigra, and the subthalamic nucleus. The striatum can be further subdivided into the caudate nucleus and the putamen, see Figure 1.2 for illustration. The basal ganglia are important in the regulation of motor performance, especially at the beginning of a voluntary movement and at rest where it decreases the muscle tone. Disorders of the basal ganglia result in difficulty in initiating a movement, e.g., walking, or in uncontrolled movements like a slight shaking of the hands. Cerebral palsy and Parkinson's disease are examples of basal ganglia disorders.
The primary motor cortex is located in the posterior part of the frontal lobe. This region controls many voluntary movements, especially the fine motor movements of the hand. Both the basal ganglia and the primary motor cortex affect movements of the contralateral body. This is expressed by motion of the left hand activates the right motor cortex and vice versa.

1.3 MRI Physics

The body consists of about 70% water. The protons of the body are like small magnets and these are aligned along the magnetic field, B_0, when placed in the magnetic resonance (MRI) scanner. The protons are not fully aligned due to molecular tumbling, thermal motion and nuclear actions, but the equilibrium magnetization is along B_0. The protons precess in a cone around the magnetic field. This differs from a compass needle, which will oscillate in a plane around north. This difference is due to the rotation of the protons.
In a MRI scanner the body is placed in a strong magnetic field which aligns the protons along the magnetic field. Radio waves perturbing the equilibrium magnetization are applied and when the radio waves are turned off the precession of the magnetic dipole can be measured. The radio waves have to have the right frequency to make resonance and hence to have an effect. The frequency that changes the angle from equilibrium magnetization is called the Larmor frequency and is given by $f = \gamma B_0$, where γ is the gyromagnetic ratio. Radio waves are applied until a 90 degree rotation has occurred. When the system returns to thermal equilibrium radio waves are emitted from the protons and these can be detected by the coil in the scanner. This routine is continued when the magnetization is back in equilibrium. The period where the protons are vibrating before they are back in equilibrium is called relaxation and the relaxation-time is the time it takes the protons to get back in equilibrium. The relaxation occurs in two different time scales, the magnetization perpendicular to the magnetic field (the transversal magnetization) and the magnetization along the magnetic field (the longitudinal magnetization). The transversal magnetization decreases exponentially on a time scale T_2: $S \propto \exp(-TE/T_2)$, where TE is the echo time. The echo time is the time from the magnetization is in the transversal plane till we start measuring the signal.
Due to field inhomogeneities (variations in B_0) the protons will, after a while, point in all directions transversally, even if they were aligned to some extent after excitation. This process is called dephasing. As the measured signal is proportional to the transversal net magnetization, inhomogeneity causes loss of signal. The larger the field inhomogeneity the faster the dephasing. The time it takes to dephase is denoted T_2^*. Field inhomogeneity can, i.a., be a result of variations in magnetic properties on a microscopic scale, e.g., around capillaries. T_2^* is the most important relaxation time when understanding contrast in functional magnetic resonance images. The longitudinal magnetization, M_z approaches equilibrium on a time scale T_1: $M_z \propto 1 - \exp(t - T_1)$. The signal measured by the MRI scanner is given by Equation (1.1).

$$S = PD(1 - \exp(-TR/T_1))\exp(-TE/T_2^*) \tag{1.1}$$

Different types of tissue have different relaxation times. Several factors influence the signal, such as proton density (water content), local nuclear environment, flow, perfusion, diffusion, neuronal activation, metabolic properties, etc.. Solids have a short T_2 and a long T_1, while liquids have both a long T_1 and a long T_2. When the contrast in an image primarily comes from T_1-differences the image is said to be T_1-weighted, while it is T_2-weighted when the contrast mostly comes from T_2-differences.[24]

1.4 Functional Magnetic Resonance Imaging and the BOLD-signal

Functional magnetic resonance imaging (fMRI) measures brain activity by detecting changes in the blood flow. As neurons are activated they need more glucose, which results in an increased blood flow to the activated neurons. The changes in the hemodynamic response, which is measured during fMRI, is a measure of the activation of the neurons. Hence, fMRI shows the brain regions activated when a specific task is performed.
During synaptic activity vasodilators are released. This makes the blood flow increase to a larger extend than necessary to provide enough oxygen and glucose for the increased energy production. Blood flow and glucose consumption increases during activation while oxygen consumption only increases to a minor degree[18]. This uncoupling between blood flow and oxygen consumption leads to an increase in the oxy- to deoxyhaemoglobin concentration in the activated regions.
When using the blood-oxygenation level dependent (BOLD) signal we take advantage of the difference in magnetization in oxygen-rich (zero magnetic moment) and oxygen-poor (magnetic moment) blood. The activated oxygen-rich areas result in a longer relaxation time, T_2^*, which increases the intensity of the MRI signal [34].

1.5 Diffusion

Molecules in a solution are all subject to random movements, called Brownian motion. This motion can be illustrated by a drop of ink in a glass of water. When the ink is dropped into the water it starts to diffuse due to the Brownian motion. After some time the colour of the water becomes homogeneous, because

the surface of the glass restricts further diffusion of the ink. The concentration of the ink reaches a state of equilibrium and the net diffusion is zero, but it is still subject to Brownian motion. The molecules are subject to free diffusion until they reach the surface of the glass which restrict their motion. In free diffusion the mean diffusivity is isotropic, meaning that the molecules are moving an equal distance in all directions. When the molecules are subject to restrictions, the diffusion becomes anisotropic, meaning that the molecules move a longer distance in one direction than in another. Isotropic diffusion can be illustrated as a ball, while anisotropic diffusion can be seen as an ellipsoid.

1.6 Diffusion Weighted Magnetic Resonance Imaging

In diffusion weighted MRI (DWI) the movement of molecules, mostly water, is measured. The water molecules are subject to Brownian motion and the displacement of the molecules can be measured using DWI.
During each $TE/2$ period diffusion weighted gradients are applied. The gradient applied before the refocusing pulse induces a phase shift of the proton spins dependent upon their location. For a short time period the molecules are allowed to diffuse. After the refocusing pulse diffusion weighted gradients are applied again to rephase the proton spins. The displacement of the molecules can be found from the loss of phase. If the molecules have not moved, the phase shifts due to the two gradient pulses will cancel out[7]. The diffusion weighted spin echo sequence is illutrated in Figure 1.3.
The applied gradient is proportional to the displacement in the direction of the gradient. To measure the displacement in more directions, diffusion weighted gradients are applied in 61 directions.

The diffusion weighted signal can be expressed by Equation (1.2)[35].

$$S = PD(1 - \exp(-TR/T_1)) \exp(-TE/T_2^*) \exp(-bD) \tag{1.2}$$

where PD is the proton density, representing the water concentration, TR and TE are the repetition time and the echo time, respectively, T_1 and T_2^* are the relaxation times. D is the diffusion coefficient, representing the Brownian motion

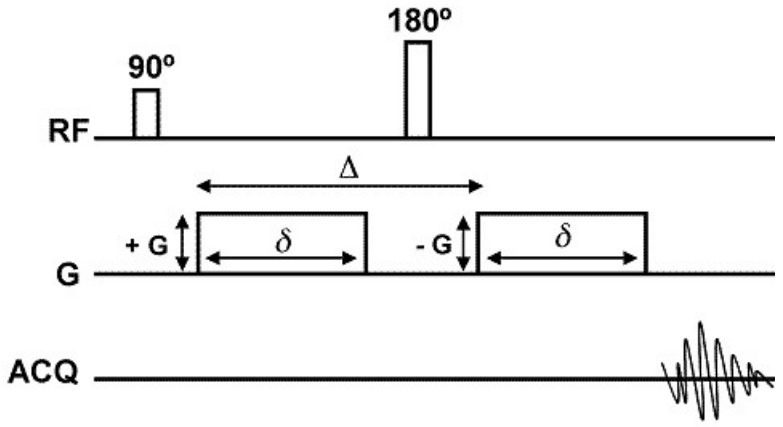


Figure 1.3: Schematic diagram of the diffusion weighted spin echo sequence. The top row shows the radio frequency (RF) pulses. The second row shows the diffusion weighted gradients with strength G. Δ is the time between the gradients and δ is the duration of the gradient. The first gradient shifts the phase of the spins, while the second gradient rephases the spins. The bottom row shows the measured echo signal.[4]

of water molecules and b is the diffusion weighting factor given by Equation (1.3).

$$b = \gamma^2 G^2 \delta^2 \left(\Delta - \frac{\delta}{3} \right) \tag{1.3}$$

where γ is the gyromagnetic ratio, G is the gradient and δ is the duration of the gradient.
By applying two gradients with different diffusion weighting, e.g., b_1 and b_2 the diffusion coefficient D can be obtained by dividing the two signals obtained by (1.2) as given by Equation (1.4).

$$D = -\ln \left(\frac{S_2}{S_1} \right) \frac{1}{(b_2 - b_1)} \tag{1.4}$$

By solving Equation (1.4) in each voxel we can obtain an apparent diffusion coefficient (ADC) map, where the intensity in each voxel is proportional to the extend of diffusion in that voxel. Bright regions in an ADC map express

fast diffusion, while dark regions express slow diffusion. It is called the "apparent" diffusion coefficient because the measured diffusion coefficient may not be the real diffusion coefficient. To accurately find the direction with greatest ADC it is necessary to measure the diffusion in a huge number of directions. This is not practical, so to simplify the problem Diffusion Tensor Imaging was introduced[35].

1.7 Diffusion Tensor Imaging

Diffusion Tensor Imaging (DTI) was introduced in the 1990s by Basser et al. [9]. DTI takes advantage of the fact that the diffusion of water in the brain is anisotropic.
As described earlier, the axons of the white matter in the brain have a myelin coating. Among other reasons for anisotropy this coating makes them more permeable in the direction parallel to the fibres than in the perpendicular direction. As a result, water diffuse faster along the fibres than in the direction perpendicular to the fibres.
In DTI the measurements of the ADC along different axes are fitted to a 3D ellipsoid. The ellipsoid represents the average diffusion distance in each direction. The length and the orientation of the axes of the ellipsoid are the eigenvalues and the eigenvectors, respectively. To calculate the ellipsoid, ADC measurements along six axes only are needed. In DWI the ADC is measured in more than six directions, and to convert the measurements to the six parameters, namely the diffusion along each direction, a 3×3 symmetric matrix, called a tensor, is used. The diffusion along each direction is estimated for each voxel. The direction of greatest diffusion is found by calculating the eigenvalues of the diffusion tensor.
A widely used measure of diffusion anisotropy is the fractional anisotropy (FA) [41], which is a convenient index because it is scaled from 0 (isotropic) to 1 (anisotropic) [35]. FA is given by

$$FA = \sqrt{\frac{1}{2}} \frac{\sqrt{((\lambda_1 - \lambda_2)^2 + (\lambda_2 - \lambda_3)^2 + (\lambda_3 - \lambda_1)^2)}}{\sqrt{\lambda_1^2 + \lambda_2^2 + \lambda_3^2}},$$

where λ_1, λ_2 and λ_3 are the eigenvalues of the diffusion tensor.

The diffusion tensor is also used to make tractography, either probabilistic or deterministic. The local fiber orientation is assumed to be along the direction of the longest axis of the diffusion tensor, corresponding to the direction of the

eigenvector with the biggest eigenvalue (v_1). To find a streamline a seed voxel is required as a starting point.
In deterministic tractography the streamline propagates from the seed voxel in the direction of v_1 in each voxel until it reaches a region with low anisotropy or a specified termination mask.
In probabilistic tractography the propagation of streamlines from the seed voxel is based on the connection probability between two points. The probability is defined as the integral over all possible trajectories connecting the two points. As this integral is too complex to be solved analytically instead a Monte Carlo (MC) approach can be used [40]. The integral is approximated by sampling a given number of streamlines between the two points resulting in a limited number of all possible trajectories between the two points. The connection probability is found by averaging this approximation.
A probability density function (pdf) describing the uncertainty of the fibre orientation in each voxel is needed to generate probabilistic streamlines. A probabilistic streamline starting from point A is created by randomly sampling an orientation from the local pdf and move a small step along this direction. This is repeated until a certain stopping criteria is reached. To produce probabilistic streamlines from a point A to any point in the brain a large number of streamlines starting in point A is generated and averaged.

1.8 Integration of Multimodal fMRI

The most investigated multimodal neuroimaging technique involving fMRI combines electroencephalography (EEG) and fMRI. EEG measures the electrical activity of the brain in the cerebral cortex, while fMRI measures the haemodynamic changes throughout the brain. Therefore EEG-fMRI measures both the neuronal and the haemodynamic activity, which are two important components of the neurovascular coupling mechanism.
EEG was first recorded in an MR environment in 1993 [28]. The first study, where EEG and fMRI was acquired simultaneously, was presented in 2000 by Goldman et al. [20].
Because of the difference in both spatial and temporal resolution between fMRI (high spatial resolution) and EEG (high temporal resolution) the combination was believed to have a high spatiotemporal resolution.
A challenge combining fMRI and EEG is that the electrodes used to measure the EEG signal are magnetic, and thus interference between the strong magnetic field in the MRI scanner and the EEG acquisition system arise. Another challenge is that even a slight motion of the EEG-electrodes within the MRI scanner can induce a significant electromotive force and induced electric cur-

rents can heat up the electrode leads and harm the patient.[23]
The EEG-fMRI technique is being used worldwide to evaluate normal brain function as well as brain function associated with recognized brain disorders and is a crucial diagnostic tool in epilepsy. The research in the EEG-fMRI technique has come so far that commercial devices are available.

In recent years there has been an increasing interest in combining DWI and fMRI. The reason is that it is anticipated that the integration of the two modalities can minimize the weaknesses of each of the modalities alone. It is the combination of DWI and fMRI which is explored in this project.
Over the last decade the studies combining DWI and fMRI have increased from three in 2000-2002 to seven in 2003-2005 for clinical applications, and from zero in 2000-2002 to nine in 2003-2005 outside clinical settings[42]. The first study where both DWI and fMRI data was collected was published by Werring et al. in 1998[49].
A number of studies are combining structural and functional data in different ways. The simplest way to combine the two modalities is by collecting both structural and functional images from the same subject and compare them in a qualitative manner [50]. In this type of approach no direct comparison of the two modalities is performed.
Another, and the most common, way to combine the two modalities is to superimpose the functional activation maps with maps of, e.g., fractional anisotropy, diffusivity, or reconstructed fibre tracts, derived from DTI. Differences and similarities are examined by visual inspection.
Statistical modelling methods to combine the two modalities, reduce the amount of information by only considering the most interesting patterns, e.g., by using the significant activations in the fMRI images as seeds for the tractography.[42].

Although there has been an increasing interest in combining DWI and fMRI in the past years, there is only a few studies where the analysis of the two modalities is fully integrated.
Bowman et al.[10] have recently published a work where DWI and fMRI are fully integrated. They calculate the functional distances between the fMRI signals, and the structural distances from the structural connectivity probabilities. The distance measurements are a measure of the dissimilarities between BOLD fMRI profiles in different regions, and the structural dissimilarities, respectively. By combining these two measures they can calculate the anatomically weighted functional distances. To cluster the data they use the average linkage method. They find that the proposed method produces more strongly autocorrelated functional connectivity networks, compared to conventional unweighted clustering.

1.9 Aim of the Thesis

The motivation for the project is that the understanding of the function and the connectivity of the brain that underlies normal and adaptive behaviours is the foundation for understanding the pathology and the pathophysiology of neurological diseases. An investigation of the connection between the functionality and the structural connectivity in healthy humans is a step in the direction of understanding the complex networks of the brain.

The aim of this project is to integrate the two MRI modalities, DWI and fMRI, to investigate the networks of the motor cortex and the striatum. These networks are investigated to understand the organization of the motor networks in the healthy brain, to later model the changes related to degenerative diseases as Parkinson's. To integrate the two modalities the stochastic blockmodel, the Infinite Relational Model (IRM), is used. The partition of the motor cortex and the striatum and the communication patterns between the motor cortex and the striatum, when the subjects are at rest and when they perform left and right finger tapping, respectively, are compared both for fMRI data alone and for combined fMRI and DWI data.

1.10 Content of the Thesis

The thesis comprises the following chapters. In Chapter 2 an introduction to community detection and graph theory is given and other measures used throughout the thesis are presented. Furthermore the IRM is described and the influence of the hyperparameters in the model are investigated. At last the predictability and the reproducibility of the IRM are compared to three other stochastic block models. In Chapter 3 the datasets used in the analysis are described and a description of how the data is processed is given. Chapter 4 contains the results and a discussion of the results is given in Chapter 5. Chapter 6 draws conclusions over the project and offers future perspectives.

Chapter 2

Methods and Community Detection Review

Structure exists everywhere, both in nature and in man-made networks. Commonly known networks are the World Wide Web, airline connections, epidemic spread, social networks like Facebook, and connections between neurons in the brain, just to mention a few. All these networks can be modelled using graphical models and their real world "graph like" structures can be identified using community detection.

2.1 Graph Theory

A graph model consists of nodes (also called vertices) connected by edges (also called links). Each node represents a random variable or a group of random variables, and each link represents a connection between two nodes. The order of a graph is given by the number of nodes, and the size of a graph is given by the number of links.

A graph can be described by an adjacency matrix, A, where the entities corre-

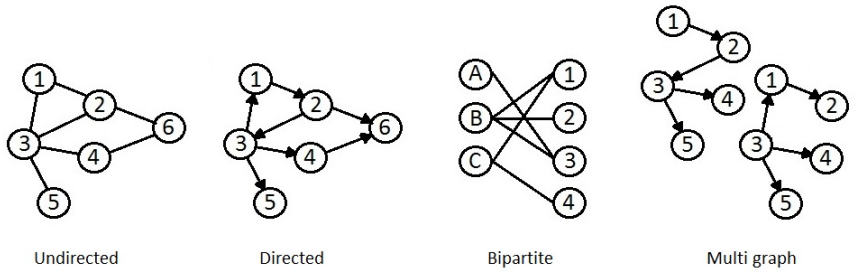


Figure 2.1: Illustration of different types of graphs. From left: an undirected graph, a directed graph, a bipartite graph and a multi graph. Based on [3].

spond to the connections between the nodes. If there is a link from node i to node j, the element $A(i, j) = 1$ and if no link exists between the two $A(i, j) = 0$. The graph can also be weighted, resulting in a non-binary adjacency matrix. In such case, the element $A(i, j)$ will take on the strength or number of connections between i and j.
A graph can be directed, undirected, bipartite or a multigraph, for illustration see Figure 2.1. In an undirected graph the adjacency matrix is symmetric, while this is not the case when the graph is directed. A bipartite network is a network with two different types of nodes and links that only connect nodes of different types. A multi graph consists of multiple graphs with the same set of nodes, but with different edge labels.

2.2 Community Detection

In large random graphs the node degrees are distributed according to the normal distribution, but many real networks are scale-free, meaning that the degree distribution follows a power-law. In 1999 Barabási and Albert presented a model that generates scale-free networks[8]. It incorporates two important concepts, which exist widely in real networks: Growth and preferential attachment. Preferential attachment is a "rich-gets-richer" phenomenon, meaning that nodes with many links are more likely to get even more links than nodes with few links, as the graph grows.
Because of the preferential attachment real networks are inhomogeneous, and have a high level of order, and organization. These inhomogeneities are expressed by groups of nodes which are densely connected with the other nodes in the group, but only sparsely connected to nodes outside the group. Such groups

are called communities.
A reason to detect communities in a graph is that it allows for classification of the nodes within the community. In context with the brain, voxels in the same cluster will probably be related either functionally or structurally.
One of the challenges finding structures in networks are validation of the clustering. How can we know if it is the real communities we find? It is suggested to generate synthetic data and test if the known structure can be detected. Another way to evaluate the inference is by link prediction. If some of the links in the graph is treated as missing it can be seen how well the model predicts the structure[21].

For a simple illustration of how a network can consist of communities, Zachary's Karate Club dataset[53] is used as an example. The dataset consists of 34 nodes, corresponding to the members of the club, and 78 undirected connections describing the friendships of the members outside the club. There are 60 members of the club, but only 34 of the members have friends within the club, leaving the other members as disconnected nodes in the graph. Only the 34 members with friends within the club are included in the analysis.
The members are forming two groups (communities), because the instructor and the president disagree whether or not the instructor should receive a pay rise. In the original paper Zachary predicts the two parts the club will split into. Figure 2.2 visualizes Zachary's Karate Club dataset. Nodes with the same color belong to the same cluster. Node 1 represents the instructor and node 34 represents the president.

2.3 K-means Clustering

A simple way to identifying groups, or clusters, in a dataset is by K-means clustering. The goal is to partition the dataset $\mathbf{x}_1, ..., \mathbf{x}_N$ into K clusters in such a way that the sum of squares of the distances from each point to the center of its cluster, $\boldsymbol{\mu}_k$, is minimized.
The assignment of each data point is described by a set of variables $r_{nk} \in \{0, 1\}$, where $k = 1, ..., K$ is the cluster the data point $\mathbf{x}_n$ is assigned to. If a data point is assigned to cluster k then $r_{nk} = 1$, and $r_{nj} = 0$ for $j \neq k$.
The sum of squares of the distances from each point to the center of its cluster

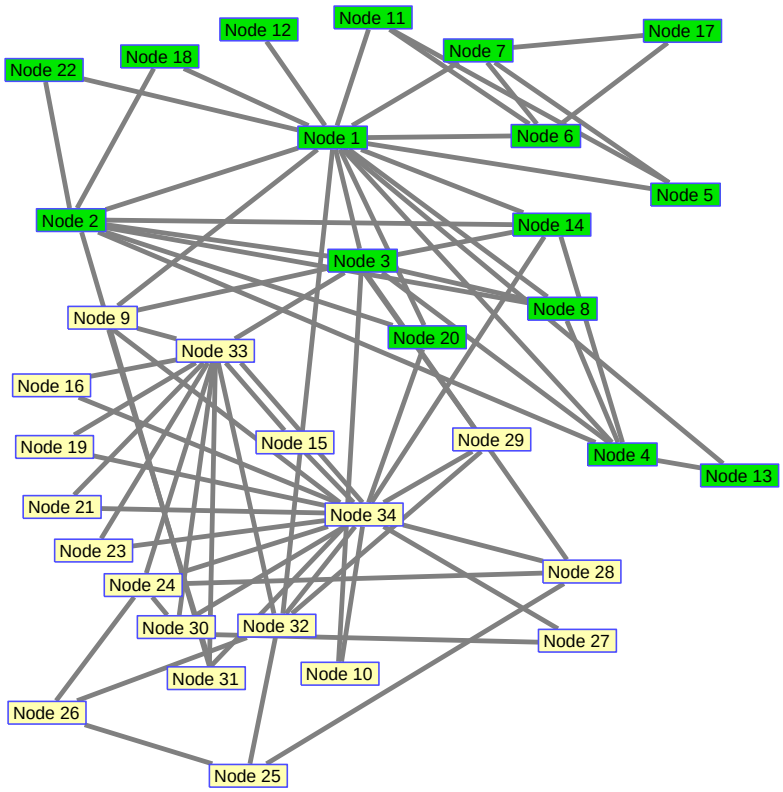


Figure 2.2: The original partition of Zachary's Karate Club dataset [53].

is called a distortion measure, and is given by Equation (2.1).

$$J = \sum_{n=1}^{N} \sum_{k=1}^{K} r_{nk} \|\mathbf{x}_n - \boldsymbol{\mu}_k\|^2 \tag{2.1}$$

To minimize Equation (2.1), we need to optimize $\boldsymbol{\mu}_k$ and r_{nk}. This is done through an iterative procedure, where each iteration consists of two steps. First we choose some initial values for $\boldsymbol{\mu}_k$. In the first step of each iteration, J is minimized with respect to r_{nk}, keeping $\boldsymbol{\mu}_k$ fixed, and in the second step J is minimized with respect to $\boldsymbol{\mu}_k$, keeping r_{nk} fixed. These two steps are repeated until a certain stopping criterion is reached.
J is a linear function of r_{nk} and the terms involving different ns are independent. We can optimize for each n seperately by choosing r_{nk} to be 1 for the value of k which gives the minimum value of $\|\mathbf{x}_n - \boldsymbol{\mu}_k\|^2$. The optimization of r_{nk} can be expressed as given by Equation (2.2).

$$r_{nk} = \begin{cases} 1 & \text{if } k = argmin_j \|\mathbf{x}_n - \boldsymbol{\mu}_j\|^2 \\ 0 & \text{otherwise} \end{cases} \tag{2.2}$$

When optimizing with respect to $\boldsymbol{\mu}_k$, r_{nk} is held fixed. J is a quadratic function of $\boldsymbol{\mu}_k$, and can be minimized by setting the derivative with respect to $\boldsymbol{\mu}_k$ to zero, and solve for $\boldsymbol{\mu}_k$

$$2 \sum_{n=1}^{N} r_{nk}(\mathbf{x}_n - \boldsymbol{\mu}_k) = 0 \Leftrightarrow \boldsymbol{\mu}_k = \frac{\sum_n r_{nk} \mathbf{x}_n}{\sum_n r_{nk}} \tag{2.3}$$

The denominator in Equation (2.3) is equal to the number of points assigned to cluster k, so J is minimized with respect to $\boldsymbol{\mu}_k$ by setting $\boldsymbol{\mu}_k$ equal to the mean of all the data points, $\mathbf{x}_n$, assigned to the cluster k.

One of the main drawbacks of K-means clustering is that you need to specify the number of clusters in the dataset. Another drawback is that a data point is always assigned to the cluster with the nearest center, even if the data point lies roughly midway between two cluster centres. This is despite the nearest cluster perhaps is not the most appropriate assignment.

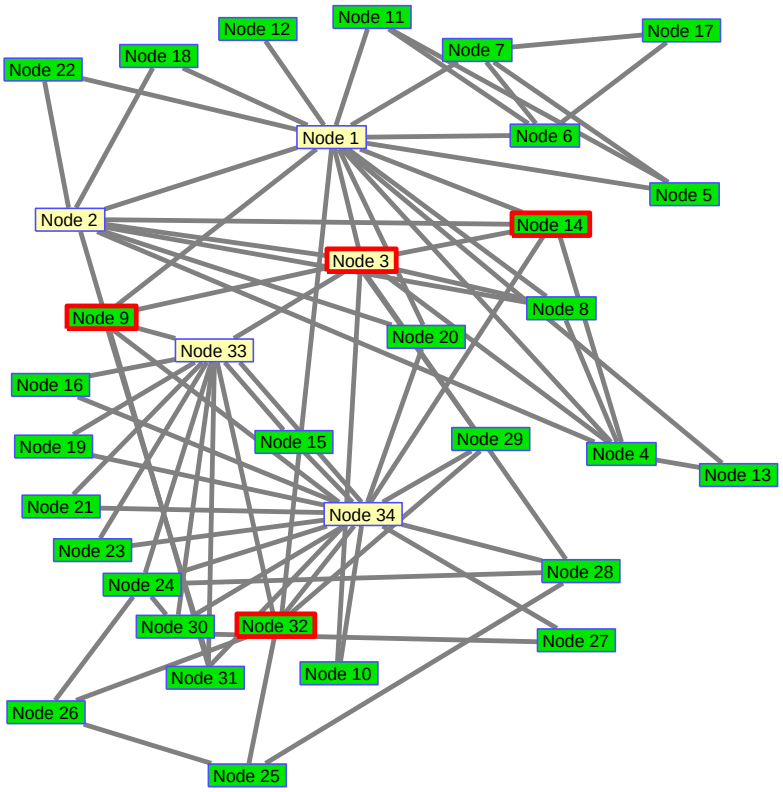


Figure 2.3: The partition of Zachary's Karate Club dataset using the K-means algorithm and the euclidean distance. The number of cluster $K = 2$ and the algorithm is initialized by assigning each node to one of the two clusters randomly. Nodes with red edges denote Hubs[11].

The Zachary's Karate Club dataset is used to test the K-means algorithm. To find the distance between the nodes, the euclidean distance is calculated between each pair of rows in the adjacency matrix. The number of clusters, K, is set to two, because the original partition contains two clusters. To initiate the algorithm each node is assigned to one of the two clusters randomly. The final clustering is illustrated in Figure 2.3. The result of the K-means algorithm is compared with the original partition of the graph (Figure 2.2). It can be seen that the K-means algorithm fails to find the original partition. Nodes with red edges denote Hubs, which are nodes connecting multiple clusters with association to any of them less than or equal to 50%[11].

2.4 The Relational Model

The main difference between data clustering and community detection is that communities in graphs are defined by a dense connection within the community, and only a sparse connection to nodes outside the community. On the other hand data clustering communities are sets of points close to each other defined by the distance, or a measure of similarity, between each pair of points.

The Relational model was presented by Codd in 1969[12] and 1970[13]. He presented a way to organise data into two-dimensional tables, called relations. Originally the relational model was developed for databases, and presented a way to efficiently store, access, and modify information. The relational model helps us design these relation tables, which makes it easy to efficiently store, access, and modify information.
A relation table contains information about the relations in the dataset. The columns of the table are called attributes, and the rows are called tuples. An important part of the design process is to select the attributes (properties of the objects) that can be kept in a table without redundancy, meaning that two rows contains the same information.
The relational model uses keys to order data, or relate data to other relations. The primary key is used to uniquely identify each column in the table. Foreign keys relate data in one relation to the primary key of another relation.
Analysis of relational structures has its primary focus on the pattern of relations between the actors (nodes) involved, but it is strongly affected by the selected attributes.[38]

2.5 The Blockmodel

A way to get a simpler description of a graph is by blockmodeling. In blockmodeling the goal is to decompose a graph into groups of nodes with similar properties. Nodes are usually grouped in clusters of equivalence [17]. There is two kinds of equivalence, structural and regular. Two nodes are said to be structurally equivalent if they have identical relationships to all the other nodes. Regular equivalence means that two nodes have similar relationships to the nodes of the other groups. Nodes which are structurally equivalent are also regular equivalent, but the inverse is not true.

Data studied by relational analysis consists of R relational variables measured

on g nodes, where

$$A_{ijr} = \begin{cases} c & \text{if node } i \text{ relates to node } j \text{ at level c on variable } r,\ i \neq j; \\ 0 & \text{otherwise.} \end{cases}$$

These variables are collected into R sets of $(g \times g)$ matrices $A_1, A_2, ..., A_3$, called sociomatrices or adjacency matrices. The standard approach is to permute the rows and the columns of all the adjacency matrices simultaneously to reveal patterns with respect to entries. This is done so that nodes, which are approximately equivalent, are assigned to the same block, meaning that nodes which are heavily interconnected, should form a block or a community [6]. The optimal partition is when all submatrices of the adjacency matrix only contains either zeros or ones. In real networks the nodes are seldom structurally equivalent, resulting in submatrices containing both zeros and ones. This is a deterministic approach and relies on algorithms to find the optimal partition.

2.6 The Stochastic Blockmodel

While the blockmodel is an algorithmic approach, the stochastic blockmodel is a model-based approach. It involves the discovery of the block structure as a part of the model search strategy. Both of the approaches are searching for the optimal partition of the nodes, such that nodes which are heavily connected form a block or a community. The nodes are rearranged to display the blocks in the diagonal of the adjacency matrix, which is representing the network. The connections between the nodes in different blocks appear in much sparser off-diagonal blocks. The difference between the algorithmic and the stochastic approach is that the stochastic approach searches for a partition which maximizes a statistical criterion linked to the model, such as a likelihood function. On the other hand most algorithmic approaches search to maximize an ad hoc criteria related to the "density" of links within and between blocks. The first model-based approach was presented by Holland et al. in 1983[27], but the first stochastic blockmodel involving the discovery of the block structure as a part of the model search strategy was presented by Nowicki and Snijders in 1997[45]. The statistical or stochastic blockmodels are also known as detection of community structure in the statistical physics and computer science litterature.

In the simplest stochastic blockmodel, each of n nodes is assigned to one of the K blocks, groups, or communities. Undirected edges are placed independently between pairs of nodes with probabilities that are a function only of the group

memberships of the nodes. If we denote by g_i the group to which node i belongs, then we can define a $K \times K$ matrix of probabilities, such that the matrix element (g_i, g_j) is the probability of an edge between nodes i and j. A stochastic blockmodel need two sets of parameters, the $K \times K$ matrix containing the average probability that nodes in block B_1 have connections to nodes in block B_2, and a mapping function between nodes and blocks [19]. One limitation of the stochastic blockmodel is the number of clusters has to be specified. To overcome this problem a stochastic blockmodel with an infinite number of clusters was presented by Kemp et al. [31].

2.7 The Infinite Relational Model

The Infinite Relational Model (IRM) [31, 52] can discover the structure in a relational dataset that appears quite different on the surface. One of the advantages of the IRM is that you do not need to know the number of clusters in advance. The IRM is able to adapt to the data by changing the number of clusters, as more data is provided, hence the term infinite.

Since the IRM can have an infinite number of clusters, it is important to have a prior on the number of clusters, which introduce no more clusters than necessary to explain the data. Therefore the Chinese Restaurant Process (CRP) is used as a prior.
The CRP is a discrete-time stochastic process. Imaging a Chinese restaurant with an infinite number of tables. When the first costumer arrives all tables are unoccupied, and he starts a new table. When the second costumer arrives, he can choose to be seated at the same table as the first costumer, or to start a new table. This process continues till all costumers are seated. The probability of sitting at an existing table or starting a new table, respectively, is:

$$P(z_i = k | z_1, ..., z_{i-1}) = \begin{cases} \dfrac{n_k}{i - 1 + \alpha} & n_k > 0 \\ \dfrac{\alpha}{i - 1 + \alpha} & k \textit{ is a new cluster} \end{cases}$$

where z_i is the i^{th} costumer, n_k is the number of costumers seated at table k, and α is a parameter, affecting the number of clusters generated. The customers and the tables correspond to the nodes and the clusters, respectively, in

the terms of graph theory.
The distribution on z induced by the CRP is exchangeble, meaning that the probability of the resulting partition does not depend on the order in which nodes are assigned to clusters. The distribution only depends on the size of the clusters, and not on the order in which the nodes are assigned to the clusters.

The probability of observing a link between two nodes is π_{ij}, and $(1 - \pi_{ij})$ for observing a non-link. This corresponds to getting either heads or tails when a coin is tossed.
The generative model for generating relations from clusters is given by (2.4). It is assumed that the relations are binary-valued functions.

$$\begin{aligned} \mathbf{Z} &\sim \mathrm{CRP}(\alpha) \\ \boldsymbol{\eta}^{(n)}(a,b) &\sim \mathrm{Beta}(\boldsymbol{\beta}_{+}(a,b), \boldsymbol{\beta}_{-}(a,b)) \\ \mathbf{A}^{(n)}(i,j) &\sim \mathrm{Bernoulli}(\mathbf{z}_i^T \boldsymbol{\eta}^{(n)} \mathbf{z}_j), \end{aligned} \tag{2.4}$$

where $a, b \in \mathbb{N}$. $\mathbf{Z}$ is the clustering assignment matrix of size cluster $\times$ nodes and $\mathbf{z}_j$ is the jth row of $\mathbf{Z}$. $\boldsymbol{\eta}^{(n)}(a,b)$ specifies the probability that a link exists between cluster a and cluster b, and have the size cluster $\times$ cluster. $\boldsymbol{\beta}_{+}(a,b)$ is a two element vector of pseudo link counts within and between communities, and $\boldsymbol{\beta}_{-}(a,b)$ is a two element vector of pseudo non-link counts within and between communities. $\mathbf{A}^{(n)}(i,j)$ is the binary adjacency matrix. $\mathbf{A}^{(n)}(i,j) = 1$ if a link exists between node i and node j, else $\mathbf{A}^{(n)}(i,j) = 0$.

The parameter α in the CRP affects the number of clusters generated. If α takes on a great number, it is more likely that a new cluster is generated, which will result in many small clusters. On the other hand, if α takes on a small number, almost all nodes will be assigned to the same cluster, resulting in a few huge clusters. This is illustrated in Figure 2.4-2.6, where the value of α is varied while the values of the other hyperparameters are kept constant: $\boldsymbol{\beta}_{+}(a,b) = \boldsymbol{\beta}_{-}(a,b) = [1\ 1]$.

The size of the graph affects how well the structure is estimated and how well missing links and non-links are predicted. If the graph is small, the model does not have enough information to find the true structure neither to predict the missing links. If the graph contains a lot of nodes, plenty of information is available to detect the true structure and to predict links. The limitation of the

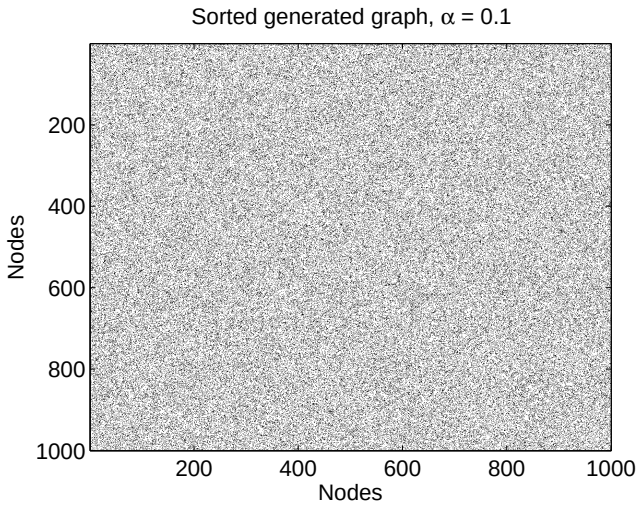


Figure 2.4: The sorted generated graph for $\alpha = 0.1$. $\boldsymbol{\beta}_+(a,b) = \boldsymbol{\beta}_-(a,b) = [1\ 1]$. The graph contains 1000 nodes. The black dots denote links.

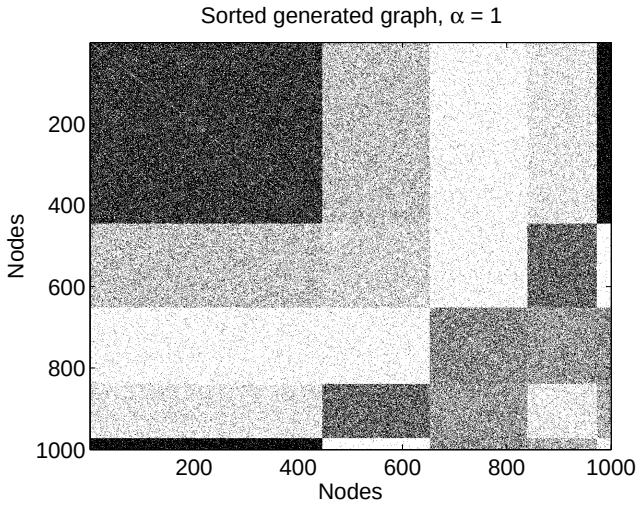


Figure 2.5: The sorted generated graph for $\alpha = 1$. $\boldsymbol{\beta}_+(a,b) = \boldsymbol{\beta}_-(a,b) = [1\ 1]$. The graph contains 1000 nodes. The black dots denote links.

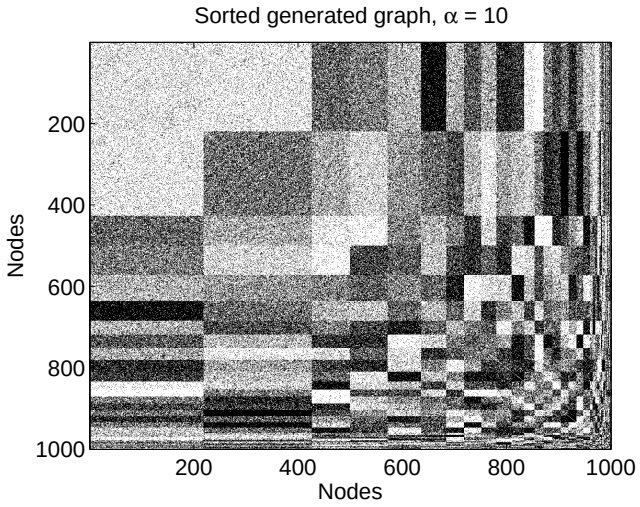


Figure 2.6: The sorted generated graph for $\alpha = 10$. $\boldsymbol{\beta}_+(a,b) = \boldsymbol{\beta}_-(a,b) = [1\ 1]$. The graph contains 1000 nodes. The black dots denote links.

size of the graph is the time consumption, when the graph is inferred.

If the prior on links within communities $\boldsymbol{\beta}_+(1)$ and the prior on non-links between communities $\boldsymbol{\beta}_-(2)$ take on a large number, and the prior on links between communities $\boldsymbol{\beta}_+(2)$ and the prior on non-links within communities $\boldsymbol{\beta}_-(1)$ take on a small number, the generative model will generate a graph with links in the diagonal only, see Figure 2.7 for illustration. On the other hand if the prior on links within communities $\boldsymbol{\beta}_+(1)$ and the prior on non-links between communities $\boldsymbol{\beta}_-(2)$ take on a small number, and the prior on links between communities $\boldsymbol{\beta}_+(2)$ and the prior on non-links within communities $\boldsymbol{\beta}_-(1)$ take on a large number, the generative model will generate a graph with links in the off-diagonal only, see Figure 2.8 for illustration. If the priors on links are uniform, $\boldsymbol{\beta}_+(a,b) = \boldsymbol{\beta}_-(a,b) = [1\ 1]$, there will be an equal number of links and non-links in the diagonal and the off-diagonal, see Figure 2.9.

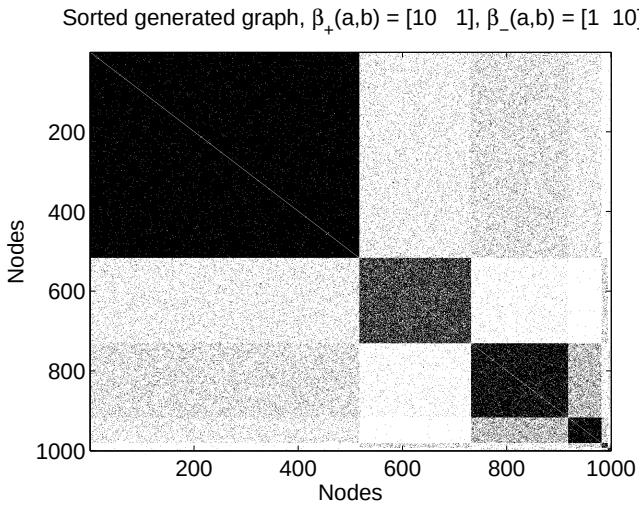


Figure 2.7: The sorted generated graph for $\boldsymbol{\beta}_+(a,b) = [10\ 1]$ and $\boldsymbol{\beta}_-(a,b) = [1\ 10]$ and $\alpha = 1$. The graph contains 1000 nodes. The black dots denote links. The generated graph contains almost only links in the diagonal.

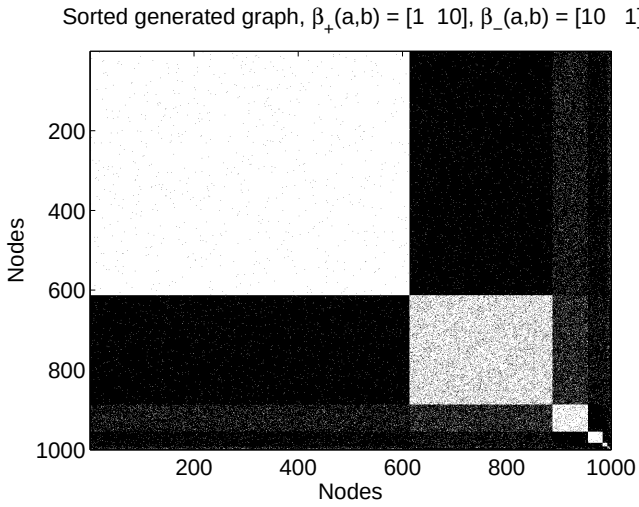


Figure 2.8: The sorted generated graph for $\boldsymbol{\beta}_+(a,b) = [1\ 10]$ and $\boldsymbol{\beta}_-(a,b) = [10\ 1]$ and $\alpha = 1$. The graph contains 1000 nodes. The black dots denote links. The generated graph contains almost only links in the off-diagonal.

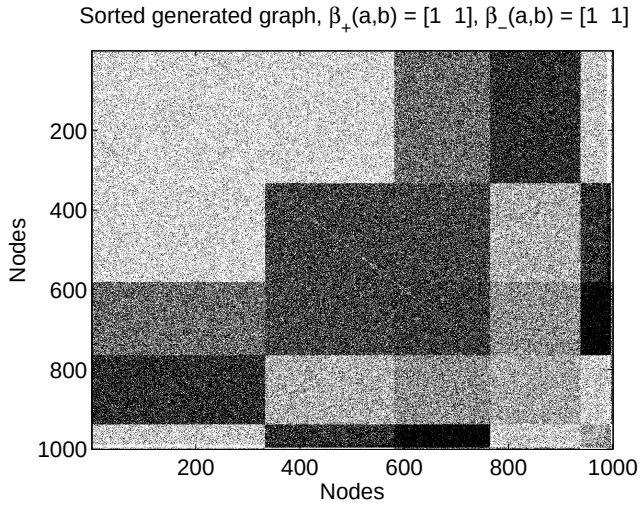


Figure 2.9: The sorted generated graph for $\boldsymbol{\beta}_+(a,b) = [1\ 1]$ and $\boldsymbol{\beta}_-(a,b) = [1\ 1]$ and $\alpha = 1$. The graph contains 1000 nodes. The black dots denote links. The generated graph contains almost and equal number of links and non-links in the diagonal and the off-diagonal.

2.8 Inference of the Infinite Relational Model

To infer the IRM, we use the posterior likelihood. This can be found from the joint likelihood. Lets say we have N adjacency matrices $\mathbf{A}^{(1)}, ..., \mathbf{A}^{(N)}$. Each adjacency matrix $\mathbf{A}^{(n)}$ have a corresponding group link probability matrix $\boldsymbol{\eta}^{(n)}$, while the assignment matrix $\mathbf{Z}$ is shared across the adjacency matrices. We assume that the graphs are independent over subjects. The joint likelihood for the IRM is

$$
\begin{aligned}
P(\mathbf{A}^{(1)}, ..., \mathbf{A}^{(N)}, \mathbf{Z}, \boldsymbol{\eta}^{(n)} \mid \alpha, \boldsymbol{\beta}_+, \boldsymbol{\beta}_-) = \\
\prod_n \prod_{ij} \mathrm{Bern}(\mathbf{A}^{(n)}(i,j) \mid \mathbf{z}_i^T \boldsymbol{\eta}^{(n)} \mathbf{z}_j) \cdot \\
\prod_{ab} \mathrm{Beta}(\boldsymbol{\eta}^{(n)}(a,b) \mid \boldsymbol{\beta}_+(a,b), \boldsymbol{\beta}_-(a,b)) \cdot \mathrm{CRP}(\mathbf{Z} \mid \alpha)
\end{aligned}
$$

If the posterior has the same functional form as the prior, the prior is conjugate to the likelihood function. As $\boldsymbol{\eta}^{(n)}$ is the conjugated prior, we can integrate $\boldsymbol{\eta}^{(n)}$ out analytically

$$
\begin{aligned}
&P(\mathbf{A}^{(1)}, ..., \mathbf{A}^{(N)}, \mathbf{Z} \mid \alpha, \boldsymbol{\beta}_+, \boldsymbol{\beta}_-) \\
&= \int \prod_n \prod_{ij} (\mathbf{z}_i^T \boldsymbol{\eta}^{(n)} \mathbf{z}_j)^{\mathbf{A}^{(n)}(i,j)} (1 - \mathbf{z}_i^T \boldsymbol{\eta}^{(n)} \mathbf{z}_j)^{1-\mathbf{A}^{(n)}(i,j)} \cdot \\
&\quad \prod_{ab} \frac{\Gamma(\boldsymbol{\beta}_+(a,b) + \boldsymbol{\beta}_-(a,b))}{\Gamma(\boldsymbol{\beta}_+(a,b))\Gamma(\boldsymbol{\beta}_-(a,b))} (\boldsymbol{\eta}^{(n)})^{\boldsymbol{\beta}_+(a,b)-1} (1 - \boldsymbol{\eta}^{(n)})^{\boldsymbol{\beta}_-(a,b)-1} \cdot \mathrm{CRP}(\mathbf{Z} \mid \alpha) d\boldsymbol{\eta}^{(n)} \\
&= \mathrm{CRP}(\mathbf{Z} \mid \alpha) \int \prod_n \prod_{ab} \prod_{ij} \boldsymbol{\eta}^{(n)}(a,b)^{\mathbf{A}^{(n)}(i,j)\mathbf{z}_{ai}\mathbf{z}_{bj}} (1 - \boldsymbol{\eta}^{(n)}(a,b))^{(1-\mathbf{A}^{(n)}(i,j))\mathbf{z}_{ai}\mathbf{z}_{bj}} \cdot \\
&\quad \frac{\Gamma(\boldsymbol{\beta}_+(a,b) + \boldsymbol{\beta}_-(a,b))}{\Gamma(\boldsymbol{\beta}_+(a,b))\Gamma(\boldsymbol{\beta}_-(a,b))} \boldsymbol{\eta}^{(n)}(a,b)^{\boldsymbol{\beta}_+(a,b)-1} (1 - \boldsymbol{\eta}^{(n)}(a,b))^{\boldsymbol{\beta}_-(a,b)-1} d\boldsymbol{\eta}^{(n)} \\
&= \mathrm{CRP}(\mathbf{Z} \mid \alpha) \\
&\quad \int \prod_n \prod_{ab} \boldsymbol{\eta}^{(n)}(a,b)^{\sum_{ij} \mathbf{A}^{(n)}(i,j)\mathbf{z}_{ai}\mathbf{z}_{bj}} (1 - \boldsymbol{\eta}^{(n)}(a,b))^{\sum_{ij}((1-\mathbf{A}^{(n)}(i,j))\mathbf{z}_{ai}\mathbf{z}_{bj})} \cdot \\
&\quad \frac{\Gamma(\boldsymbol{\beta}_+(a,b) + \boldsymbol{\beta}_-(a,b))}{\Gamma(\boldsymbol{\beta}_+(a,b))\Gamma(\boldsymbol{\beta}_-(a,b))} \boldsymbol{\eta}^{(n)}(a,b)^{\boldsymbol{\beta}_+(a,b)-1} (1 - \boldsymbol{\eta}^{(n)}(a,b))^{\boldsymbol{\beta}_-(a,b)-1} d\boldsymbol{\eta}^{(n)}
\end{aligned}
$$

We use the definitions

$$\sum_{ij} \mathbf{A}^{(n)}(i,j)\mathbf{z}_{ai}\mathbf{z}_{bj} = M_+^{(n)}(a,b) \tag{2.5}$$

$$\sum_{ij} ((1 - \mathbf{A}^{(n)}(i,j))\mathbf{z}_{ai}\mathbf{z}_{bj}) = M_-^{(n)}(a,b) \tag{2.6}$$

and get

$$
\begin{aligned}
&P(\mathbf{A}^{(1)}, ..., \mathbf{A}^{(N)}, \mathbf{Z} \mid \alpha, \boldsymbol{\beta}_+, \boldsymbol{\beta}_-) \\
&= \mathrm{CRP}(\mathbf{Z} \mid \alpha) \int \prod_n \prod_{ab} \boldsymbol{\eta}^{(n)}(a,b)^{M_+^{(n)}(a,b)} (1 - \boldsymbol{\eta}^{(n)}(a,b))^{M_-^{(n)}(a,b)} \\
&\qquad \frac{\Gamma(\boldsymbol{\beta}_+(a,b) + \boldsymbol{\beta}_-(a,b))}{\Gamma(\boldsymbol{\beta}_+(a,b))\Gamma(\boldsymbol{\beta}_-(a,b))} \boldsymbol{\eta}^{(n)}(a,b)^{\boldsymbol{\beta}_+(a,b)-1} (1 - \boldsymbol{\eta}^{(n)}(a,b))^{\boldsymbol{\beta}_-(a,b)-1} d\boldsymbol{\eta}^{(n)} \\
&= \mathrm{CRP}(\mathbf{Z} \mid \alpha) \prod_n \prod_{ab} \frac{\Gamma(\boldsymbol{\beta}_+(a,b) + \boldsymbol{\beta}_-(a,b))}{\Gamma(\boldsymbol{\beta}_+(a,b))\Gamma(\boldsymbol{\beta}_-(a,b))} \\
&\qquad \int \boldsymbol{\eta}^{(n)}(a,b)^{M_+^{(n)}(a,b)+\boldsymbol{\beta}_+(a,b)-1} (1 - \boldsymbol{\eta}^{(n)}(a,b))^{M_-^{(n)}(a,b)+\boldsymbol{\beta}_-(a,b)-1} d\boldsymbol{\eta}^{(n)}
\end{aligned}
$$

We use that the integral of the probability density function is unity

$$\int \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\mu^{a-1}(1-\mu)^{b-1}d\mu = 1 \Leftrightarrow \int \mu^{a-1}(1-\mu)^{b-1}d\mu = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$$

and get the likelihood for the IRM marginalised with regard to $\boldsymbol{\eta}$

$$\begin{aligned}
&P(\mathbf{A}^{(1)}, ..., \mathbf{A}^{(N)}, \mathbf{Z} \mid \alpha, \boldsymbol{\beta}_+, \boldsymbol{\beta}_-) \\
&= \mathrm{CRP}(\mathbf{Z} \mid \alpha) \\
&\quad \prod_n \prod_{ab} \frac{\Gamma(\boldsymbol{\beta}_+(a,b) + \boldsymbol{\beta}_-(a,b))}{\Gamma(\boldsymbol{\beta}_+(a,b))\Gamma(\boldsymbol{\beta}_-(a,b))} \frac{\Gamma(M_+^{(n)}(a,b) + \boldsymbol{\beta}_+(a,b))\Gamma(M_-^{(n)}(a,b) + \boldsymbol{\beta}_-(a,b))}{\Gamma(M_+^{(n)}(a,b) + \boldsymbol{\beta}_+(a,b) + M_-^{(n)}(a,b) + \boldsymbol{\beta}_-(a,b))} \\
&= \mathrm{CRP}(\mathbf{Z} \mid \alpha) \prod_n \prod_{ab} \frac{\mathrm{Beta}(M_+^{(n)}(a,b) + \boldsymbol{\beta}_+(a,b), M_-^{(n)}(a,b) + \boldsymbol{\beta}_-(a,b))}{\mathrm{Beta}(\boldsymbol{\beta}_+(a,b), \boldsymbol{\beta}_-(a,b))} \\
&= \frac{\alpha^D \Gamma(\alpha) \prod_a \Gamma(n_a)}{\Gamma(J+\alpha)} \prod_n \prod_{ab} \frac{\mathrm{Beta}(M_+^{(n)}(a,b) + \boldsymbol{\beta}_+(a,b), M_-^{(n)}(a,b) + \boldsymbol{\beta}_-(a,b))}{\mathrm{Beta}(\boldsymbol{\beta}_+(a,b), \boldsymbol{\beta}_-(a,b))},
\end{aligned}$$

where $M_+^{(n)}(a,b)$ is the number of links and $M_-^{(n)}(a,b)$ is the number of non-links between groups a and b, D is the number of clusters and n_a is the number of of nodes assigned to the cluster D.
Using Bayes' theorem, the posterior likelihood of $\mathbf{z}_{ai}$ can be found as

$$P(\mathbf{z}_{ai} = 1 \mid \mathbf{A}^{(1)}, ..., \mathbf{A}^{(N)}, \mathbf{Z}_{\setminus i}, \alpha, \boldsymbol{\beta}_+, \boldsymbol{\beta}_-) = \frac{P(\mathbf{A}^{(n)}, \mathbf{Z}_{\setminus i}, \mathbf{z}_{ai} = 1 \mid \alpha, \boldsymbol{\beta}_+, \boldsymbol{\beta}_-)}{\sum_a P(\mathbf{A}^{(n)}, \mathbf{Z}_{\setminus i}, \mathbf{z}_{ai} = 1 \mid \alpha, \boldsymbol{\beta}_+, \boldsymbol{\beta}_-)} \tag{2.7}$$

The assignment matrix $\mathbf{Z}$ is inferred by Gibbs sampling, as proposed in [31], in combination with split-merge sampling [29, 31, 36], each column at a time from the posterior distribution in Equation (2.7).

The IRM is tested on Zachary's Karate Club dataset, as with the K-means algorithm, and the resulting partition is shown in Figure 2.10. The IRM used 200 iterations, and the partition was the same for ten different runs. The hyperparameters was $\alpha = 1$ and $\boldsymbol{\beta}_+(a,b) = \boldsymbol{\beta}_-(a,b) = [1\ 1]$. The IRM finds

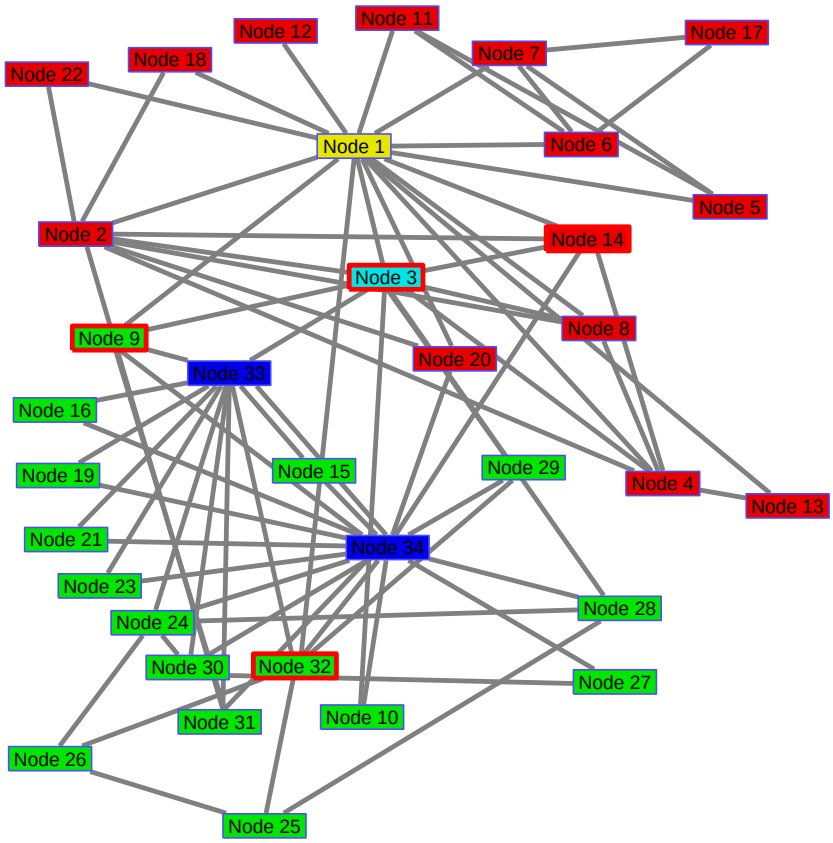


Figure 2.10: Illustration of the partition of Zachary's Karate Club dataset inferred with the IRM. Nodes with red edges (3, 9, 14, and 32) denote Hubs[11].

five clusters. Nodes 1 and 3 constitute their own clusters and node 33 and 34 constitute one cluster. However, the remaining nodes are partitioned as in the original partition (Figure 2.2). In the original partioning node 1 and 3 belong to one cluster (the red cluster in Figure 2.10), and node 33 and 34 belong to another cluster (the green cluster in Figure 2.10).

To investigate the influence of the hyperparameter α, Zachary's Karate Club dataset is inferred for different values of α in the interval $[10^{-13}\ 10^{7}]$. The other hyperparameters are held constant, $\boldsymbol{\beta}_{+}(a,b) = \boldsymbol{\beta}_{-}(a,b) = [1\ 1]$. The average number of clusters across 100 runs are plotted in a semi-logarithmic plot, see

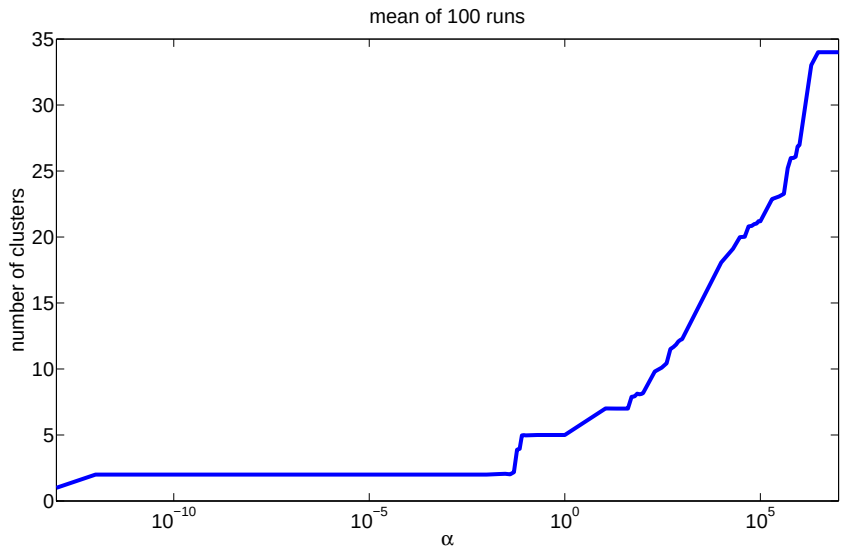


Figure 2.11: The average number of clusters across 100 runs for different values of α when the Zachary's Karate Club dataset is inferred with the IRM.

Figure 2.11. When α takes on extreme values the dataset is partitioned in, respectively, 1 ($\alpha = 10^{-13}$) and 34 ($\alpha = 3 \cdot 10^{6}$) clusters. However, in the range $\alpha = [10^{-12}\ 0.05]$ the number of clusters is constant. The two clusters found by the IRM are not the same as in the original partition, as the IRM assigns the nodes with many links to the same cluster and all the remaining nodes to another cluster. The partition of Zachary's Karate Club dataset, using $\alpha = 10^{-5}$, is shown in Figure A.1 in Appendix A.

2.9 Sampling Methods

Suppose we have a joint density $f(x, y_1, ...y_p)$ and want to obtain some characteristics, such as mean and variance, of the marginal distribution. A common way to do this is by calculating $f(x)$ by integration

$$f(x) = \int ... \int f(x, y_1, ...y_p) dy_1 ... dy_p$$

However, in many cases it is very difficult to do the integration either analytically or numerically. In such cases Markov chain Monte Carlo (MCMC) techniques, like Gibbs sampling, provide a way to approximate the value of the integral.

2.9.1 Monte Carlo

Monte Carlo methods are algorithms that can be used to numerically approximate the value of an integral by performing simulations.
Suppose we have a function and want to estimate the integral. Pick n randomly distributed points $x_1, x_2, ..., x_n$ in the interval $[a, b]$. The average value of the function can be calculated as

$$\overline{f} = \frac{1}{n} \sum_{i=1}^{n} f(x_i)$$

and the integral can be approximated as

$$\int_a^b f(x)dx = (b-a)\overline{f}$$

The approximation becomes more accurate as the number of samples, n, increases.

2.9.2 Markov Chain

A Markov chain is a mathematical system that undergoes transitions from one state to another between a finite number of possible states. It is a random and memoryless process, meaning that the next state only depends on the current state.
A Markov chain is generating a sequence of random variables $x_\tau, x_{\tau+1}, x_{\tau+2}, ...$ with the Markov property: Given the present state, the future and the past states are independent.
The probability of moving from the current state, i, to the next state, j, is given by the transition probability p_{ij}, as given in Equation (2.9.2), or the process can remain in its current state with the transition probability p_{ii}.

$$p(x_{\tau+1} = i \mid x_\tau = j) = p_{ij}$$

An example of a Markov chain is the game Monopoly. It is a Markov chain because the next state of the board only depends on the current state of the board and the next roll of the dice. It doesn't depend on how things got to their current state. In contrast are games like blackjack or poker, whose next state is dependent on the past states. They are not Markov chains, because the player can gain advantages by remembering the cards, which have already been shown, and thereby know which cards there no longer are in the deck. Any game, whose moves are determined entirely by the roll of a dice is a Markov chain.

2.9.3 Metropolis-Hastings

The Metropolis-Hastings algorithm was developed by Metropolis et al. in 1953 and generalized by Hastings in 1970. The Metropolis-Hastings algorithm is a MCMC algorithm for obtaining a sequence of random samples from a multivariate probability distribution for which direct sampling is difficult. The sequence can be used to approximate the distribution. Suppose an invariant distribution, $p(x)$, and a proposal distribution, $q(x^* \mid x)$, are given. The Metropolis-Hastings algorithm follows the following steps

1. Initialize the algorithm by picking an initial value x_τ at random
2. Pick a candidate value x^* from the proposal distribution at random
3. Accept the candidate value $x_{\tau+1} = x^*$ with the acceptance probability $A(x_\tau, x^*) = min\left\{1, \frac{p(x^*)q(x_\tau \mid x^*)}{p(x_\tau)q(x^* \mid x_\tau)}\right\}$, otherwise it remains at $x_{\tau+1} = x_\tau$

Step two and three are repeated. As the number of iterations increases the histogram of the samples approximates the target distribution.

2.9.4 Gibbs Sampling

Gibbs sampling is a special case of the Metropolis-Hastings algorithm. Suppose a joint distribution is given, for example the posterior, that we want to sample from. The Gibbs sampler can be used to sample from the joint distribution, if

the full conditional distribution for each parameter is known. The full conditional distribution is the distribution of the parameter, conditional on the known information and all the other parameters.

There are two requirements to be satisfied, in order to the Gibbs sampling procedure samples from the correct distribution. The first requirement is that the joint distribution, $p(\mathbf{x})$, is an invariant of each of the Gibbs sampling steps individually and hence of the whole Markov chain and the second requirement is that it be ergodic.

The algorithm:

1 Initialize $\{x_i : i = 1, ..., M\}$

2 For $\tau = 1, ..., T$:

- Sample $x_1^{\tau+1} \sim p(x_1|x_2^{\tau}, x_3^{\tau}, ..., x_M^{\tau})$.
- Sample $x_2^{\tau+1} \sim p(x_2|x_1^{\tau+1}, x_3^{\tau}, ..., x_M^{\tau})$.
- $\vdots$
- Sample $x_j^{\tau+1} \sim p(x_j|x_1^{\tau+1}, ..., x_{j-1}^{\tau+1}, x_{j+1}^{\tau}, ..., x_M^{\tau})$.
- $\vdots$
- Sample $x_M^{\tau+1} \sim p(x_M|x_1^{\tau+1}, x_2^{\tau+1}, ..., x_{M-1}^{\tau+1})$.

To go from x_1^{τ} to $x_1^{\tau+1}$ we have to go through the iterations $x_j^{\tau}, j = 2, ..., M$. The iteration sequence forms a markov chain with a bunch of draws of $p(\mathbf{x})$ that are approximately from our posterior distribution. Monte Carlo integration can be done on the draws to get quantities of interest.

In Figures 2.12-2.15, it is illustrated how the Gibbs sampler works by a simple example. A 2D Gaussian distribution with mean, μ, and covariance, Σ, is used for sampling.

$$\mu = \begin{bmatrix} 0 & 0 \end{bmatrix} \tag{2.8}$$
$$\Sigma = \begin{bmatrix} 0.5 & 0.4 \\ 0.4 & 0.5 \end{bmatrix}$$

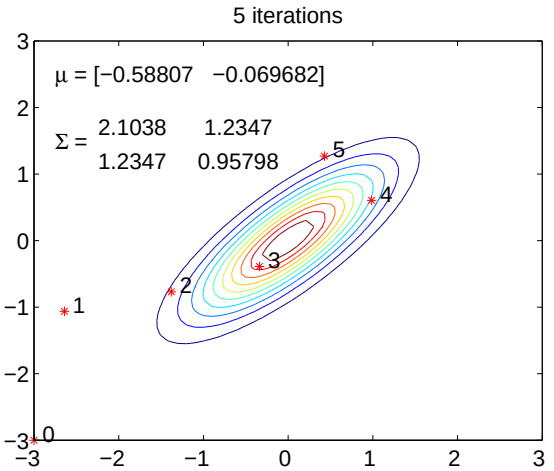


Figure 2.12: Illustration of Gibbs sampling from a 2D Gaussian distribution. Five iterations are made and the estimates of the mean and the covariance are printed in the figure.

In Figure 2.12 only five iterations are made, in Figure 2.13 10 iterations are made, in Figure 2.14 100 iterations are made, and in Figure 2.15 500 iterations are made. The mean, μ, and the covariance, Σ, of the samples are printed in each figure. As the number of samples increases, the mean and the covariance approach the true values given in (2.8). Figure 2.16 shows the log likelihood for 100 iterations averaged across 100 runs. For the first iterations the log likelihood is small, but around the tenth iteration it stabilizes. This is called the burn in period.

2.10 Mutual Information

To evaluate the reproducibility of the IRM, the Mutual Information is used. The relative entropy, or the Kullback-Leibler divergence, between the distributions $p(x)$ and $q(x)$ is a measure of the distance between the two distributions, and is given by

$$KL(p \parallel q) = \sum_{x \in X} p(x) \log \frac{p(x)}{q(x)}$$

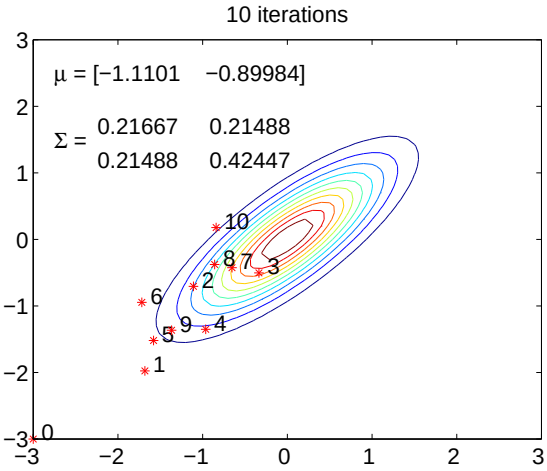


Figure 2.13: Illustration of Gibbs sampling from a 2D Gaussian distribution. Ten iterations are made and the estimates of the mean and the covariance are printed in the figure.

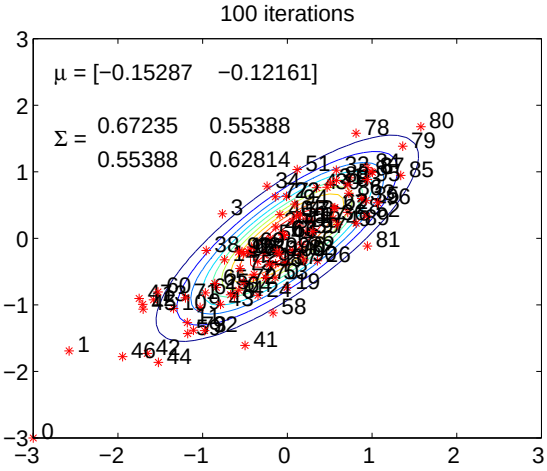


Figure 2.14: Illustration of Gibbs sampling from a 2D Gaussian distribution. 100 iterations are made and the estimates of the mean and the covariance are printed in the figure.

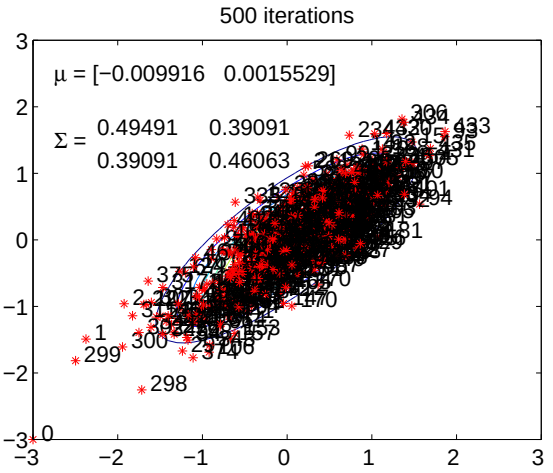


Figure 2.15: Illustration of Gibbs sampling from a 2D Gaussian distribution. 500 iterations are made and the estimates of the mean and the covariance are printed in the figure.

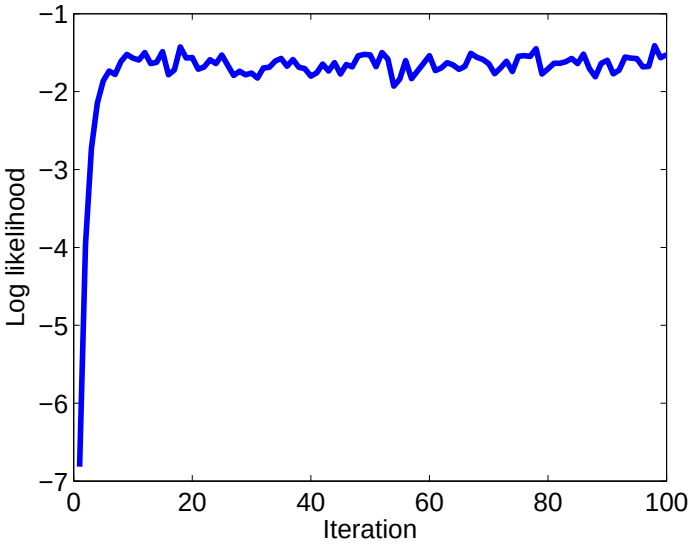


Figure 2.16: The log likelihood for 100 iterations averaged across 100 runs. For the first iterations the log likelihood is small, but around the tenth iteration it stabilises. This is called the burn in period.

Consider the joint distribution $p(x, y)$ of two sets of variables, X and Y. If the sets of variables are independent their joint distribution will factorize into the product of their marginals $p(x, y) = p(x)p(y)$. If the variables are not independent, the Kullback-Leibler divergence between the joint distribution and the product of the marginals gives a measure of how close they are to being independent. This measure is also called Mutual Information (MI) and is a measure of the amount of information that one random variable contains about another random variable. MI is given by [14]

$$\mathrm{MI}(X;Y) = KL(p(x,y) \parallel p(x)p(y)) = \sum_{y \in Y} \sum_{x \in X} p(x,y) \log \frac{p(x,y)}{p(x)p(y)} \tag{2.9}$$

The Normalized MI (NMI) is calculated as

$$\mathrm{NMI}(X;Y) = \frac{2 \cdot \mathrm{MI}(X;Y)}{\mathrm{MI}(X) + \mathrm{MI}(Y)} \tag{2.10}$$

2.11 NPAIRS Evaluation Criteria

To evaluate the models' predictability and reproducibility, the NPAIRS evaluation criteria [46] is used. The predictability is evaluated using the test log likelihood, and the reproducibility of the identified community structures is evaluated using Mutual Information between the node assignments, $\mathbf{Z}$. The data is randomly split in two equally sized sets, S1 and S2, and inference is made separately on each set.
The node assignments, $\mathbf{Z}$, and the link probabilities, $\boldsymbol{\eta}$, from the maximum a posteriori probability (MAP) solution for the model inferred using S1 are used to calculate the test log likelihood for S2 and vice versa. The average test log likelihood of the two splits is used as the predictability measure. The test log likelihood of set S2 for the model inferred using S1 is given as [5]

$$\begin{aligned} &\log L(\mathbf{Z}, \boldsymbol{\eta} \mid \mathbf{A}^{S2,(1)}, ..., \mathbf{A}^{S2,(N)}) \\ &= \frac{1}{N} \sum_{n=1}^{N} \sum_{j>1} \left[A_{i,j}^{S2,(n)} \log(\mathbf{z}_{i_r} \boldsymbol{\eta} \mathbf{z}_{j_r}^T) + (1 - A_{i,j}^{S2,(n)}) \log(1 - \mathbf{z}_{i_r} \boldsymbol{\eta} \mathbf{z}_{j_r}^T) \right] \end{aligned} \tag{2.11}$$

The reproducibility measure found by the MI between the node assignments $\mathbf{Z}^{(S1)}$ and $\mathbf{Z}^{(S2)}$ of the MAP solution for each split is calculated as [5]

$$\mathrm{MI}(\mathbf{Z}^{(S1)}, \mathbf{Z}^{(S2)}) = \sum_{i=1}^{D1} \sum_{j=1}^{D2} p(\mathbf{z}_{i_r}^{(S1)}, \mathbf{z}_{j_r}^{(S2)}) \log \left(\frac{p(\mathbf{z}_{i_r}^{(S1)}, \mathbf{z}_{j_r}^{(S2)})}{p(\mathbf{z}_{i_r}^{(S1)}) p(\mathbf{z}_{j_r}^{(S2)})} \right) \tag{2.12}$$

2.12 Validation of the Infinite Relational Model

To validate the IRM, a synthetic dataset from four different models is generated, and each dataset is inferred with all four models as in [5].
The four models are the IRM, the Infinite Diagonal Model (IDM), the model proposed in [26] which will be termed the Infinite Hofman-Wiggins (IHW) model, and Bayesian Community Detection (BCD). The structure of the link probability matrix, $\boldsymbol{\eta}$, defines the difference between IRM, IDM and IHW. As described $\boldsymbol{\eta}$ has a prior given by the Beta function with the hyperparameters $\beta_+(a, b)$ and $\beta_-(a, b)$. The IRM has different link characteristics within and between communities, see (2.13). The IDM has different link probabilities within communities, but one common link probability between communities, see (2.14). The IHW has one common link probability within communities and one common link probability between communities, see (2.15).

$$\boldsymbol{\eta}_{IRM} = \begin{bmatrix} \eta_{11} & \eta_{12} & \cdots & \eta_{1m} \\ \eta_{21} & \eta_{22} & \cdots & \eta_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ \eta_{m1} & \eta_{m2} & \cdots & \eta_{mm} \end{bmatrix} \tag{2.13}$$

$$\boldsymbol{\eta}_{IDM} = \begin{bmatrix} \eta_1 & \eta_0 & \cdots & \eta_0 \\ \eta_0 & \eta_2 & \cdots & \eta_0 \\ \vdots & \vdots & \ddots & \vdots \\ \eta_0 & \eta_0 & \cdots & \eta_m \end{bmatrix} \tag{2.14}$$

$$\boldsymbol{\eta}_{IHW} = \begin{bmatrix} \eta_c & \eta_0 & \cdots & \eta_0 \\ \eta_0 & \eta_c & \cdots & \eta_0 \\ \vdots & \vdots & \ddots & \vdots \\ \eta_0 & \eta_0 & \cdots & \eta_c \end{bmatrix} \tag{2.15}$$

The BCD model [37] is a nonparametric Bayesian generative model of networks with community structure. It differs from the other models by having a parameter, γ, describing the strength of the community structure. The parameter defines the extent to which a community has comparatively fewer external links than internal links. The BCD model recover the IRM as $\gamma \to \infty$, but γ is restricted to the interval [0 1].

When generating synthetic data, assignments for $J = 100$ nodes was drawn from the CRP with $\alpha = 5$. The link probabilities were drawn from the Beta function with $\beta_+(a, b) = [2\ 1]$ and $\beta_-(a, b) = [3\ 5]$ constrained according to Equations (2.13)-(2.15). When generating BCD data the cluster gap parameter $\gamma = [1\ 1]$. For each model 20 undirected graphs were generated. The probability matrix $\boldsymbol{\eta}$ was shared across the 20 graphs.
For each dataset 500 different half-splits were made. Each split was inferred by each of the models, and the predictability and reproducibility was calculated as described in Section 2.11.
Figure 2.17, 2.19, 2.21, and 2.24 shows the predictability-reproducibility (PR) plots for each type of generated data. In the PR plots the predictability for the random and the true model and the maximum reproducibility are shown.
The predictability for the random model is found by generating a random $\boldsymbol{\eta}$, by setting all elements equal to the average density across the 20 graphs generated. The test log likelihood is calculated for each of the 20 graphs using the random $\boldsymbol{\eta}$ and the predictability for the random model is the average test log likelihood for the 20 graphs. The predictability for the true model is the average test log likelihood for the 20 graphs using the true $\boldsymbol{\eta}$.
The maximum reproducibility is equal to the entropy given by

$$\mathrm{MI}_{\max} = \text{Entropy} = -\sum_i p(i) \log p(i), \tag{2.16}$$

where $p(i)$ is the probability of assigning a node to the ith cluster. As all the clusters have equal probability, namely $\frac{1}{N}$, where N is the number of clusters, Equation 2.16 reduces to

$$\mathrm{MI}_{\max} = -\sum_N \frac{1}{N} \log \frac{1}{N} = \log N$$

Figure 2.18, 2.20, 2.23 and 2.25 shows the number of clusters found by each of the models in each of the synthetic datasets. Figure 2.22 shows a zoom of Figure 2.21 to expose the results of the IRM, which is very close to the best

results of the BCD.
Figure 2.17 shows the result, when inferring the synthetic IRM dataset. The BCD is slightly better than the IRM, but they both have almost the same predictability as the true model and they both have better predictability and reproducibility than the IDM and the IHW. The IDM and the IHW perform equally. Figure 2.18 shows that the BCD finds the true number of clusters in the dataset, while the IRM slightly underestimates the number. Both the IDM and the IHW underestimates the number of clusters, IDM more than IHW though.
Figure 2.19 shows the result, when inferring the synthetic IDM dataset. The IDM has the best performance. The IRM has almost the same predictability, but has a little less reproducibility. The IHW performs similarly to random in predictability and has minimum reproducibility. The BCD performs in between the IRM and the IHW. Figure 2.20 shows that all the models underestimate the number of clusters in the IDM data. The IRM underestimates the most and the IDM is slightly better than the IHW and the BCD.
Figure 2.21 and 2.22 show the results, when inferring the synthetic IHW dataset. All the models have predictability close to the true model. The IHW has slightly better reproducibility than the IDM, which again is slightly better than the IRM. The best results from the BCD are similar to the results from the IRM. Figure 2.23 shows that both the IHW and the IDM find the true number of clusters while the IRM and the BCD underestimates the number.
Figure 2.24 shows the result, when inferring the synthetic BCD dataset. The BCD performs best in predictability and the second highest in reproducibility. The IHW has the highest reproducibility, while the IRM has the second highest predictability. The IDM has predictability and reproducibility equal to the IHW and the IRM, respectively. Figure 2.25 shows that the BCD finds the true number of clusters while the IRM and the IDM underestimate the number and the IHW overestimates the number.
For all four types of the synthetic data the IRM performs similarly to the model generating the data and thus does not seem to overfit the data, despite of the more simple structure in the IDM and the IHW data. Except on the IRM generated data the IRM underestimates the number of clusters more than the other models.

2.13 Permutation Test

A permutation test is a simple statistic test to test a null hypothesis. A permutation test can be used to determine whether an effect, e.g. the mean or the correlation between two variables, can be ascribed to the randomness introduced in selecting the sample, or if it reflects an effect present in the population. A

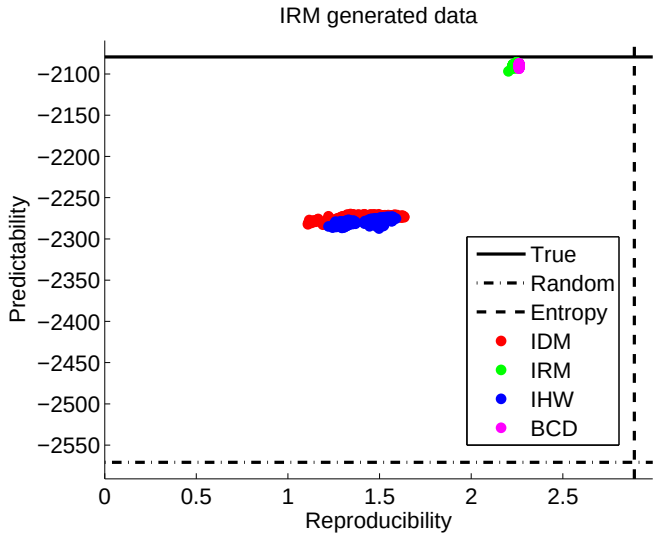


Figure 2.17: Predictability (test log likelihood) as function of reproducibility (MI) for data generated by the IRM. 500 half-splits.

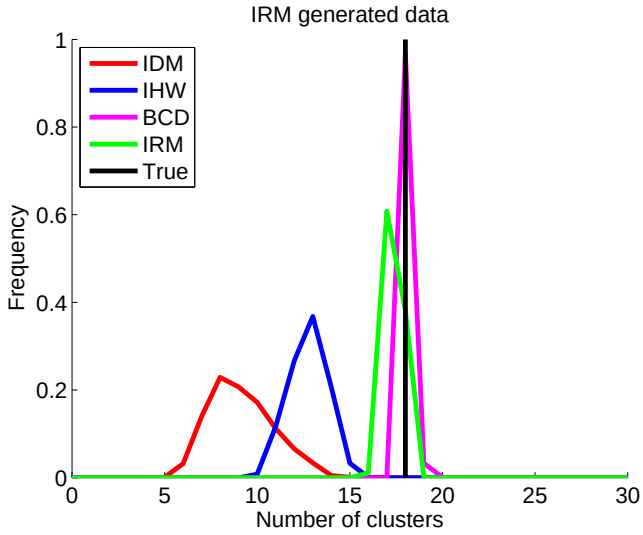


Figure 2.18: The number of components found by each of the models when inferring the IRM data.

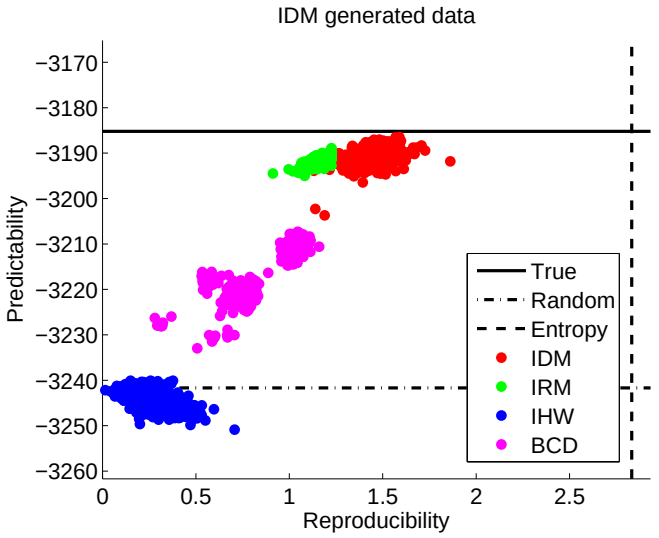


Figure 2.19: Predictability (test log likelihood) as function of reproducibility (MI) for data generated by the IDM. 500 half-splits.

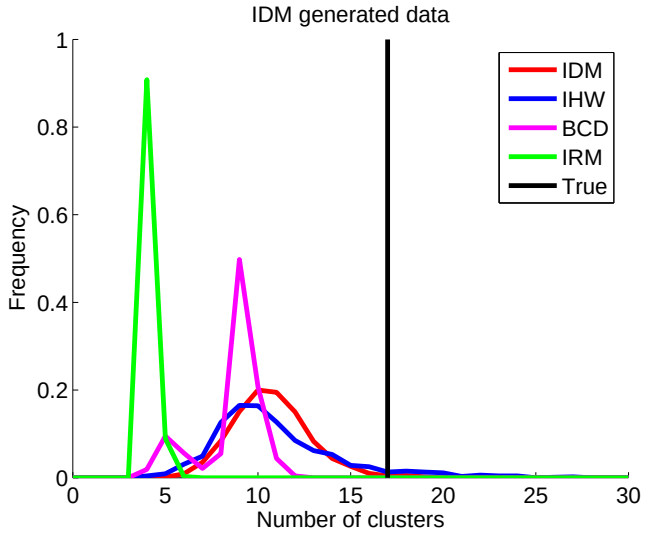


Figure 2.20: The number of components found by each of the models when inferring the IDM data.

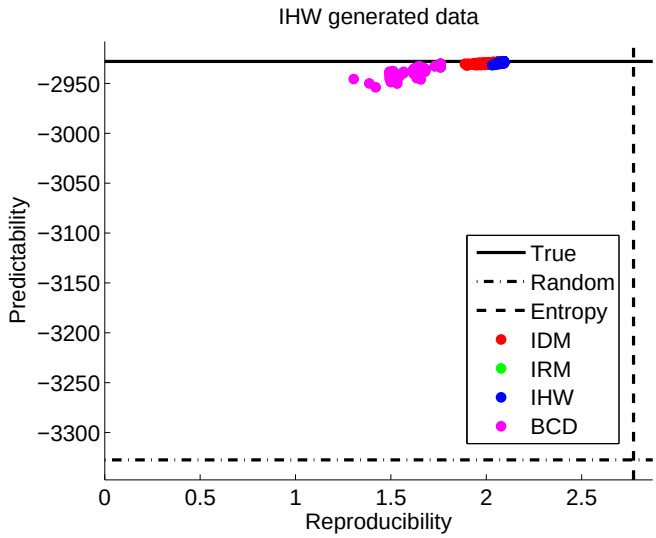


Figure 2.21: Predictability (test log likelihood) as function of reproducibility (MI) for data generated by the IHW. 500 half-splits.

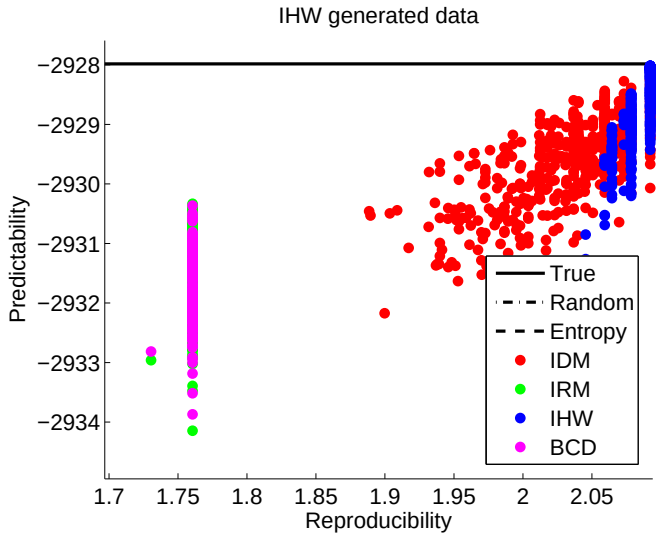


Figure 2.22: A zoom of Figure 2.21 to expose the results of the IRM. Predictability (test log likelihood) as function of reproducibility (MI) for data generated by the IHW. 500 half-splits.

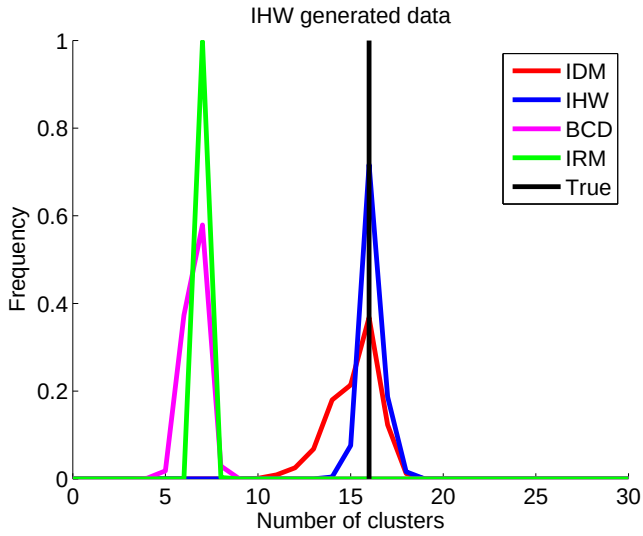


Figure 2.23: The number of components found by each of the models when inferring the IHW data.

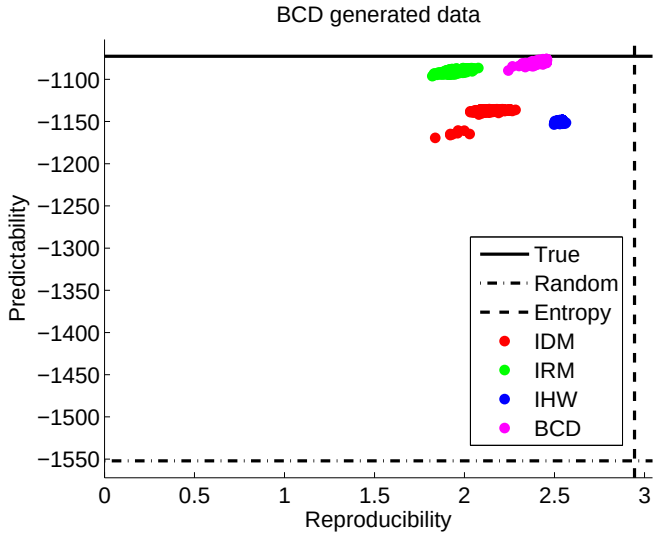


Figure 2.24: Predictability (test log likelihood) as function of reproducibility (MI) for data generated by the BCD. 500 half-splits.

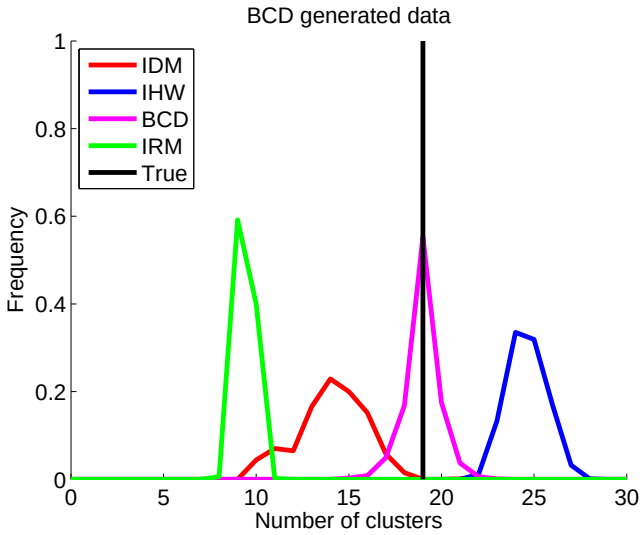


Figure 2.25: The number of components found by each of the models when inferring the BCD data.

permutation test is performed as

1. Select a statistic that measure the effect, you are looking for, e.g. the mean or the correlation, and calculate the observed value T_{Obs}.

2. Construct the sampling (permutation) distribution by resampling. This is done by splitting the population in two groups randomly, and calculate the statistic, T, between the two new groups. To obtain a distribution the resampling is repeated hundreds of times. The number of resamples is denoted N_{Tot}.

3. Calculate the p-value of accepting the null hypothesis. The p-value is equal to the proportion of samples that exceeds the observed value for the true split.

$$p = \frac{N_{T>T_{\text{Obs}}} + 1}{N_{\text{Tot}} + 1}$$

One advantage of a permutation test is that a permutation test does not depend on the distribution, where as the student's t-test require the distribution to be roughly normal. If the sampling distribution is normal, the p-value for the

permutation test is very similar to the p-value obtained by the two-sample t-test [25]. The normality of the sampling distribution can be checked directly by looking at the permutation distribution. The disadvantage of a permutation test is that it has low power. You need at least 19 resamples to get a p-value of 0.05.

The permutation test requires the two populations to have identical distributions when the null hypothesis is true. Not only the same mean, but also the same spreads and shapes are required. However, it is usually reasonable to assume that the distributions are approximately the same if the null hypothesis is true [25].

Chapter 3

Data and Processing of the Data

3.1 The Data

fMRI data from 29 healthy volunteers is used in the project. The median age in years was 45 in the range 22-69. Most of the participants were right-handed as revealed by the Edinburgh Inventory[39].

The MRI measurements were performed on a 3T Magnetom Trio scanner. The measurements were recorded with a standard single-channel birdcage head-coil using a T_2^*-weighted echo planar imaging (EPI) sequence (TR=2490 ms, TE=30 ms; flip angle 90°). In total 240 volumes were acquired over 10 minutes (42 contiguous axial slices; slice thickness=3 mm; FOV=111 192x192 mm; 64x64 acquisition matrix; voxel size=3x3x3 mm).
The cardiac cycle was monitored with an infrared pulse oximeter attached to the index finger and the respiratory frequency was monitored with a pneumatic thoracic belt.

The subjects were instructed to perform alternating left and right finger tapping. In the left and right conditions there is a visual cue (red/green blinking

dot 1 Hz) to pace movements. In between each finger tapping block there is a resting block, corresponding to baseline, where the subjects have to focus on a fixation cross in the middle of the screen. The duration of each finger tapping block is 20 s, which is approximately 8 volumes. The first 11 volumes are baseline followed by 8 volumes of left finger tapping, 4 volumes of baseline, 8 volumes of right finger tapping and so on. The onsets of the left finger tapping in volumes are [12 36 60 84 108 132 156 180 204 228] and the onsets of the right finger tapping is [24 48 72 96 120 144 168 192 216 240].
To compensate for the delay in the haemodynamic response function, the two volumes in the beginning of each block is treated as baseline and one volume after the block is used as either left or right finger tapping. As this is not possible in the last block with right finger tapping it is disregarded from the analysis together with the last block of left finger tapping, to get an equal number of blocks with left and right finger tapping.

High-resolution three-dimensional structural MRI scans of the brain was acquired using an eight-channel surface head-coil (Invivo, FL, USA). A saggital 120 magnetisation prepared rapid acquisition gradient echo (MPRAGE) sequence (TR=1550 ms, TE=3.04, inversion time (IT)=800 ms; 192 slices; 1 mm isotropic resolution; field of view (FOV)=256 mm; flip-angle=9°) was acquired. The structural scans were used in the pre-processing of the functional images.

To obtain the diffusion data set DWI data was obtained from 22 of the healthy volunteers. Each subject underwent two DWI sessions. For each session diffusion along 61 directions were recorded with $b = 1200$ s/mm^2. Additionally, 10 images with $b = 0$ were obtained.
Only the second DWI session is used in the analysis with the IRM, because it contains more streamlines between cortex and striatum after the tractography than the first session.

The acquisition of the data sets was performed by MD Anne-Marie Dogonowski and PhD Kristoffer Hougaard Madsen at Copenhagen University Hospital Hvidovre.

3.2 Statistical Parametric Mapping

To process the fMRI data, the program *Statistical Parametric Mapping 8* (SPM) [16] is used. SPM is a common tool for analysis of brain data. SPM is a voxel-

based approach used to identify functionally specialized brain responses and to make some comments about regionally specific responses to experimental factors, employing classical inference. In order to assign an observed response to a particular brain structure, or cortical area, the data must conform to a known anatomical space. The pre-processing steps in SPM realigns and map the time-series of images into some standard space, in this particular case, the Montreal Neurological Institute (MNI) standard space [15].

3.3 Pre-processing of the fMRI Data

The first step in the pre-processing of the fMRI data is to realign the images. This procedure realigns the time-series of the images acquired from one subject. The aim with the realignment is to remove movement artefacts. The subject will always move in the scanner, and hence, there will always be movement artefacts. The artefacts can be either correlated or uncorrelated with the task, but in both cases it will affect the sensitivity of the analysis. In SPM the translation in the x-, y-, and z-direction, and the rotation (pitch, roll, and yaw) for each image are estimated. The first image for each subject is used as reference, and all the images are re-sliced, including the mean image.

There is no general agreement upon whether or not the data should be slice timed. Neither, if it should be performed before or after the realignment. The argument for doing it, is to correct for differences in slice acquisition times. It is intended to correct for the staggered order of the slice acquisition that is used doing the scanning. The correction is necessary to make the data on each slice correspond to the same point in time. Without correction the time between two adjacent slices can be up to TR/2 in the case of interleaved scans. The slice timing procedure is important in event-related designs, but in block designs, where each block lasts for several TRs, there is no need to introduce the interpolation errors, even though the interpolation errors introduced by the slice timing are generally small. Since the current design is a block design, the data is not slice timed.

In the second step the SPM toolbox Slice-Wise (written by PhD Kristoffer Hougaard Madsen) is used to model the physiological effects from pulse, heartbeat, and respiratory movements, and the residuals are estimated. The images of the residuals are used in the subsequent processing.

In the next step, the functional images are co-registered to the structural image. The mean of the functional images is estimated, and the structural image and the mean functional image are co-registered, such that the mutual information is maximized. After the co-registration there is anatomical correspondence between the functional images and the structural image. The functional images are T_2^*-weighted, which make them sensitive to blood-flow, blood-volume, and the degree of oxygenation. The structural image is T_1-weighted, which gives large signal-variation between different kinds of tissue, i.e. it shows the anatomy of the brain.

In the normalization step, the images are mapped into the standard space, the MNI space. In the normalization, two different voxelsizes are selected, namely $3 \times 3 \times 3$ mm^3 and $2 \times 2 \times 2$ mm^3. The reasons for the voxelsize of $3 \times 3 \times 3$ mm^3 are that it is the same voxelsize as used in the acquisition of the data and with this voxelsize a reasonable graphsize is obtained. This voxelsize is used when analysing the fMRI graphs alone. The voxelsize of $2 \times 2 \times 2$ mm^3 is used to make the resolution of the fMRI and the DWI images the same, so this voxelsize is used in the combined analysis of the fMRI and DWI data. To prevent the graphs from becoming too big a smaller threshold is used in the group analysis.

In the last step of the pre-processing of the fMRI images, the images are smoothed with an 8 mm Full-Width Half-Maximum Gaussian smoothing kernel. The images are smoothed to improve the signal-to-noise ratio. The smoothing can prevent noise, if the images are not completely aligned or normalized.

3.4 Model Specification, Model Estimation and Inference

To perform statistical analysis, SPM uses the General Linear Model (GLM). The statistical analysis corresponds to modelling the data, to partition observed neurophysiological responses into components of interest, confounds, and errors, and making inferences about the interesting effects in relation to the error variance.
The pre-processed images are specified and estimated. The realignment parameters and their first and second derivatives are used as multiple regressors in the design matrix.

3.4.1 The Design Matrix

The equation for the GLM is given by

$$\mathbf{Y} = \mathbf{X}\beta + \epsilon \tag{3.1}$$

Where $\mathbf{Y}$ is the observed data, $\mathbf{X}$ is the design matrix, β is the real data and ϵ is the noise.
The design matrix $\mathbf{X}$ contains the explanatory variables and defines the experimental design. It contains one row for each scan and one column for each explanatory variable or effect. The design matrix can also contain columns with time-series that are used to remove low frequency variations in the signals due to artifacts, e.g. biorhythms or signals from the scanner.
Figure 3.1 shows the design matrix for subject HC01. The design matrix contains 29 effects. The first four effects define the experimental conditions, left and right finger tapping, and their derivatives. The effects R1-R24 are motion related effects and the last effect is the constant regressor.

3.4.2 Inference

For each subject the contrasts Left vs. Baseline, Right vs. Baseline, Left vs. Right, and Right vs. Left are investigated. The contrast Left vs. Baseline shows the activations, when the subject performs left finger tapping, while the contrast Right vs. Baseline shows the activations when the subject performs right finger tapping. The contrast Left vs. Right shows the activations which is greater during left finger tapping, than during right finger tapping, and the contrast Right vs. Left shows the opposite.
In order to find the significant activations in the group, inference is made, using the contrast of parameter estimates for each subject. A 1-sample T-test for each of the contrasts are performed.

The resulting contrasts Left vs. Baseline and Right vs. Baseline, from the group analysis, are used to find the regions of interests (ROIs) in the motor cortex and the striatum.
To find the ROIs in the motor cortex in the images with the resolution $3 \times 3 \times 3$ mm^3, a threshold, corrected for multiple comparisons using the Family Wise Error (FWE), $p_{FWE} < 0.001$ is used. The p-value corresponds to a T-value equal to 7.4140 for the contrast Left vs. Baseline and a T-value equal to 7.4171

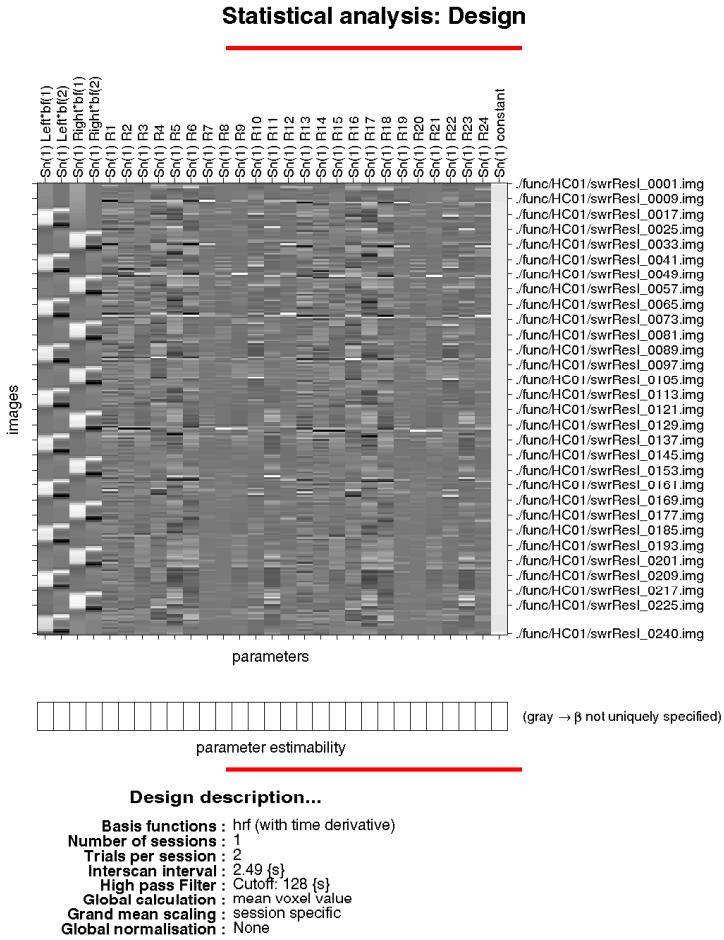


Figure 3.1: The design matrix for subject HC01. The design matrix contains 29 effects. The first four effects define the experimental conditions and their derivative, left and right finger tapping, respectively. The effects R1-R24 are motion related effects. The last effect is the constant regressor.

for the contrast Right vs. Baseline. Figure 3.2 and 3.3 show the significant activations for the two contrasts.
For the images with the resolution $2 \times 2 \times 2$ mm^3 a threshold of $p_{FWE} < 10^{-6}$ is chosen, corresponding to a T-value equal to 10.9395 for both the contrasts Left vs. Baseline and Right vs. Baseline.
To find the ROIs in the striatum a threshold of $p < 0.01$ (unc.) for both resolutions is chosen. This threshold corresponds to a T value equal to 2.4671 for both contrasts.

3.5 Obtaining the Graphs for the fMRI Data

The images of the T-statistics for the contrasts Left vs. Baseline and Right vs. Baseline are loaded into the SPM toolbox xjview one at a time, and the T value is set. Significant activations for the motor cortex and the striatum are saved as masks.
The activated regions in the striatum contains a lot of voxels belonging to other parts of the brain, because of the high p-value and the lack of correction. To ensure a mask only contains voxels belonging to the striatum, the mask is multiplied with a template-mask of the AAL regions [47] of the putamen and the caudate. This procedure is performed using the imcalc option in SPM for both hemispheres.
Both hemispheres contain significant activations for both contrasts. To ensure that each voxel only appears once in the graph, the masks of the cortex are combined using the imcalc option in SPM, resulting in one mask for each hemisphere. In total there are 1610 voxels in the graphs for the images with resolution $3 \times 3 \times 3$ mm^3, 1137 voxels in the cortex and 473 voxels in the striatum, respectively. The graphs for the images with resolution $2 \times 2 \times 2$ mm^3 contain 2526 voxels, 1126 voxels in the cortex and 1400 voxels in the striatum, respectively.

The time-serie for each voxel in the four masks is extracted for all subjects for each of the three states (Baseline, Left and Right). Because every second block in the design contains the state Baseline, there are 103 volumes with Baseline and only 63 volumes with Left finger tapping and Right finger tapping, respectively. The correlations between the time-series are calculated, and three correlation matrices are obtained for each subject.

The correlation matrices are thresholded to obtain a certain graph density. Since we are interested in the correlation between cortex and striatum, the links within

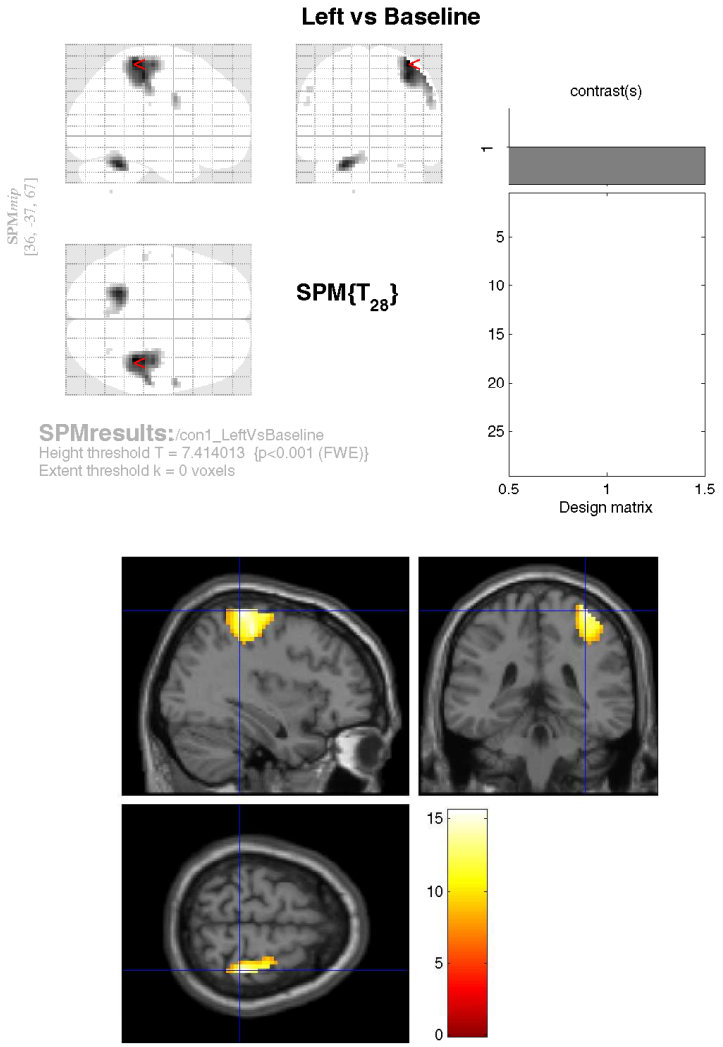


Figure 3.2: The significant activations ($p_{FWE} < 0.001$) in the group analysis for the contrast Left vs. Baseline. The upper left part of the figure shows the maximum intensity plot. The upper right part shows the design matrix and the lower part of the figure shows the significant activations on sections of the brain.

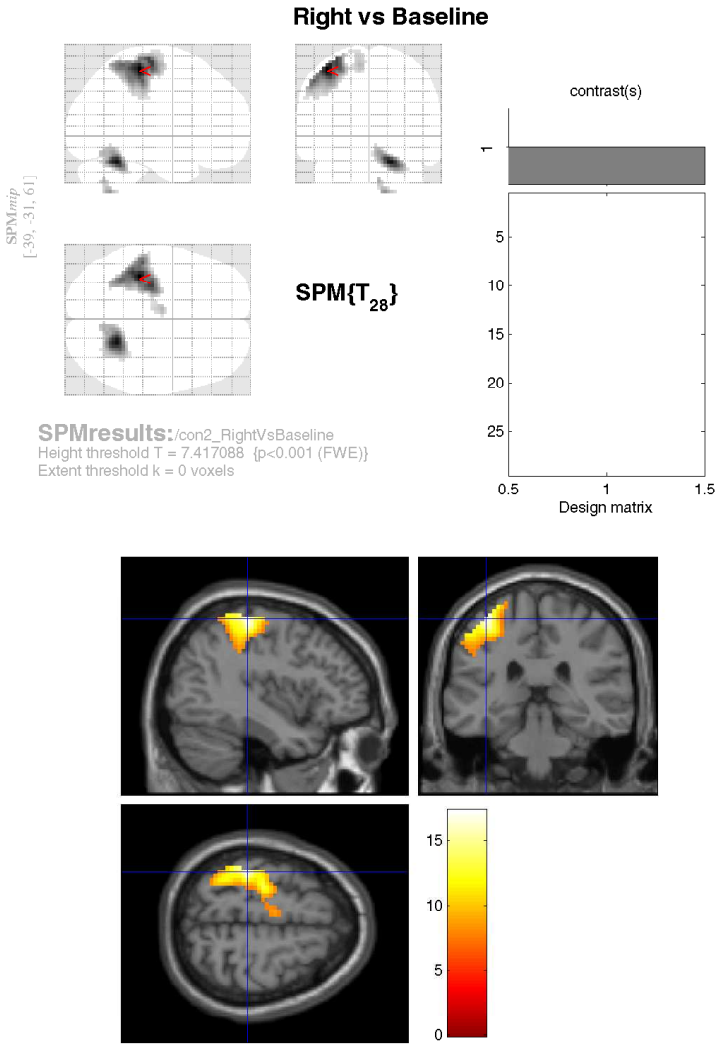


Figure 3.3: The significant activations ($p_{FWE} < 0.001$) in the group analysis for the contrast Right vs. Baseline. The upper left part of the figure shows the maximum intensity plot. The upper right part shows the design matrix and the lower part of the figure shows the significant activations on sections of the brain.

cortex and striatum, respectively, are treated as missing. The reason is that the correlations within cortex and striatum, respectively, are higher than the correlations between cortex and striatum. If the whole correlation matrix is thresholded, almost no correlations between cortex and striatum will remain. The thresholded correlation matrices are used as input to the IRM and inference is made. In the analysis of the fMRI graphs a graph density of 3% is used, and in the combined analysis of the fMRI and DWI graphs, a graph density of 0.5% is used. When inferring the data with the IRM the hyperparameters are set to: $\beta_+(a,b) = \beta_(a,b) = [1\,1]$ and $\alpha = \log J$, where J is the number of nodes in the graph. The initial number of clusters is set to 50 and the number of iterations is 200.

3.6 Processing of the DWI Data

FMRIB's Software Library (FSL) [30, 51, 44, 48] is a comprehensive library of analysis tools for fMRI, MRI, and DWI brain data. To process the DWIs the FMRIB's Diffusion Toolbox (FDT), which is a part of the FSL, is used.

The processing pipeline for DWIs consists of six steps:

1. Any study or scanner-specific pre-processing, such as coversion from DICOM to NIFTI or removal of images affected by large artifacts.
2. Eddy current correction
3. Fitting of the probabilistic diffusion model on corrected data using bedpostx
4. Brain extraction using BET
5. Registration
6. Probabilistic tractography run on the output of bedpostx using probtrackx2

Eddy current correction corrects for distortions in the DWIs using affine registration to a reference volume. The distortions are due to eddy currents in the gradient coils, induced by the changing magnetic field, which stretches and shears the DWIs, and due to head motions.

BEDPOSTX stands for Bayesian Estimation of Diffusion Parameters Obtained using Sampling Techniques and the X stands for modelling crossing fibres. Bedpostx runs MCMC sampling to build up distributions on diffusion parameters at each voxel. Bedpotstx generates all the files necessary for running probabilistic tractography.
The first three steps, the study and scanner-specific pre-processing, the eddy current correction and bedpostx, were already performed on the dataset by Kasper Winther Andersen, and I have chosen to use his results.

The Brain Extraction Tool (BET)[43] is used to extract a betted image of the brain from the structural image for each subject. BET deletes non-brain tissue, such that the betted image only contains brain tissue.
In the registration step, transformations between the three spaces: diffusion-, structural-, and standard-space, are made. The registration is used, when the seed voxels in the probabilistic tractography are not in the diffusion space. The seed masks, ROIs, found from the fMRI data, are in the standard space. Linear registration FLIRT is applied to obtain the transformation matrices. To make the linear registration, the structural image with BET applied is used.

In the last step of the processing of the DWI data, within FDT, probabilistic tractography is carried out. In this step, the streamlines between each voxel-pair in the ROIs are found. The masks found by the analysis of the fMRI data are combined to one single seed mask. Probtrackx2 generates connectivity distributions from all voxels in the seed mask, by sampling from the distribution of voxel-wise principal diffusion direction, each time computing a streamline through these local samples to generate a probabilistic streamline or a sample from the distribution on the location of the true streamline. By taking many such samples, FDT is able to build up the histogram of the posterior distribution on the streamline location or the connectivity distribution[2]. The output is a matrix with a value for each voxel representing the connectivity value between that voxel and the seed voxel. Many of the voxels will have a value of zero. The maximum value a voxel can take is the number of samples in the tractography. The number of samples is chosen to be 5,000.

3.7 Obtaining the Graphs for the DWI data

In order to obtain undirected graphs, the connectivity matrices are averaged, $\overline{\mathbf{C}} = \frac{1}{2}(\mathbf{C} + \mathbf{C}^T)$. The connectivity matrices are processed in the same way as

the correlation matrices obtained from the fMRI data. The links between voxels within cortex and striatum, respectively, are treated as missing. Regardless of the 5,000 samples 4 subjects have less than 1000 links between cortex and striatum, corresponding to a graph density less than 0.06%. The connectivity matrices are thresholded, such that the resulting graph density is 0.5%. Subjects with too few links to obtain a graph density of 0.5% are disregarded from the analysis. The thresholded connectivity graphs from the remaining 14 subjects are used as input to the IRM and inference is made. When inferring the data with the IRM the hyperparameters are set to: $\beta_+(a,b) = \beta_(a,b) = [1\,1]$ and $\alpha = \log J$, where J is the number of nodes in the graph. The initial number of clusters is set to 50 and the number of iterations is 200.

3.8 The Experiments

Two hypotheses are tested. The first hypothesis is that IRM finds the same clusters in the different states. The second hypothesis is that IRM finds the same communication pattern in the different states. Both hypotheses are tested both on the fMRI graphs alone and on fMRI and DWI graphs together. The experiments are listed in Table 3.1.
For testing the reproducibility and predictability of the IRM, ten different half-splits are generated for each of the four experiments. For each split the model is inferred and the MI and the NMI between the two sets are calculated. To evaluate the predictability the test log likelihood between the two splits are calculated, as described in section 2.11. The predictability is evaluated for experiment 1 and 2 (Table 3.1) only. The reason is that the $\boldsymbol{\eta}$ matrices have to be identical to calculate the test log likelihood, which we cannot restrict when testing the hypothesis that IRM finds the same communication pattern ($\boldsymbol{\eta}$) in the different states.

3.9 Analysis of the Results

The outputs from the IRM are analysed using a permutation test. To test hypothesis 1 and 2 (Table 3.1): the clustering is pair-wise the same in the different states, all the assignment matrices are pair-wise compared. As a distance measuare, the NMI is used. As an example the hypothesis $\mathbf{Z}_B = \mathbf{Z}_L$ is used. For each state we have ten assignment matrices, because of the ten resamples,

#	Hypothesis	Data	Graph density	# of subjects
1	$\mathbf{Z}_B = \mathbf{Z}_L = \mathbf{Z}_R$	fMRI	3%	29
2	$\mathbf{Z}_B = \mathbf{Z}_L = \mathbf{Z}_R = \mathbf{Z}_{DWI}$	fMRI + DWI	0.5%	14
3	$\boldsymbol{\eta}_B^n = \boldsymbol{\eta}_L^n = \boldsymbol{\eta}_R^n$	fMRI	3%	29
4	$\boldsymbol{\eta}_B^n = \boldsymbol{\eta}_L^n = \boldsymbol{\eta}_R^n = \boldsymbol{\eta}_{DWI}^n$	fMRI + DWI	0.5%	14

Table 3.1: A summary of the experiments carried out. $\mathbf{Z}_B, \mathbf{Z}_L, \mathbf{Z}_R$ and $\mathbf{Z}_{DWI}$ are the assignment matrices for the different states and $\boldsymbol{\eta}_B^n, \boldsymbol{\eta}_L^n, \boldsymbol{\eta}_R^n$ and $\boldsymbol{\eta}_{DWI}^n$ are the link probability matrices for the different states and subjects.

$\mathbf{Z}_{B1}, ..., \mathbf{Z}_{B10}$ and $\mathbf{Z}_{L1}, ..., \mathbf{Z}_{L10}$. First the mean NMI between the two groups is calculated as given by Equation (3.2).

$$\overline{d} = \frac{1}{2N} \sum_{n=1}^{N} \sum_{m=1}^{N} \mathrm{NMI}(Z_{Bn}, Z_{Lm}), \tag{3.2}$$

where $\mathrm{NMI}(Z_{Bn}, Z_{Lm})$ is the NMI between two assignment matrices, calculated as described in section 2.10.
The two groups are permutated, such that each group contains assignment matrices from both states, and the mean NMI between the two new groups is calculated. This is repeated 1000 times, meaning that the most significant p-value, which can be obtained is 0.001. All the calculated mean NMIs make up a distribution which can be compared to the mean NMI for the true split.

A permutation test is also performed to test if the link probability matrices between states are more similar than random. As a distance measure the correlation coefficient between two matrices is used. Because the $\boldsymbol{\eta}$ matrices have different sizes in the different resamples, the correlation coefficient cannot be calculated across resamples. Instead the mean correlation coefficient between two states for each subject is calculated. This is repeated for each resample and the average correlation coefficient between two states for each subject is found. The permutations are made by randomly permute one of the link probability matrices and the calculation is repeated. 1000 permutations are performed, meaning that the most significant p-value, which can be obtained is 0.001.

To test hypothesis 3 and 4 (Table 3.1): it is pair-wise the same communication pattern in the different states, 1000 adjacency matrices for each subject and

state is generated using the MAP solution across the ten resamples. The test log likelihood for the generated adjacency matrices for each state is calculated, to find the state with highest test log likelihood. In all calculations the link probability matrix corresponding to the given subject is used. A histogram of the test log likelihoods for each state is made. If the adjacency matrices can be classified as the correct state, based on the test log likelihood, it indicates that the communication pattern differs between the states.

To investigate the differences and similarities in the communication pattern, a paired t-test is made to find the cluster pairs with the most significant differences between the states. The cluster pairs are visualised by superimposing the clusters on brain sections of a brain template (anatomical MRI).

An outline of the Matlab functions used in the project is given in Appendix C.

Chapter 4

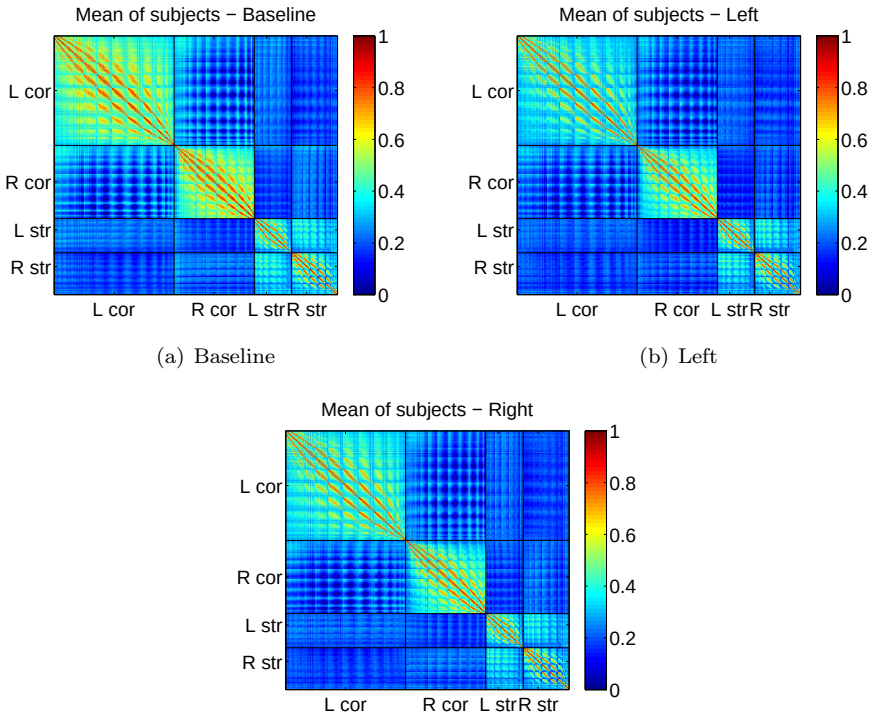
Results

4.1 Correlation, Connectivity and Adjacency Matrices

The average correlation matrix across subjects for each of the three fMRI states, Baseline, Left, and Right, are shown in Figure 4.1.

As the correlation matrices are symmetric, we consider the upper triangular matrices only and the links within cortex and striatum, respectively, are considered as missing. This results in a matrix with a structure similar to the one shown in Figure 4.2(a) (subject HC06 - Baseline) for each subject and state.
The adjacency matrix, made by thresholding, for subject HC06 - Baseline, with a graph density of 3% is shown in Figure 4.2(b). The number of links (nz) is 16,134.

Figure 4.3(a) shows the connectivity matrix for subject HC07, outputted from probtrackx2. Probtrackx2 is only given one union mask containing the four ROIs. Probtrackx2 reads the mask from the bottom to the top, such that the streamlines from alternating the left and right striatum are found first, and afterwards, the streamlines from alternating the left and right cortex are found. Therefore, the order of the voxels in the connectivity matrices is not the same



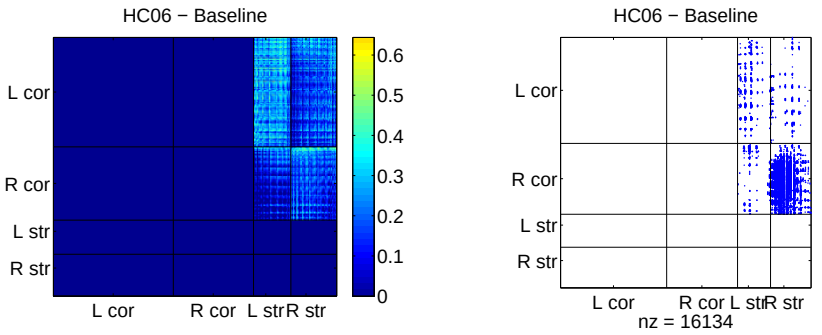
(a) Baseline

(b) Left

(c) Right

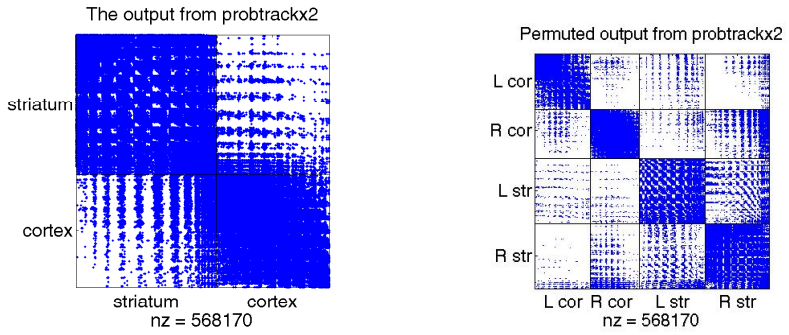
Figure 4.1: The averaged correlation matrix, across subjects, for state (a) Baseline, (b) Left, and (c) Right. The black lines indicate the boundaries between the masks.

as in the fMRI graphs. The connectivity matrices are permuted, such that the order of the voxels corresponds to the order in the fMRI graphs. Figure 4.3(b) shows the permuted connectivity matrix for subject HC07. The connectivity matrix outputted from probtrackx2 is not symmetric, so in order to make the graph symmetric it is averaged, and the result is shown in Figure 4.4(a). Now where the connectivity matrices are symmetric, we consider the upper triangular matrices only. The links within cortex and striatum, respectively, are considered as missing, resulting in the matrix in Figure 4.4(b). Figure 4.4(c) shows the thresholded connectivity matrix for subject HC07. The thresholded connectivity matrix has a link density of 0.5%.



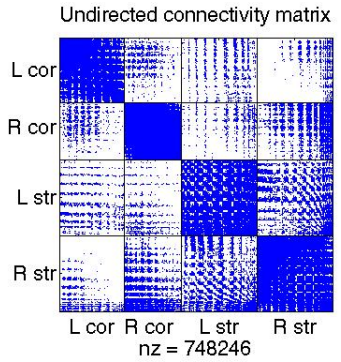
(a) (b)

Figure 4.2: (a) The upper triangular correlation matrix for subject HC06 - Baseline. The links within cortex and striatum, respectively, are treated as missing. (b) The adjacency matrix for subject HC06, Baseline, with a graph density of 3%. The number of links (nz) is 16,134. The black lines indicate the boundaries between the ROIs.

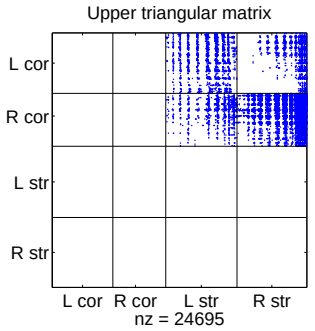


(a) Subject HC07 (b) Subject HC07

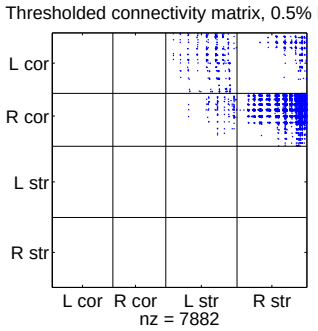
Figure 4.3: (a) Output from probtrackx2. It shows which nodes there are connected by a probabilistic streamline. (b) A permutation of the output from probtrackx2 to match the order of the voxels in the fMRI graphs. The black lines indicate the boundaries between the ROIs.



(a) Subject HC07



(b) Subject HC07



(c) Subject HC07

Figure 4.4: (a) The graph in Figure 4.3(b) is made symmetric by averaging. (b) The upper triangular connectivity matrix. Streamlines within cortex and striatum, respectively, are treated as missing. (c) The thresholded connectivity matrix, containing 0.5% links. The black lines indicate the boundaries between the ROIs.

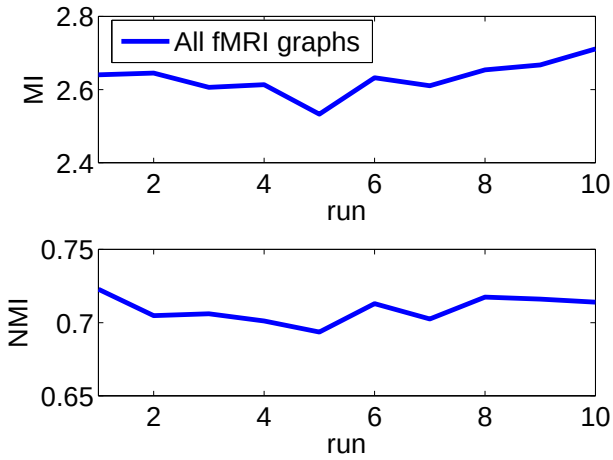


Figure 4.5: MI and NMI between splits when all the fMRI graphs are inferred simultaneously.

4.2 fMRI Data

Figure 4.5 shows the MI and the NMI between splits for the ten runs, when all the fMRI graphs are inferred simultaneously. It shows that the result found by the IRM is stable across runs.
Figure 4.6 shows the MI, the NMI, and the test log likelihood between splits for the ten runs when the fMRI graphs for the three states are inferred separately. Again, the IRM is stable across the runs. The NMI is stable across runs, but lower than the NMI when all the fMRI graphs are inferred simultaneously. The IRM has equally good reproducibility and predictability in the three fMRI states.

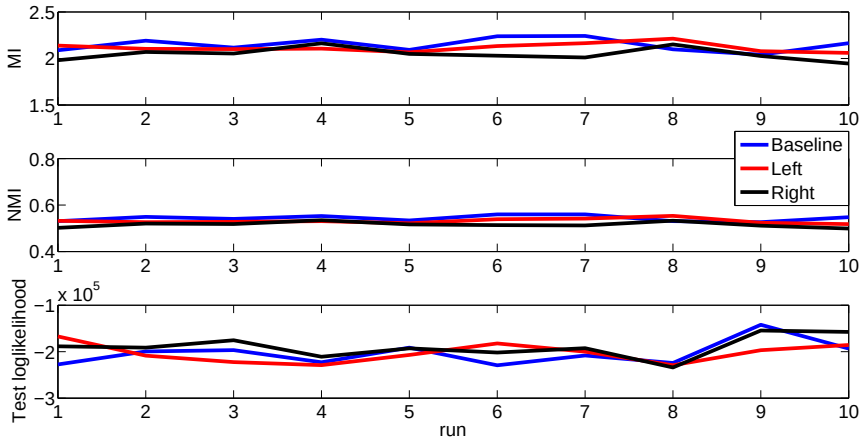


Figure 4.6: MI, NMI and test log likelihood between splits when the fMRI graphs from the different states are inferred separately.

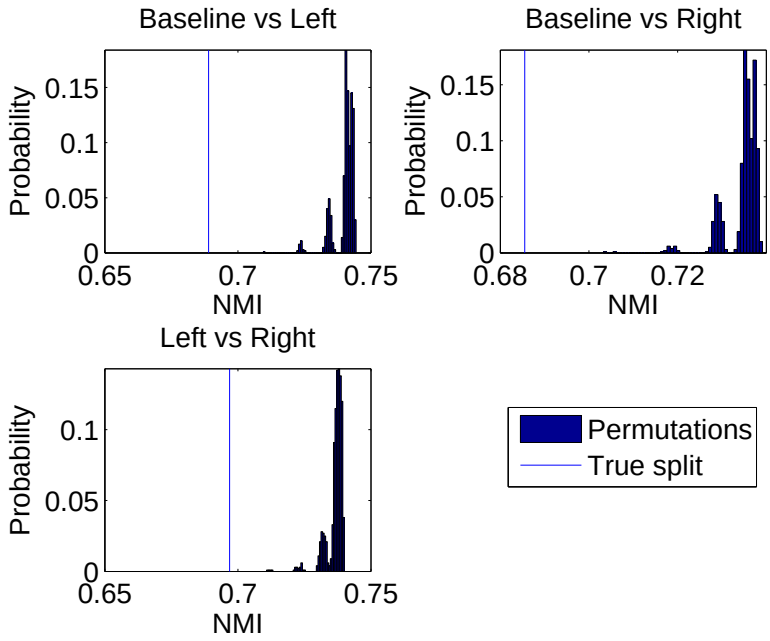


Figure 4.7: The histograms of the permutation test for testing if the clustering is the same in the different fMRI states.

4.2.1 The Clustering in the Different States

Figure 4.7 shows the result of the permutation test for testing if the IRM finds the same clusters in the three different fMRI states. The NMI for the true split is lower than for the permutations, meaning that we can reject the null hypothesis with a p-value of 0.001. The result shows that the assignment matrices are more alike within states than between states.

Figure 4.8 shows the five biggest clusters found in the state Baseline, Left, and Right, respectively. Only a few sections of the brain is shown to clarify the location of the clusters. The progress of the clusters throughout the brain in the three states, is shown in Figure B.1-B.3 in Appendix B. The white regions mark the ROIs. In the state Baseline the biggest cluster (red) contains voxels from left cortex only, while in both the other states it contains voxels from both right cortex and right striatum. The second largest cluster (cyan) contains voxels from both left cortex, left striatum, and right striatum in both the states

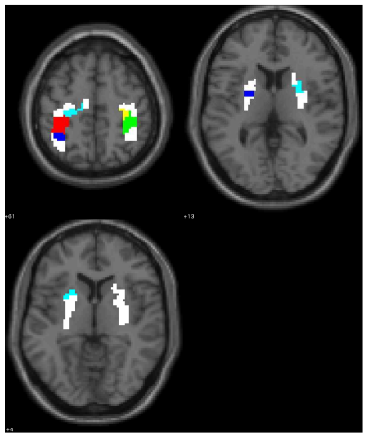
Cluster	Baseline	Left	Right
Red	L cor	R cor + R str	R cor + R str
Cyan	L cor + L str + R str	R cor	L cor + L str + R str
Green	R cor	L cor	L cor
Blue	L cor + L str	R cor	R cor
Yellow	R cor	L cor + R str	R cor

Table 4.1: Listing of which region the voxels in the five biggest clusters belongs to in the three different states. L cor indicates left cortex, while R str indicates right striatum.

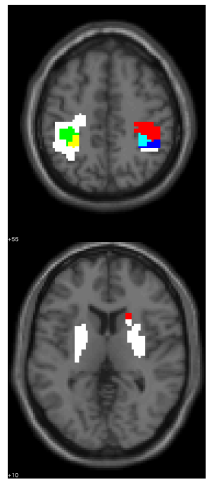
State	Left cortex	Right cortex	Left striatum	Right striatum
Baseline	16	13	13	23
Left	21	9	15	19
Right	21	12	16	17

Table 4.2: The number of clusters found in each of the four ROIs in each of the three states.

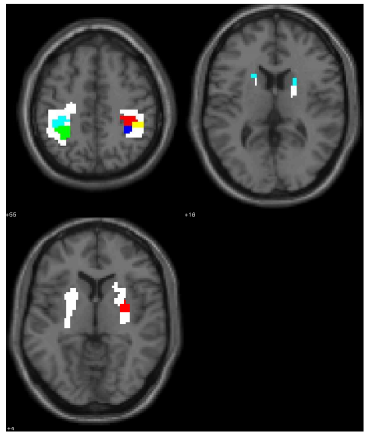
Baseline and Right. In the state Left the cluster contains voxels from right cortex only. The third biggest cluster (green) contains voxels from right cortex only in the state Baseline and from left cortex only in the states Left and Right. The fourth biggest cluster (blue) contains voxels from right cortex in the states Left and Right and voxels from left cortex and left striatum in state Baseline. The fifth biggest cluster (yellow) contains voxels from right cortex in states Baseline and Right and voxels from left cortex and right striatum in the state Left. An overview is given in Table 4.1, starting with the biggest cluster.
The number of clusters found in each of the four masks in each of the three states are listed in Table 4.2.



(a) Baseline



(b) Left



(c) Right

Figure 4.8: The five biggest clusters found in the state Baseline (a), Left (b), and Right (c). The order of the clusters regarding their size in descending order is red, cyan, green, blue and yellow. The white regions mark the ROIs.

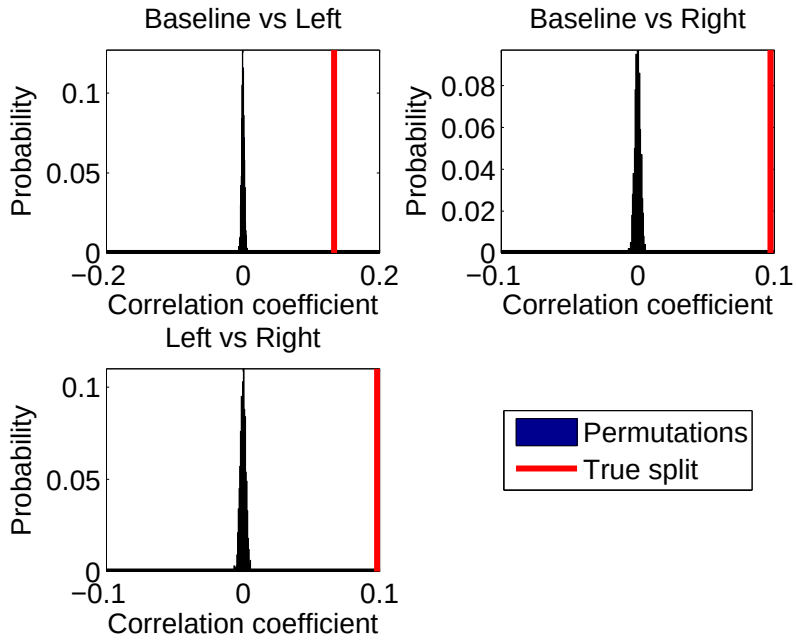


Figure 4.9: The result of the permutation test for testing if the correlation of the link probability matrices between states is different from random.

4.2.2 The Communication Pattern in the Different States

Figure 4.9 shows the result of the permutation test made to investigate if the correlation of the link probability matrices between states is different from random. The result shows that the correlation between all states are higher than random, even though the correlation coefficient is small for the true split.

To investigate if there is different connectivity in the three states, 1000 adjacency matrices for each subject and state is generated using the MAP solution across the ten runs. The test log likelihood for the generated adjacency matrices for each state is calculated. A histogram of the test log likelihoods for each state is made. Figure 4.10-4.12 shows the histogram of the test log likelihoods when the adjacency matrices are generated using the link probability matrix from state Baseline, Left, and Right, respectively. If each adjacency matrix is classified as the state with the biggest test log likelihood, all matrices are correctly classified.

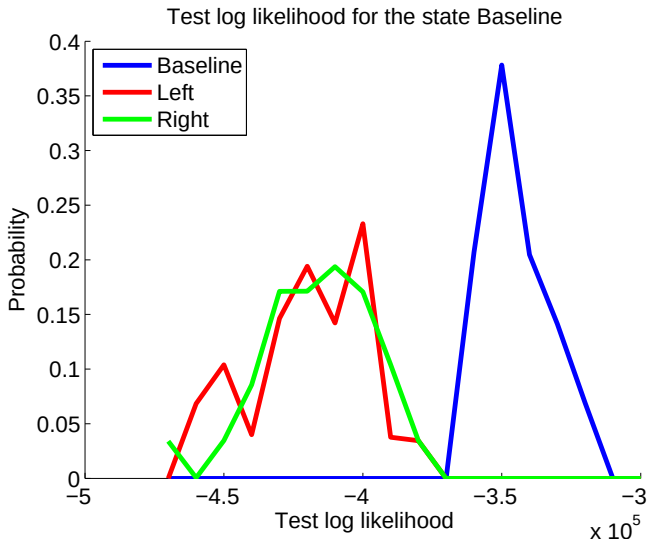


Figure 4.10: Histogram of the test log likelihoods for the three states when the adjacency matrices are generated using the link probability matrix from state Baseline.

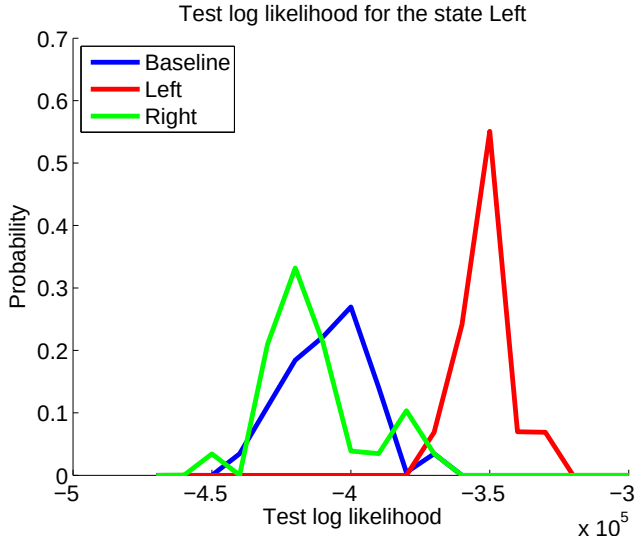


Figure 4.11: Histogram of the test log likelihoods for the three states when the adjacency matrices are generated using the link probability matrix from state Left.

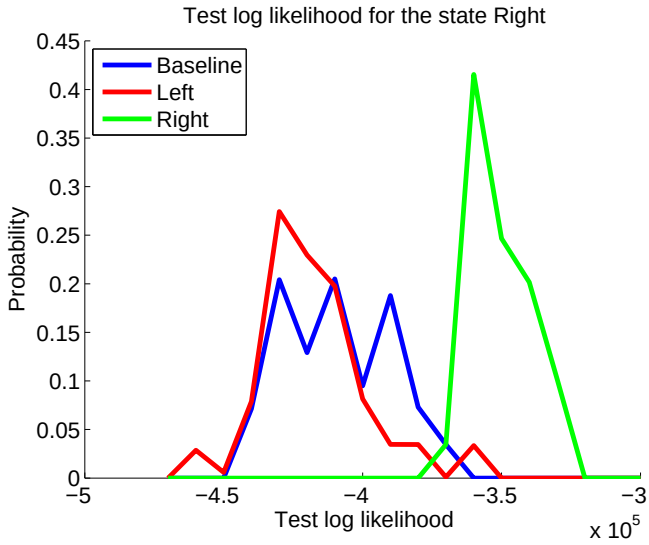


Figure 4.12: Histogram of the test log likelihoods for the three states when the adjacency matrices are generated using the link probability matrix from state Right.

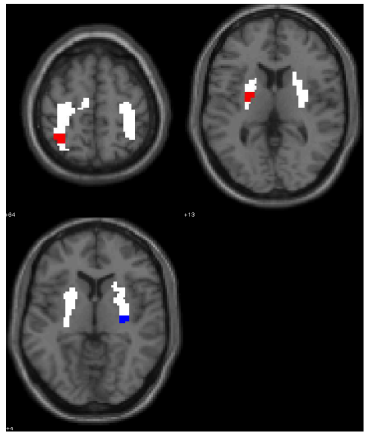
Using a paired t-test the clusters with most significant difference in the link probability between the states are found. Figures 4.13-4.15 show the three cluster pairs with most significant difference in link probability between the states Baseline and Left along with a corresponding boxplot, showing the difference. On the boxplots, the central mark is the median, the edges of the box are the 25th and 75th percentiles, the whiskers extend to the most extreme datapoints considered not to be outliers, and the outliers are plotted individually. Similar Figures are shown for comparing the states Baseline and Right (Figures 4.16-4.18) and the states Left and Right (Figures 4.19-4.21). The progress of the clusters throughout the brain is shown in Appendix B. The number of cluster pairs (noc) with significance difference between the states is listed in Table 4.3.

Figure 4.13(a) shows the two clusters with the most significant ($p = 0.001$) difference in link probability between the states Baseline and Left. The first cluster (red) contains voxels from left cortex and left striatum and the second cluster (blue) contains voxels from right striatum only. Figure 4.13(b) shows that there is a higher probability for a link between the two clusters in the state Left than in the state Baseline, but also that there is greater deviation between subjects in the state Left than in the state Baseline.

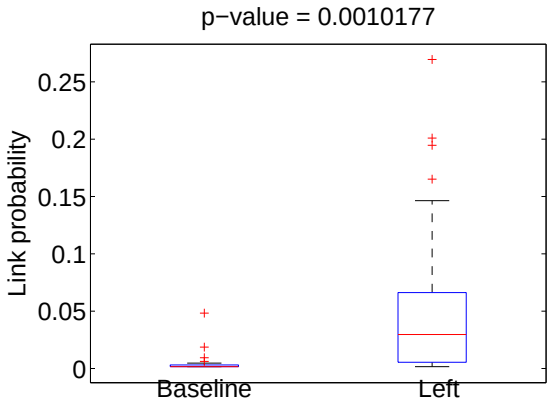
States	noc, $p = 0.05$	noc, $p = 0.01$
Baseline vs Left	49	5
Baseline vs Right	60	15
Left vs Right	30	7

Table 4.3: The number of cluster pairs (noc) with significant difference in link probability between the different states.

Figure 4.14(a) shows the two clusters with the second most significant ($p = 0.0032$) difference in link probability between the states Baseline and Left. The first cluster (red) contains voxels from left cortex and the second cluster (blue) contains voxels from left cortex and right striatum. Figure 4.14(b) shows that there is a little higher probability for a link between the two clusters in the state Left than in the state Baseline, but again the deviation across subjects are greater in the state Left.
Figure 4.15(a) shows the two clusters with the third most significant ($p = 0.0064$) difference in link probability between the states Baseline and Left. The first cluster (red) contains voxels from left cortex and the second cluster (blue) contains voxels from right striatum. Figure 4.14(b) shows that there is a higher probability for a link between the two clusters in the state Left than in the state Baseline, and that the deviation across subjects is greater in the state Left.
In summary, all the three pair of clusters with most significant difference in link probability between the state Baseline and the state Left have one cluster in left cortex and one cluster in the right posterior putamen. The probability for a link between left cortex and the right posterior putamen is higher in the state Left, than in the state Baseline.

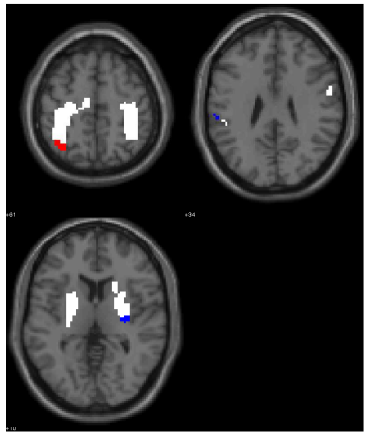


(a) The clusters superimposed on a brain template (anatomical MRI). The white regions indicate the ROIs and the red and blue regions mark the clusters.

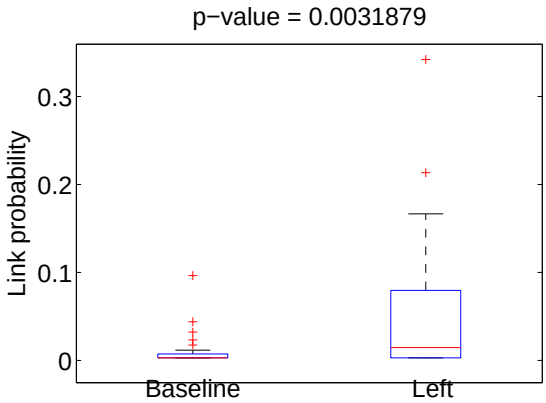


(b) The difference in link probability between the states. The p-value is given in the title.

Figure 4.13: The cluster pair with the most significant difference between the states Baseline and Left.

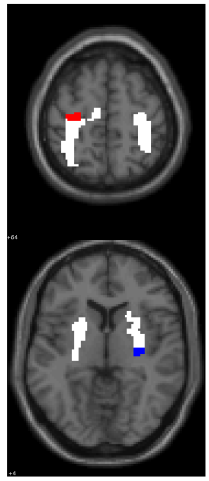


(a) The clusters superimposed on a brain template (anatomical MRI). The white regions indicate the ROIs and the red and blue regions mark the clusters.

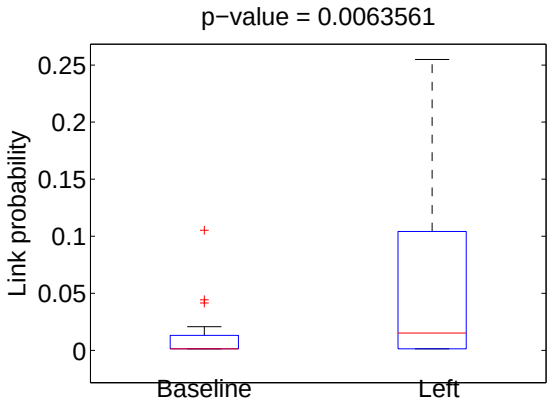


(b) The difference in link probability between the states. The p-value is given in the title.

Figure 4.14: The cluster pair with the second most significant difference between the states Baseline and Left.



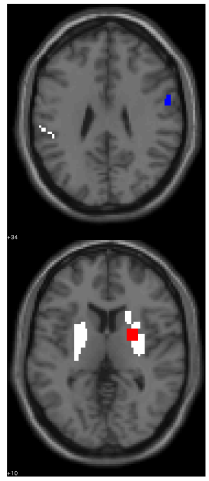
(a) The clusters superimposed on a brain template (anatomical MRI). The white regions indicate the ROIs and the red and blue regions mark the clusters.



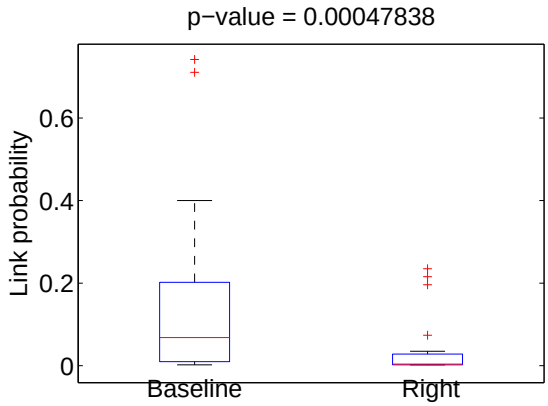
(b) The difference in link probability between the states. The p-value is given in the title.

Figure 4.15: The cluster pair with the third most significant difference between the states Baseline and Left.

Figure 4.16(a) shows the two clusters with the most significant ($p = 0.00048$) difference in link probability between the states Baseline and Right. The first cluster (red) contains voxels from right striatum and the second cluster (blue) contains voxels from right cortex. Figure 4.16(b) shows that there is a higher probability for a link between the two clusters in the state Baseline than in the state Right.
Figure 4.17(a) shows the two clusters with the second most significant ($p = 0.0013$) difference in link probability between the states Baseline and Right. The first cluster (red) contains voxels from left cortex and the second cluster (blue) contains voxels from left cortex and left striatum. Figure 4.17(b) shows that the median has the same probability for a link between the two clusters in both states, but the deviation across subjects are greater in the state Right.
Figure 4.18(a) shows the two clusters with the third most significant ($p = 0.0028$) difference in link probability between the states Baseline and right. The first cluster (red) contains voxels from right cortex and the second cluster (blue) contains voxels from left striatum. Figure 4.16(b) shows that there is a higher probability for a link between the two clusters in the state Baseline than in the state Right.

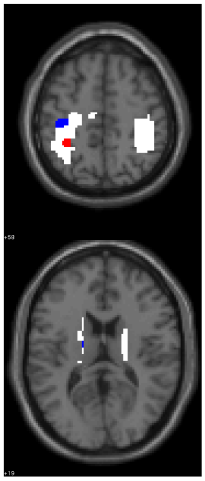


(a) The clusters superimposed on a brain template (anatomical MRI). The white regions indicate the ROIs and the red and blue regions mark the clusters.

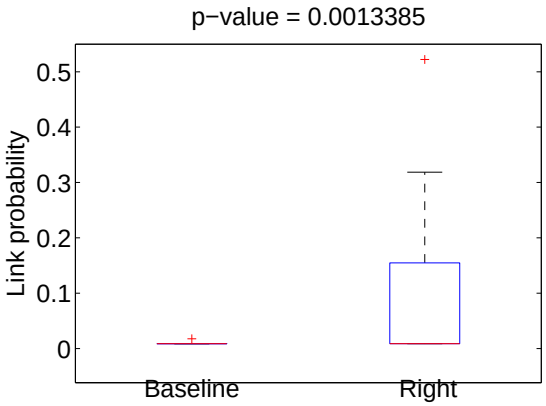


(b) The difference in link probability between the states. The p-value is given in the title.

Figure 4.16: The cluster pair with the most significant difference between the states Baseline and Right.

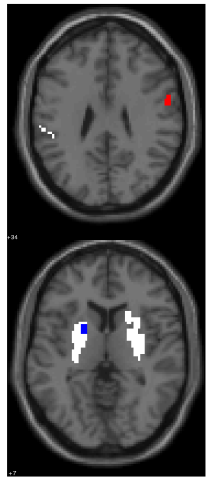


(a) The clusters superimposed on a brain template (anatomical MRI). The white regions indicate the ROIs and the red and blue regions mark the clusters.

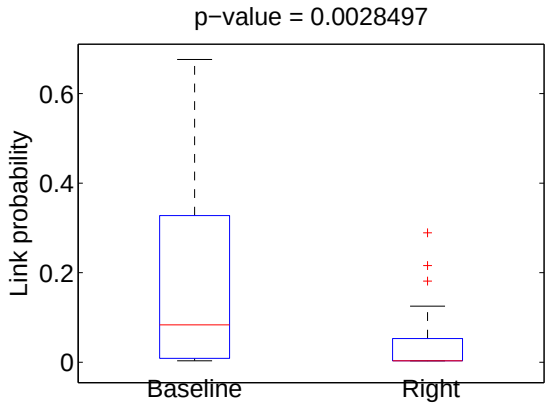


(b) The difference in link probability between the states. The p-value is given in the title.

Figure 4.17: The cluster pair with the second most significant difference between the states Baseline and Right.



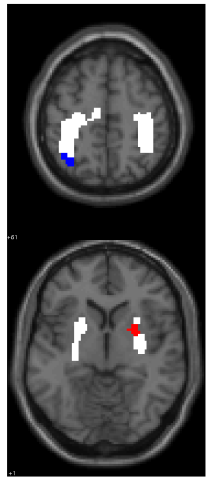
(a) The clusters superimposed on a brain template (anatomical MRI). The white regions indicate the ROIs and the red and blue regions mark the clusters.



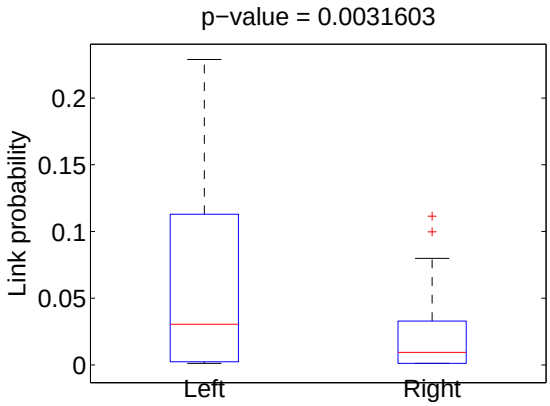
(b) The difference in link probability between the states. The p-value is given in the title.

Figure 4.18: The cluster pair with the third most significant difference between the states Baseline and Right.

Figure 4.19(a) shows the two clusters with the most significant ($p = 0.0032$) difference in link probability between the states Left and Right. The first cluster (red) contains voxels from right striatum and the second cluster (blue) contains voxels from left cortex. Figure 4.19(b) shows that there is a higher probability for a link between the two clusters in the state Left than in the state Right.
Figure 4.20(a) shows the two clusters with the second most significant ($p = 0.0044$) difference in link probability between the states Left and right. The first cluster (red) contains voxels from left cortex and left striatum and the second cluster (blue) contains voxels from right striatum. Figure 4.20(b) shows that there is a higher probability for a link between the two clusters in the state Left than in the state Right, but also that the deviation across subjects is greater in the state Left.
Figure 4.20(a) shows the two clusters with the third most significant ($p = 0.0056$) difference in link probability between the states Left and Right. The first cluster (red) contains voxels from right cortex and left striatum and the second cluster (blue) contains voxels from right striatum. Figure 4.21(b) shows that there is a higher probability for a link between the two clusters in state Right than in the state Left.

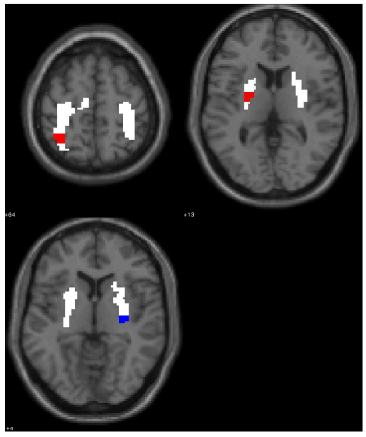


(a) The clusters superimposed on a brain template (anatomical MRI). The white regions indicate the ROIs and the red and blue regions mark the clusters.

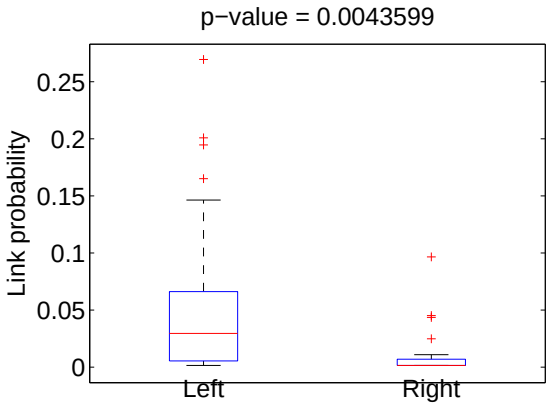


(b) The difference in link probability between the states. The p-value is given in the title.

Figure 4.19: The cluster pair with the most significant difference between the states Left and Right.

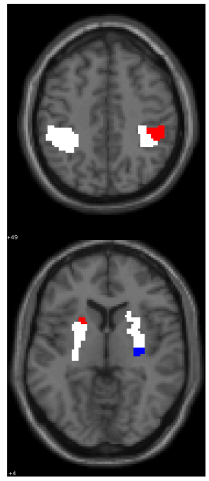


(a) The clusters superimposed on a brain template (anatomical MRI). The white regions indicate the ROIs and the red and blue regions mark the clusters.

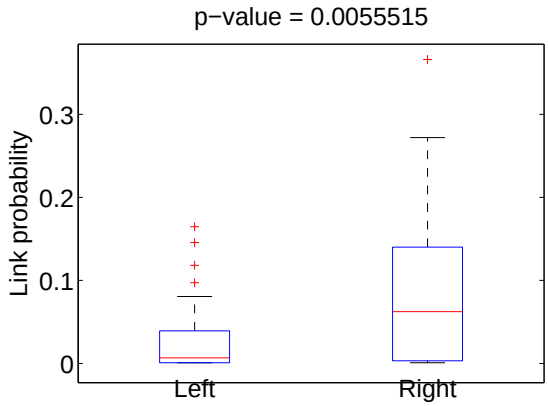


(b) The difference in link probability between the states. The p-value is given in the title.

Figure 4.20: The cluster pair with the second most significant difference between the states Left and Right.



(a) The clusters superimposed on a brain template (anatomical MRI). The white regions indicate the ROIs and the red and blue regions mark the clusters.



(b) The difference in link probability between the states. The p-value is given in the title.

Figure 4.21: The cluster pair with the third most significant difference between the states Left and Right.

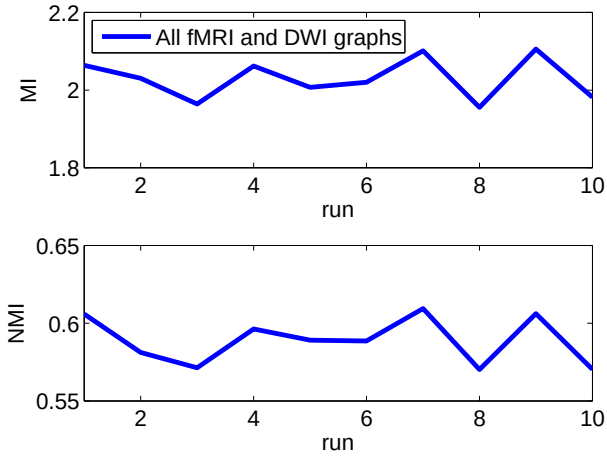


Figure 4.22: MI and NMI between splits when all the fMRI and DWI graphs are inferred simultaneously.

4.3 Combined fMRI and DWI Data

Figure 4.22 shows the MI and the NMI between splits for each of the ten runs when all fMRI and DWI graphs are inferred simultaneously.
Figure 4.23 shows the MI, the NMI, and the test log likelihood between splits for each of the ten runs when the graphs for the three fMRI states and the DWI graphs are inferred separately. All the measures are stable across the ten runs for all the states. The MI is higher for the fMRI states than for the DWI graphs, but the NMI and the test log likelihood are very similar for all the four states.

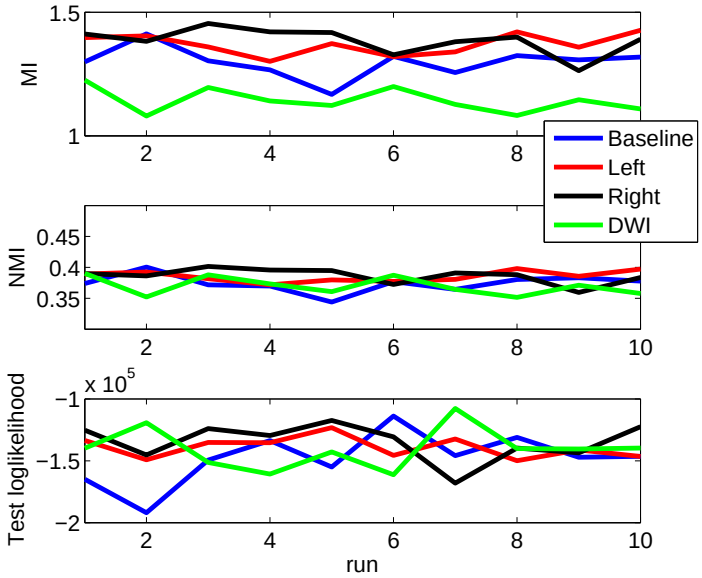


Figure 4.23: MI, NMI and test log likelihood between splits when the fMRI graphs from the different states and the DWI graphs are inferred separately.

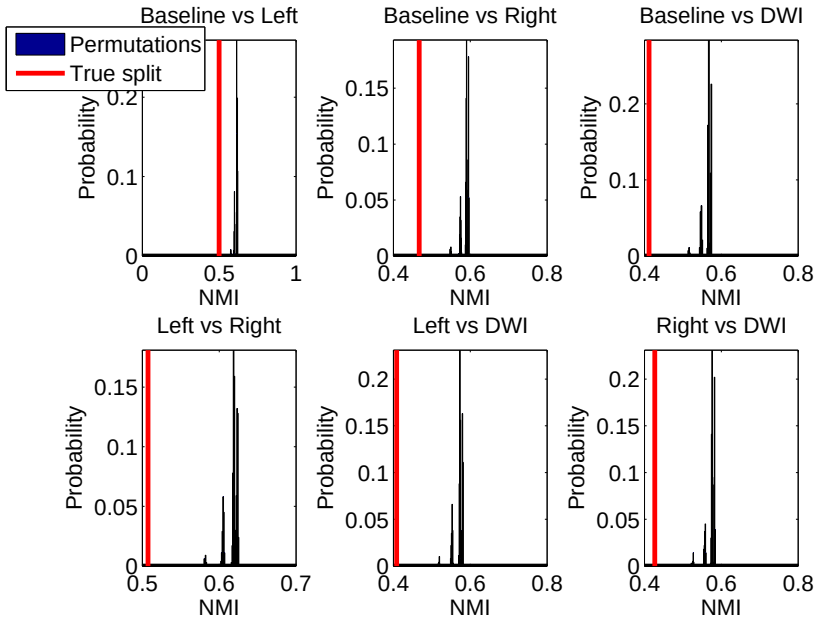


Figure 4.24: The result of the permutation test for testing if the same clusters are found in the different states.

4.3.1 The Clustering in the Different States

Figure 4.24 shows the result of the permutation test when testing if the same clusters are found in the different states. For all combinations the permutations have higher NMI than the true split, meaning that the clustering is more similar within the states than between the states. With this result the null hypothesis (same clustering in all states) can be rejected at a significance level of $p = 0.001$. The NMI between the three fMRI states and DWI (the true split in the top right, the center, and the right bottom in Figure 4.24) are a little smaller than the NMI between the fMRI states, meaning that the clustering in the three fMRI states is more similar to each other than to the clustering in the DWI graphs.

To investigate the differences in the clustering the number of clusters (MAP solution) in each ROI for each of the four states are found and listed in Table 4.4. Furthermore the three biggest clusters are visualised for each of the four states in Figure 4.25. The progress of the clusters throughout the brain is shown

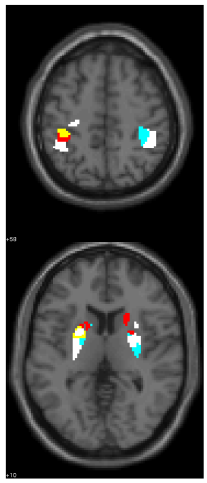
State	Left cortex	Right cortex	Left striatum	Right striatum
Baseline	13	12	16	21
Left	15	7	20	20
Right	18	12	21	17
DWI	10	16	16	13

Table 4.4: The number of cluster found in each of the four ROIs in each of the four states.

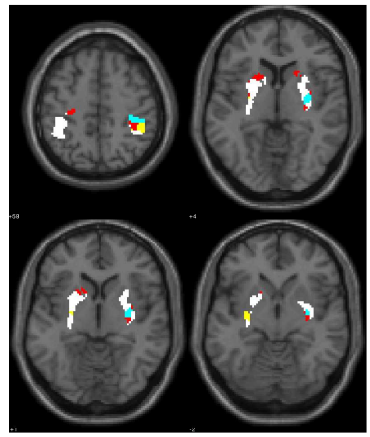
Cluster	Baseline	Left	Right	DWI
Red	L cor + L str + R str	All	All	All
Cyan	R cor + L str + R str	R cor + R str	L cor + L str	R cor + R str
Yellow	L cor + L str	R cor + L str	L cor + L str	R cor + R str

Table 4.5: Listing of which region the voxels in the three biggest clusters belongs to in the four different states. L cor indicates left cortex, while R str indicates right striatum.

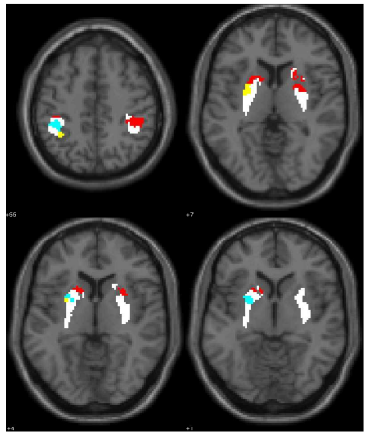
in Appendix B. An overview of which ROI, the voxels contained in the clusters belong to, is given in Table 4.5.



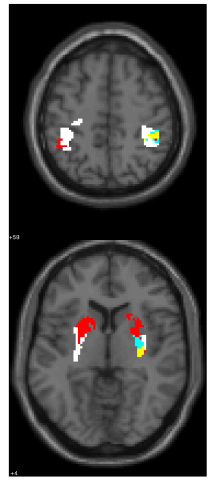
(a) Baseline



(b) Left



(c) Right



(d) DWI

Figure 4.25: The three biggest clusters found in the state Baseline (a), Left (b), Right (c), and (d) DWI. The order of the clusters regarding their size in descending order is red, cyan, and yellow. The white regions mark the ROIs.

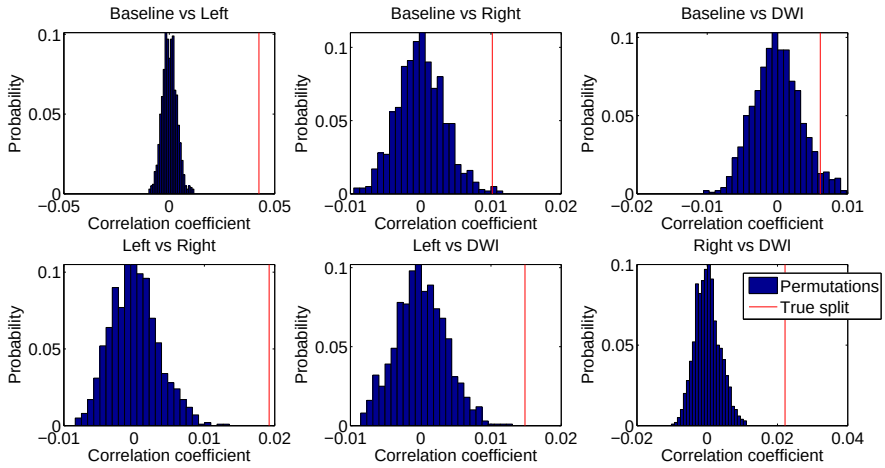


Figure 4.26: The result of the permutation test for testing if the communication pattern between states is correlated more than random.

4.3.2 The Communication pattern in the Different States

Figure 4.26 shows the result of the permutation test for testing if the communication pattern between the states is correlated more than random. It shows that the states Baseline and Left, Left and Right, Left and DWI, and Right and DWI are more correlated than random at a significance level of $p = 0.001$. The states Baseline and Right are more correlated than random at a significance level of $p = \frac{5+1}{1000+1} = 0.006$ and the states Baseline and DWI are more correlated than random at a significance level of $p = \frac{43+1}{1000+1} = 0.044$.

To investigate if there is different communication in the four states, 1000 adjacency matrices for each subject and state are generated using the MAP solution across the ten runs. The test log likelihood for the generated adjacency matrices for each state is calculated. A histogram of the test log likelihoods for each state is made. Figure 4.27-4.30 shows the histogram of the test log likelihoods when the adjacency matrices are generated using the link probability matrix from state Baseline, Left, Right, and DWI, respectively. If each adjacency matrices are classified as the state with the biggest log likelihood, all the matrices are correctly classified.

The number of cluster pairs with significant difference in link probability vary a lot across states. The significant cluster pairs are found using a paired t-test.

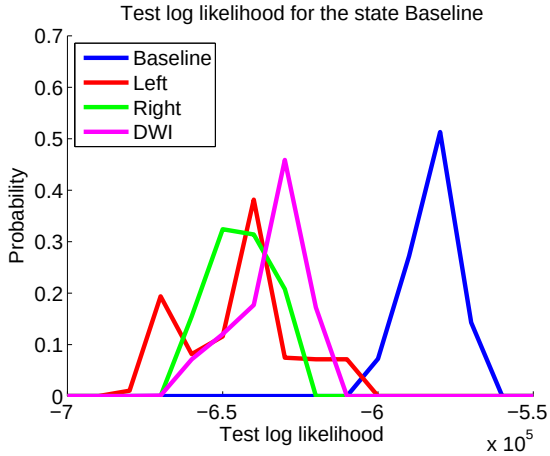


Figure 4.27: Histogram of the test log likelihoods for the four states when the adjacency matrices are generated using the link probability matrix from state Baseline.

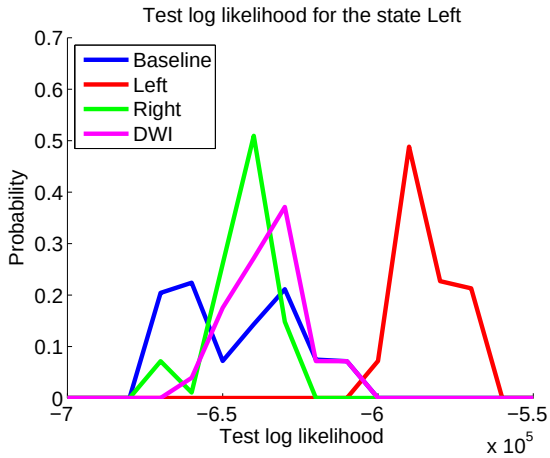


Figure 4.28: Histogram of the test log likelihoods for the four states when the adjacency matrices are generated using the link probability matrix from state Left.

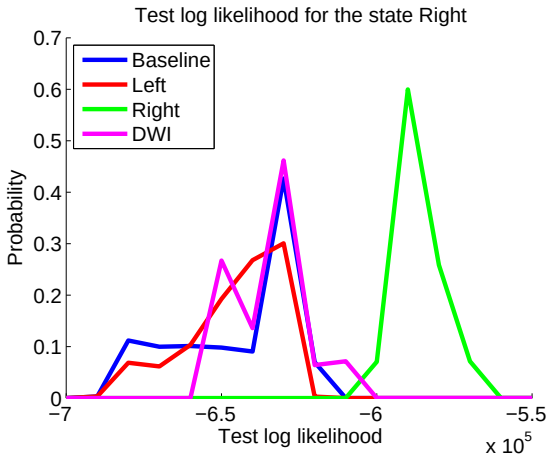


Figure 4.29: Histogram of the test log likelihoods for the four states when the adjacency matrices are generated using the link probability matrix from state Right.

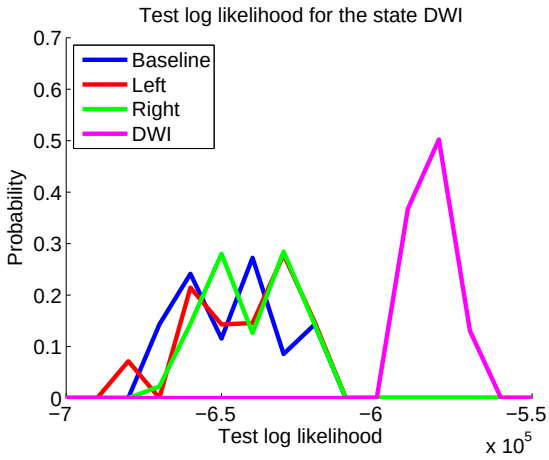


Figure 4.30: Histogram of the test log likelihoods for the four states when the adjacency matrices are generated using the link probability matrix from state DWI.

States	noc
Baseline vs Left	0
Baseline vs Right	2
Baseline vs DWI	25
Left vs Right	1
Left vs DWI	32
Right vs DWI	21

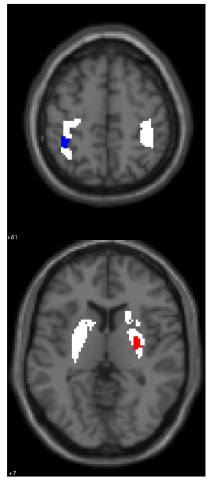
Table 4.6: The number of cluster pairs (noc) with significant difference ($p = 0.05$) in link probability between the different states.

For a significance level $p = 0.05$ the number of cluster pairs (noc) is listed in Table 4.6. The three most significant cluster pairs for the different states (if they exist) are visualised in Figure 4.31-4.39 together with boxplots showing the difference.

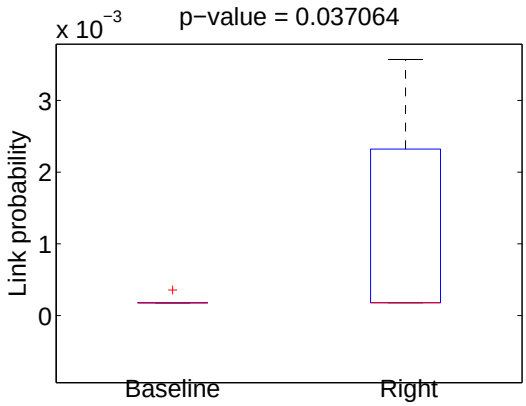
Figure 4.31(a) shows the cluster pair with the most significant difference in link probability between the states Baseline and Right. The first cluster (red) contains voxels from right striatum and the second cluster (blue) contains voxels from left cortex. Figure 4.31(b) shows the difference in the link probability. The median are almost the same in the two states, but the deviation is greater in the state Right than in the state Baseline. For both states the link probability is small.
Figure 4.32(a) shows the cluster pair with the second most significant difference in link probability between the states Baseline and Left. The first cluster (red) contains voxels from right striatum and the second cluster (blue) contains voxels from left cortex. Both clusters are small, containing only 12 voxels each. Figure 4.32(b) shows that the median link probability in the two states are almost the same, but the deviation across the subjects is bigger in the state Right. Again the link probability is quite small.

Figure 4.33(a) shows the only cluster pair with significant difference in link probability between the states Left and Right. The first cluster (red) contains voxels from left cortex and right striatum and the second cluster (cyan) contains voxels from the left posterior putamen. Figure 4.33(b) shows that there is a higher probability for a link between the two clusters in the state Left than in the state Right, and also that the variation across subjects is greater in the state Left.

Figure 4.34(a) shows the cluster pair with the most significant difference in the

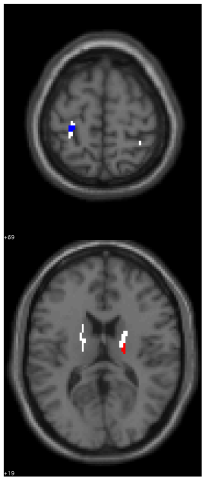


(a) The clusters superimposed on a brain template (anatomical MRI). The white regions indicate the ROIs and the red and blue regions mark the clusters.

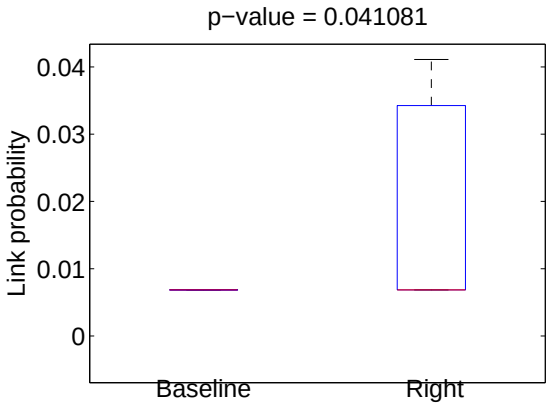


(b) The difference in link probability between the states. The p-value is given in the title.

Figure 4.31: The cluster pair with the most significant difference between the states Baseline and Right.

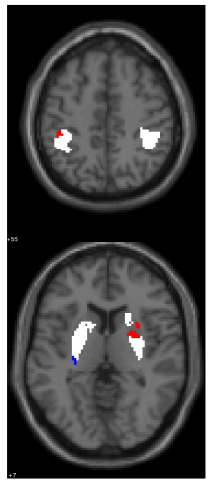


(a) The clusters superimposed on a brain template (anatomical MRI). The white regions indicate the ROIs and the red and blue regions mark the clusters.

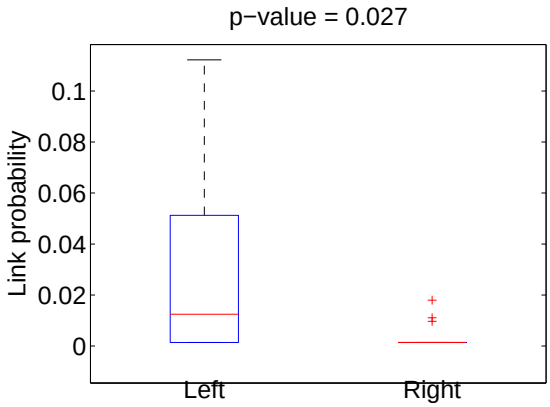


(b) The difference in link probability between the states. The p-value is given in the title.

Figure 4.32: The cluster pair with the second most significant difference between the states Baseline and Right.



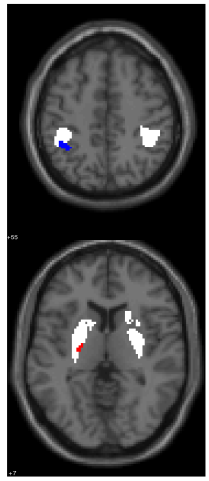
(a) The clusters superimposed on a brain. The white regions indicate the ROIs and the red and blue regions mark the clusters.



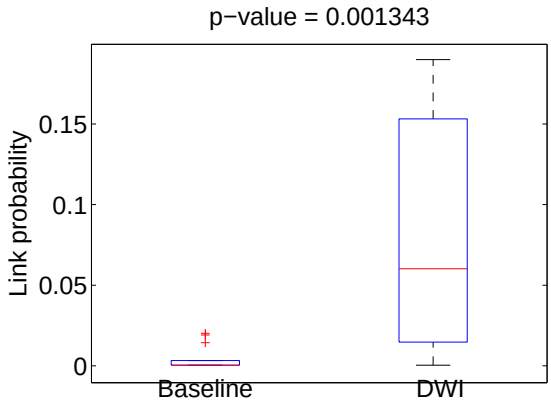
(b) The difference in link probability between the states. The p-value is given in the title.

Figure 4.33: The cluster pair with the most significant difference between the states Left and Right.

link probability between the states Baseline and DWI. The first cluster (red) contains voxels from left striatum and the second cluster contains voxels from left cortex. Figure 4.34(b) shows that there is highest probability for a link between the two clusters in the state DWI. It also shows that the probability is almost zero for all subjects in the state Baseline, while there is greater variation across subjects in the state DWI.
Figure 4.35(a) shows the cluster pair with the second most significant difference in the link probability between state Baseline and DWI. The first cluster (red) contains voxels from left cortex and left striatum and the second cluster (cyan) contains voxels from left striatum. Figure 4.35(b) shows that there is a higher link probability and a greater variation across subjects in the state DWI than in the state Baseline.
Figure 4.36(a) shows the cluster pair with the third most significant difference in the link probability between the states Baseline and DWI. The first cluster (red) contains voxels from right striatum and the second cluster (cyan) contains voxels from right cortex. Figure 4.36(b) shows that there is a higher link probability and a greater variation across subjects in the state DWI than in the state Baseline.

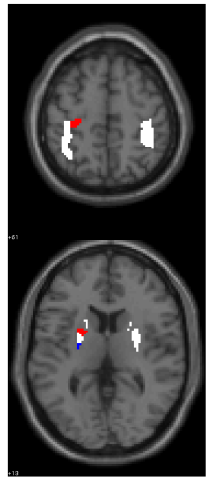


(a) The clusters superimposed on a brain template (anatomical MRI). The white regions indicate the ROIs and the red and blue regions mark the clusters.

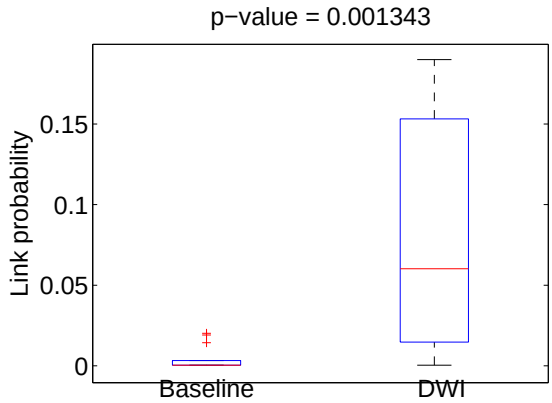


(b) The difference in link probability between the states. The p-value is given in the title.

Figure 4.34: The cluster pair with the most significant difference between the states Baseline and DWI.

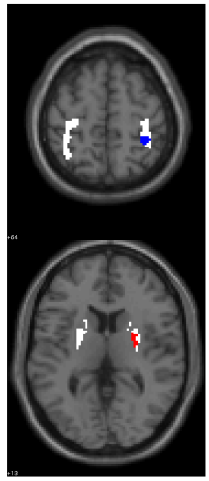


(a) The clusters superimposed on a brain template (anatomical MRI). The white regions indicate the ROIs and the red and blue regions mark the clusters.

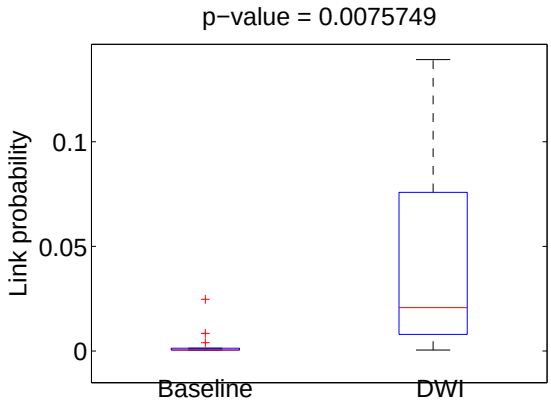


(b) The difference in link probability between the states. The p-value is given in the title.

Figure 4.35: The cluster pair with the second most significant difference between the states Baseline and DWI.



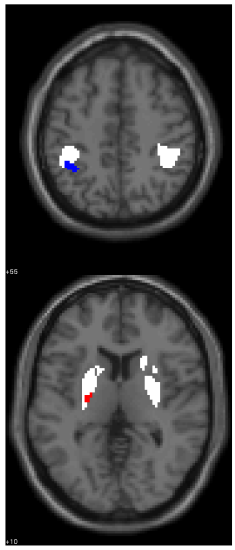
(a) The clusters superimposed on a brain template (anatomical MRI). The white regions indicate the ROIs and the red and blue regions mark the clusters.



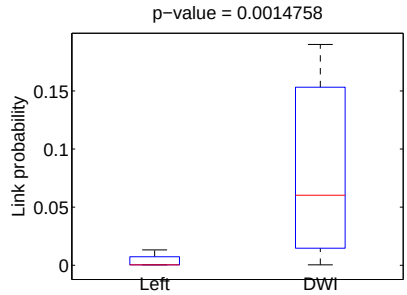
(b) The difference in link probability between the states. The p-value is given in the title.

Figure 4.36: The cluster pair with the third most significant difference between the states Baseline and DWI.

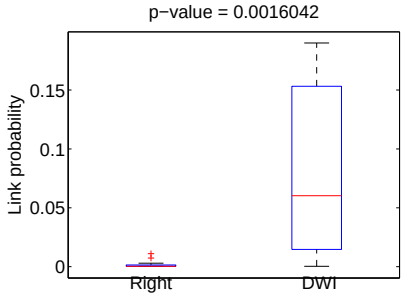
The three cluster pairs with most significant difference in the link probability between the states Left and DWI are the same as between the states Right and DWI. The three cluster pairs and the differences are shown in Figures 4.37-4.39. Figure 4.37(a) shows the cluster pair with the most significant difference in the link probability. The clusters contain voxels from left cortex and left striatum, respectively. Figures 4.37(b) and 4.37(c) show that the link probability is higher and that there is a greater variation across the subjects in the state DWI than in the states Left and Right.
Figure 4.38(a) shows the cluster pair with the second most significant difference in the link probability. The clusters contain voxels from right cortex and right striatum, respectively. Figure 4.38 shows that the link probability is higher and that there is a greater variation across the subjects in the state DWI than in the states Left and Right.
Figure 4.39(a) shows the cluster pair with the third most significant difference in the link probability between the states Left and DWI. The clusters contain voxels from right cortex and right striatum, respectively. Figure 4.39(b) shows that the link probability is higher and that there is greater variation across the subjects in the state DWI than in the state Left.



(a) The clusters superimposed on a brain template (anatomical MRI). The white regions indicate the ROIs and the red and blue regions mark the clusters.

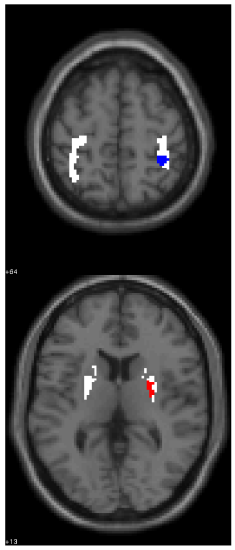


(b) Left vs. DWI

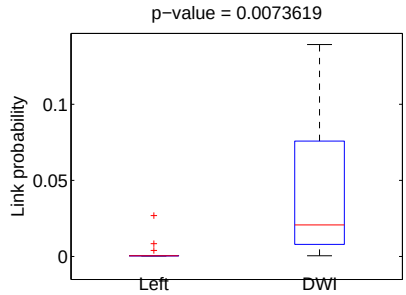


(c) Right vs. DWI

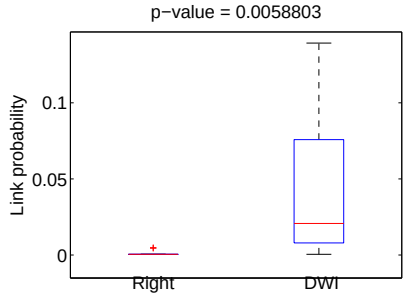
Figure 4.37: (a) The cluster pair with the most significant difference between the states Baseline and DWI and the difference in link probability across subjects between (b) Left and DWI and (c) Right and DWI.



(a) The clusters superimposed on a brain template (anatomical MRI). The white regions indicate the ROIs and the red and blue regions mark the clusters.

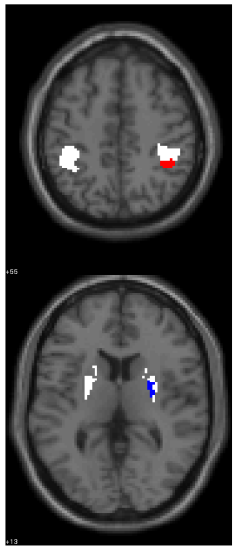


(b) Left vs. DWI

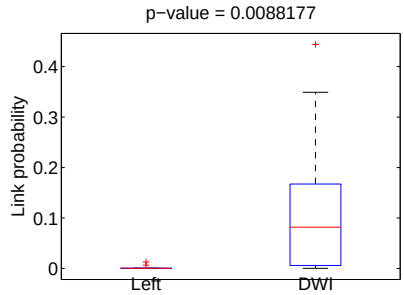


(c) Right vs. DWI

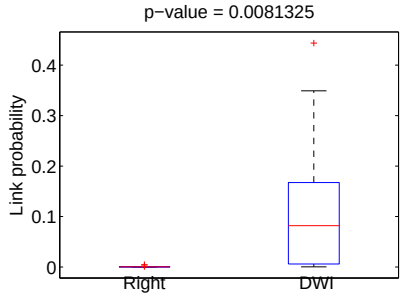
Figure 4.38: (a) The cluster pair with the second most significant difference between the states Baseline and DWI and the difference in link probability across subjects between (b) Left and DWI and (c) Right and DWI.



(a) The clusters superimposed on a brain template (anatomical MRI). The white regions indicate the ROIs and the red and blue regions mark the clusters.



(b) Left vs. DWI



(c) Right vs. DWI

Figure 4.39: (a) The cluster pair with the third most significant difference between the states Baseline and DWI and the difference in link probability across subjects between (b) Left and DWI and (c) Right and DWI.

Chapter 5

Discussion

5.1 Reliability of the IRM

To test the robustness of the IRM, each dataset was randomly split in two and the MI, the NMI, and the test log likelihood between the splits were calculated. The MI, the NMI, and test log likelihoods are found to be stable across runs (Figures 4.6, 4.5, 4.23, and 4.22), indicating that the IRM is stable and reliable. This is consistent with the results obtained for the synthetic data in section 2.12, where it was found that the IRM has both good reproducibility and good predictability.

The NMI is lower when inferring the fMRI graphs for the three states separately than when all the fMRI graphs are inferred simultaneously, see Figure 4.5 and 4.6. The same apply when inferring the fMRI and DWI graphs. This is due to the amount of data. The more data the IRM has to detect the structure, the better it is at reproducing the result. When all the fMRI graphs are inferred at once the IRM has 87 graphs, while only 29 graphs are available when inferring the three states separately. When inferring all the fMRI and DWI graphs the IRM has 56 graphs, and 14 graphs only when inferring the four states separately.

5.2 The Clustering

When inferring the states separately, we find that the clustering is different between the states (Figures 4.7 and 4.24). The anatomy of the brain is not changing during the acquisition of the data, only the activation of the functional modules. Thus, one would expect the IRM to find the same clustering at least in the three fMRI states. That the clustering is not the same in the fMRI states and the DWI data is more understandable, since the voxels activated simultaneously in the fMRI data do not necessarily have to be directly connected by a streamline. As the IRM does not find the same clustering, it indicates that the IRM is highly sensitive and able to adapt even to small differences in the data.

The IRM finds few clusters compared to the number of nodes in the graphs and is consistent in the number of clusters across the states. As the two extreme scenarios it could have found one cluster only, or a cluster for each voxel, but it finds between 9 and 23 clusters in each mask and in total 65-66 clusters in each fMRI state (Table 4.2).
For the combined fMRI and DWI data it finds between 7 and 21 clusters in each mask and 55-68 clusters in each state in total (Table 4.4).
That different clustering is found in the different states indicates that the communication pattern changes when the functional modules are activated.

When looking at the clustering in the analysis of the fMRI data only (Figure 4.8 and Table 4.1), it can be seen that the biggest cluster (red) in the two active states, Left and Right, contains voxels both from right cortex and right striatum, while it only contains voxels from left cortex in the state Baseline. This could indicate that there is more communication between cortex and striatum in the active states than in Baseline. On the other hand, the second biggest cluster (cyan) contains voxels from both left cortex and right striatum in the states Baseline and Right and only from right cortex in the state Left.
Since some of the clusters only differ between states with a few voxels, it is difficult to interpret the differences in the clustering.

When looking at the clustering in the combined analysis of the fMRI and DWI data (Figure 4.25 and Table 4.5), it can be seen that the biggest cluster (red) contains multiple single voxels from all masks in all four states, except from right cortex in the state Baseline.
The second biggest cluster (cyan) is in right cortex and striatum in the state Left and in left cortex and striatum in the state Right. This makes sense since

the motor cortex and the striatum is contralateral, meaning that left cortex and striatum are more activated when performing right finger tapping compared to left finger tapping and vice versa.

In summary it can be concluded that the IRM is highly sensitive to even small differences in the data, since the structure found by the IRM is sensitive to the task performed. At the same time the IRM has found a structure that is consistent across the states including DWI, since the number of clusters is stable across the states.

5.3 The Communication Pattern

The IRM is forced to have the same clustering for the different states to be able to compare the link probability matrices. By generating new graphs with the estimated clustering and link probabilities, and calculating the test log likelihood, we find that the communication pattern differs between the states (Figures 4.10-4.12 and 4.27-4.30).

5.3.1 fMRI Data

The most significant difference between the states Baseline and Left, is that a cluster in the right posterior putamen, which is motor striatum[22], is more connected to three different clusters in the left motor cortex, when performing left finger tapping compared to Baseline (Figures 4.13-4.15).
The most significant difference between the states Baseline and Right, is that a small cluster in the right inferior cortex is more connected to the right and the left putamen, respectively, in the state Baseline than in the state Right (Figures 4.16 and 4.18). In [32] it is found that the more complex the sequence of finger tapping becomes, the more bilateral is the activation of the putamen, and the region of the putamen activated grows with the complexity. Therefore, I would expect the clusters found to be more connected in the state Right than in Baseline. On the other hand, right cortex is only activated to a small degree when performing right finger tapping. Another significant difference between the states Baseline and Right is that the clusters in left cortex and left striatum are more connected in the state Right than in the state Baseline. This makes sense, since left cortex is highly activated during right finger tapping.

The most significant differences betweeen the states Left and Right are that left cortex and striatum are more connected in the state Left, while right cortex and striatum are more connected in the state Right. The reason why this result does not show the opposite could be that the clusters found are connected because neither of them are activated in the respective task.

Some of the cluster pairs make sense as in Figure 4.16(a), where two clusters in left cortex are more connected when performing right finger tapping than during baseline. It makes sense since, it is known that the left motor cortex is activated during right finger tapping. On the other hand some cluster pairs are hard to interpret, e.g., Figure 4.19 shows that a cluster in left cortex and the left putamen are more connected with a cluster in the right posterior putamen (motor striatum), when performing left finger tapping compared to performing right finger tapping. Offhand, I would expect right cortex to be more connected when performing left finger tapping compared to right finger tapping, because the right motor cortex is more activated during left finger tapping than during right finger tapping.
That a cluster pair is significantly different between two states does not necessarily mean that the clusters are highly activated during the task, it just suggests that the probability of a link between the clusters is significantly different between the two states. This could be due to one of the clusters is highly activated in one of the states and not in the other, and the other cluster in the cluster pair does not change its activation between states. In such a case it makes sense that the probability of a link between two clusters differs significantly.

5.3.2 Combined fMRI and DWI Data

The difference between DWI and the fMRI states is that there is a higher connection between cortex and striatum within a hemisphere in DWI than in the fMRI states. These could be regions in the cortex and the striatum, which are structural connected, but not both activated when performing finger tapping or rest.

I expected a correlation between the areas activated in the fMRI and the areas structurally connected in the tractography. The results show that neither the clustering nor the link probabilities between the clusters between the different states are similar.
A reason why there is no such connection can be that the graphs used in the analysis only have a link density of 0.5%. Perhaps, there are not streamlines

enough between the investigated voxels to reveal such a connection. To examine this hypothesis, analysis of different graph densities are needed. The reason why the graph density of only 0.5% is used in this project, is that there were not that many streamlines between the cortex and the striatum and choosing a higher link density would result in a reduction of subjects included in the analysis. This may be overcome by making more samples in the tractography.
Another thing to have in mind is the voxelsize. When the voxelsize is small, like $2 \times 2 \times 2$ mm^3, as in the combined analysis of the fMRI and DWI graphs, it is less probable for a streamline to reach that exact voxel, while larger voxels are more likely to be reached. On the other hand a $1 \times 1 \times 1$ mm^3 voxel contains hundreds of thousands of axonal fibres, so to model the anatomy of the brain more closely the voxelsize should be small.
One could expect that the link probability matrices for the DWI graphs and the state Baseline were more similar compared to the link probability matrices for the DWI graphs and the states Left and Right, because in neither of the states Baseline and DWI there are asymmetries or activations induced by a task. In the states Left and Right there is a reason why the regions are activated, since a motor task is solved. A reason for lack of similarity between the two types of data can be due to the different time scales.

In summary, the IRM is highly sensitive to the link probability matrix, as the graphs generated using the connectivity matrices can be correctly classified as either Baseline, Left, Right, or DWI. This also means that there is a significant difference between the four states. The results also indicate that there is a coordination between cortex and striatum, since clusters in the cortex and the striatum are connected.

Chapter 6

Conclusions

In this project, the IRM is described and its predictability and reproducibility is investigated using synthetic data generated by four different models. It is found that the predictability and the reliability of the IRM are on par with the model that generated the data, even if the data has a simple structure, meaning that the IRM does not overfit the data.
This result is strengthened by the results of the MRI data, where it is found that the IRM is stable across runs.

Analyses of fMRI and DWI data is performed using the IRM. The partition of the motor cortex and the striatum when the subjects are at rest and perform left and right finger tapping, respectively, are compared both for fMRI data alone and for combined fMRI and DWI data.
It is found that the partition of the ROIs is significantly different between the four different states. Even if the partition is significantly different between the states, the differences are hard to interpret by visual inspection. This shows that the IRM is highly sensitive, since it is sensitive to the task performed. Also, the IRM finds a structure which is consistent across states, as the number of clusters only vary a little between the states.

The IRM is forced to find a common partition of the ROIs across the states. The

communication pattern between the clusters, found by the IRM in the different states, is investigated, both for fMRI data alone and for combined fMRI and DWI data. It is found that the communication pattern between the functional modules is significantly different between the four states. This means that the link probability matrix is highly sensitive to the task performed.
The inspection of the cluster pairs with the most significant differences between the states, indicates that there is a coordination between the cortex and the striatum, as clusters in the cortex and the striatum are connected.

The results of the two analyses, fMRI data alone and combined fMRI and DWI data, cannot be directly compared, since the graph density is not the same. Also, the regions in the cortex is not the same, the resolution of the images differs between the analyses, and the number of subjects included in each analysis differs.

To further investigate the partition and the communication pattern of the motor cortex and the striatum, the analysis should be performed using different graph densities. If the IRM is provided more data, it will probably find other structures. Also, the bipartite IRM should be considered instead of the unipartite IRM.
When testing if the communication pattern is the same in the different states, the partition of the ROIs is assumed to be the same. To investigate if this assumption holds, all the graphs could be inferred simultaneously and the link probability matrix could be forced to be the same across subjects within each state. In this way the partition of the ROIs can be compared to the model where the states are inferred separately.

Appendix A

Partition of the Zachary's Karate Dataset

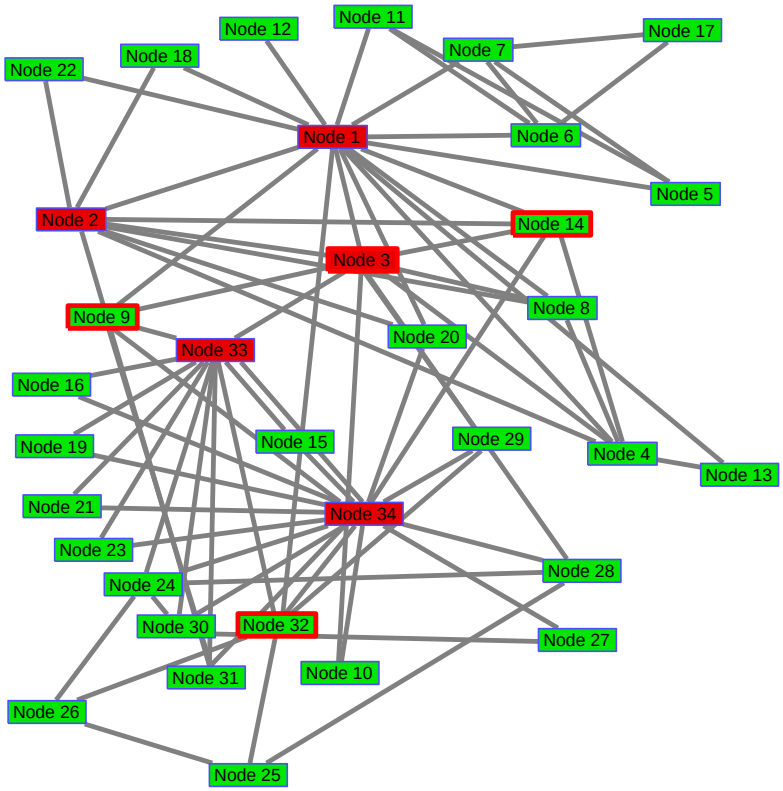


Figure A.1: Illustration of the partition of Zachary's Karate Club dataset inferred with the IRM. Nodes with red edges (3, 9, 14, and 32) denote Hubs[11]. The hyperparameters are $\alpha = 10^{-5}$ and $\boldsymbol{\beta}_+(a,b) = \boldsymbol{\beta}_-(a,b) = [1\ 1]$. Nodes with many links are assigned to the same cluster and the remaining nodes are assigned to the second cluster.

Appendix B

The Progress of the Clusters throughout the Brain

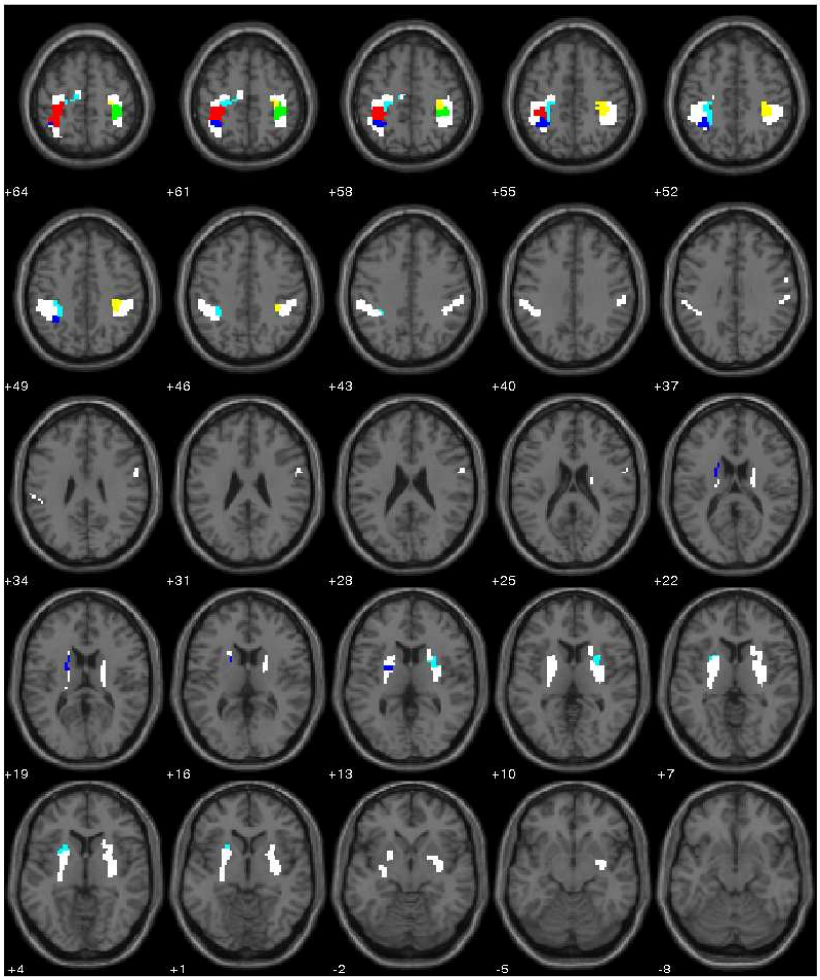


Figure B.1: The five biggest clusters found in the state Baseline. The order of the clusters regarding their size in descending order is red, cyan, green, blue, and yellow. The white regions mark the ROIs.

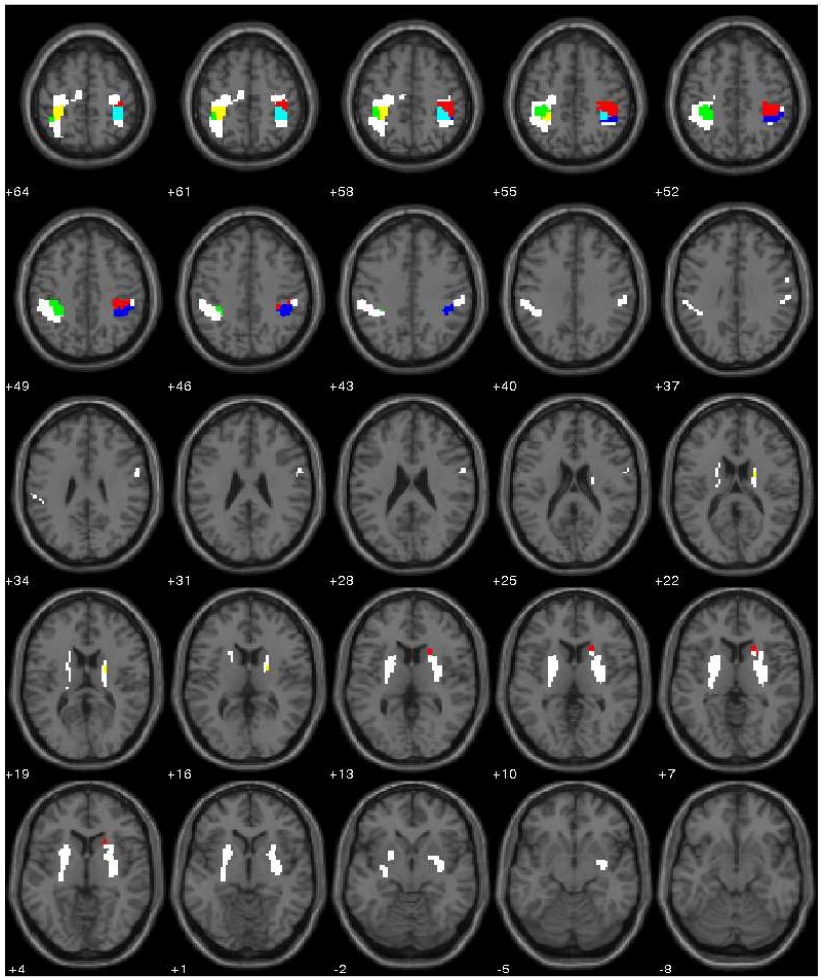


Figure B.2: The five biggest clusters found in the state Left. The order of the clusters regarding their size in descending order is red, cyan, green, blue, and yellow. The white regions mark the ROIs.

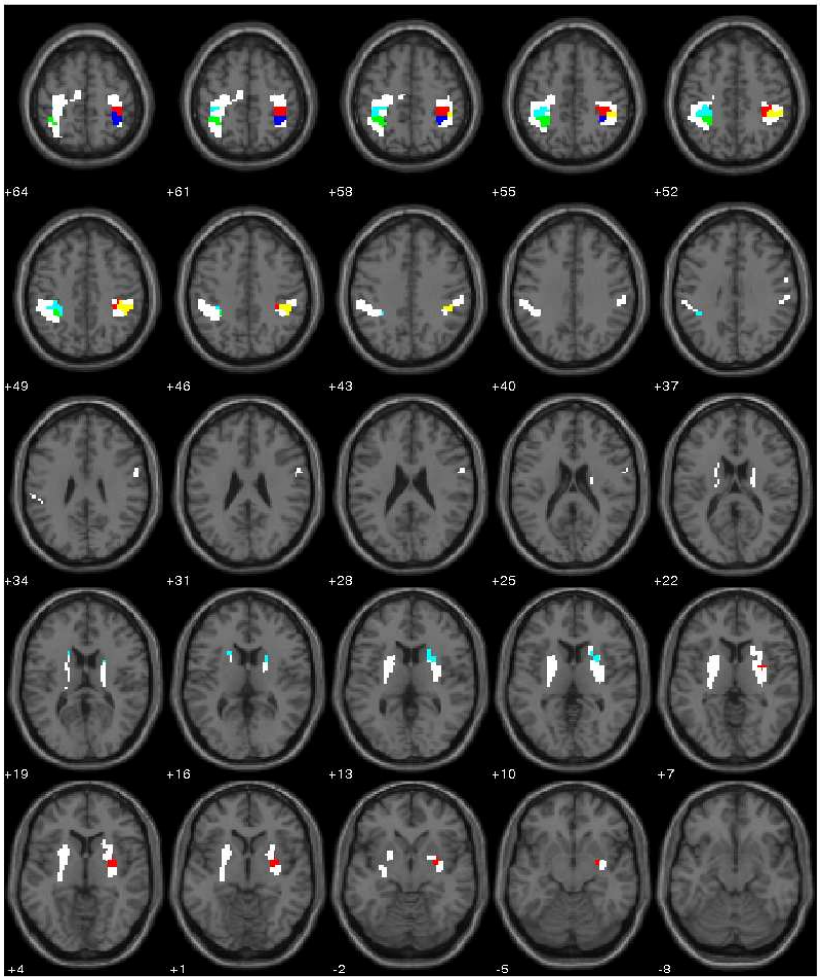


Figure B.3: The five biggest clusters found in the state Right. The order of the clusters regarding their size in descending order is red, cyan, green, blue, and yellow. The white regions mark the ROIs.

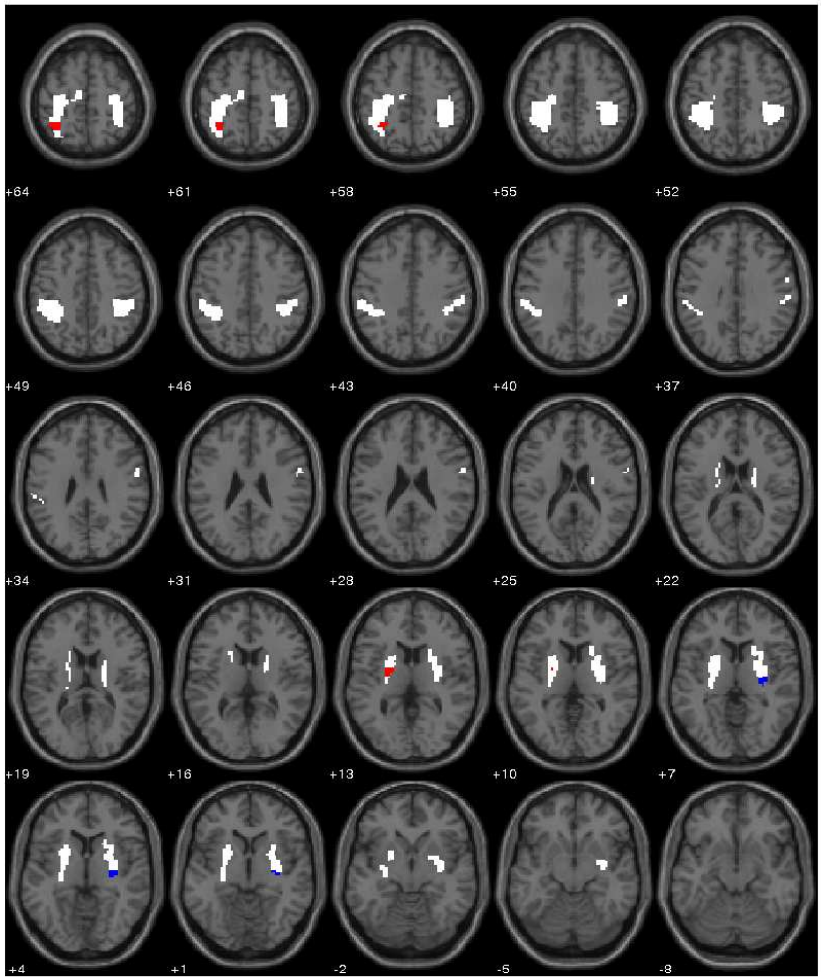


Figure B.4: The cluster pair with the most significant difference between Baseline and Left. The white regions mark the ROIs and the red and blue mark the clusters.

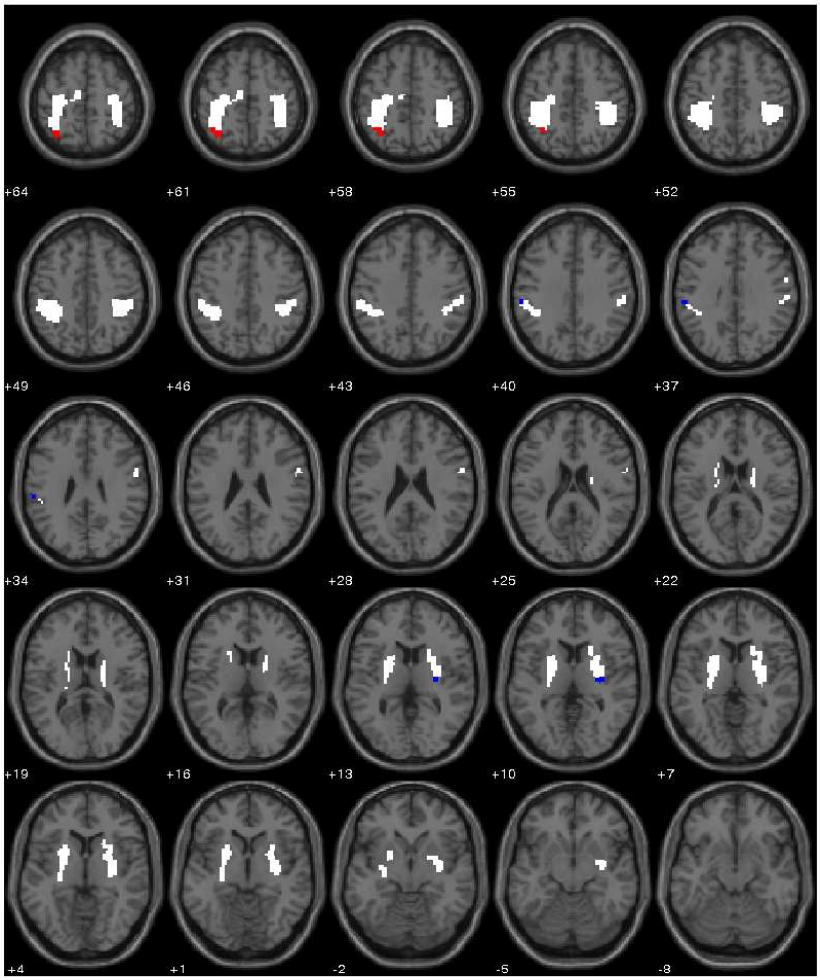


Figure B.5: The cluster pair with the second most significant difference between Baseline and Left. The white regions mark the ROIs and the red and blue mark the clusters.

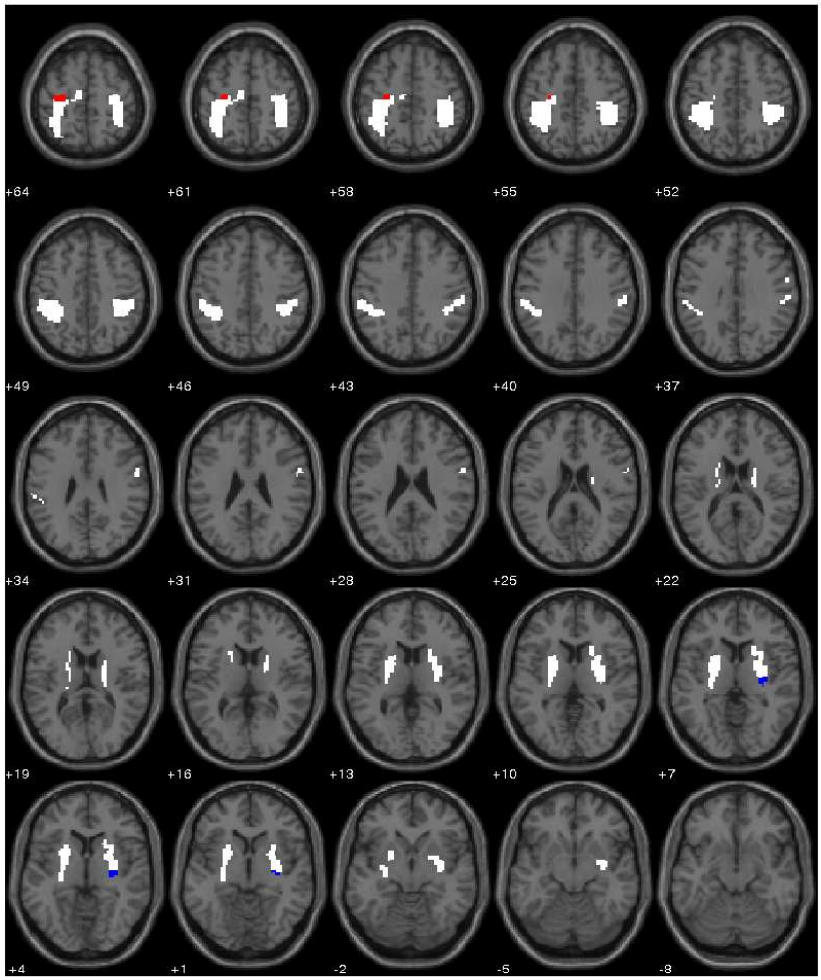


Figure B.6: The cluster pair with the third most significant difference between Baseline and Left. The white regions mark the ROIs and the red and blue mark the clusters.

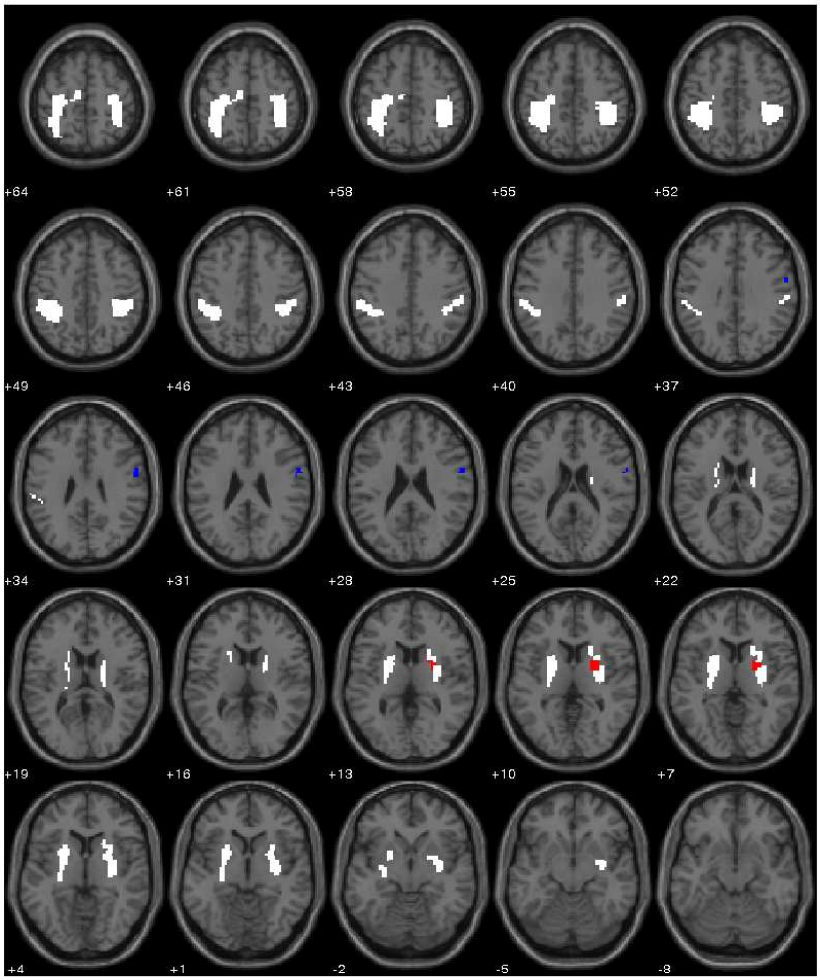


Figure B.7: The cluster pair with the most significant difference between Baseline and Right. The white regions mark the ROIs and the red and blue mark the clusters.

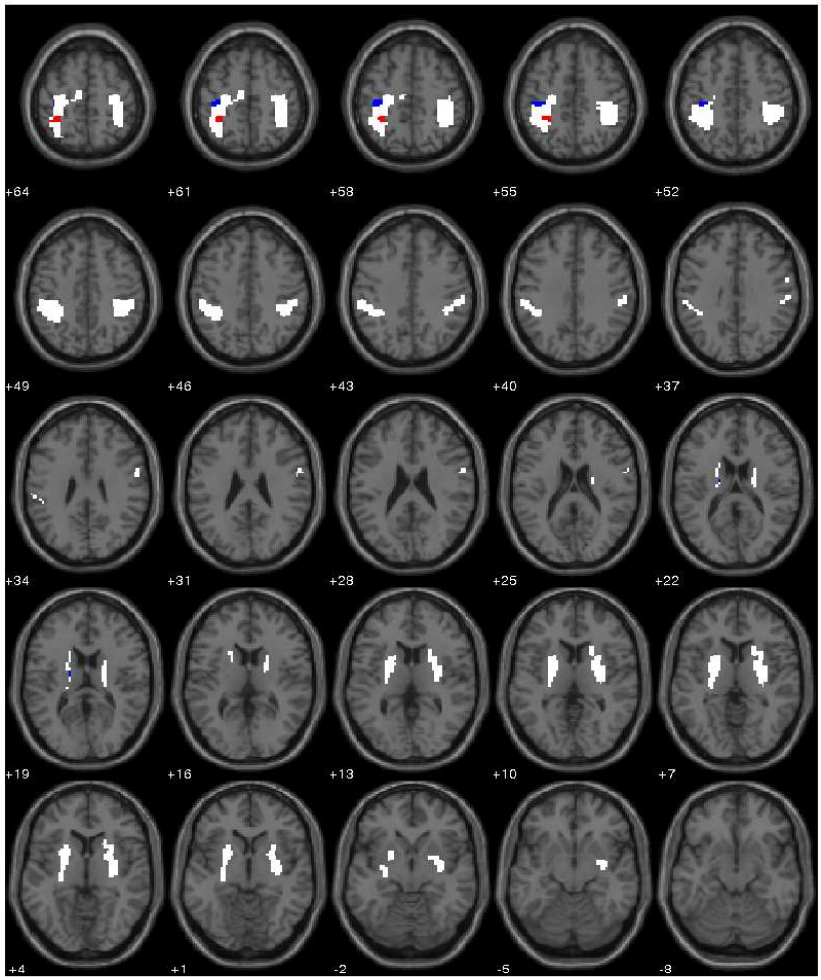


Figure B.8: The cluster pair with the second most significant difference between Baseline and Right. The white regions mark the ROIs and the red and blue mark the clusters.

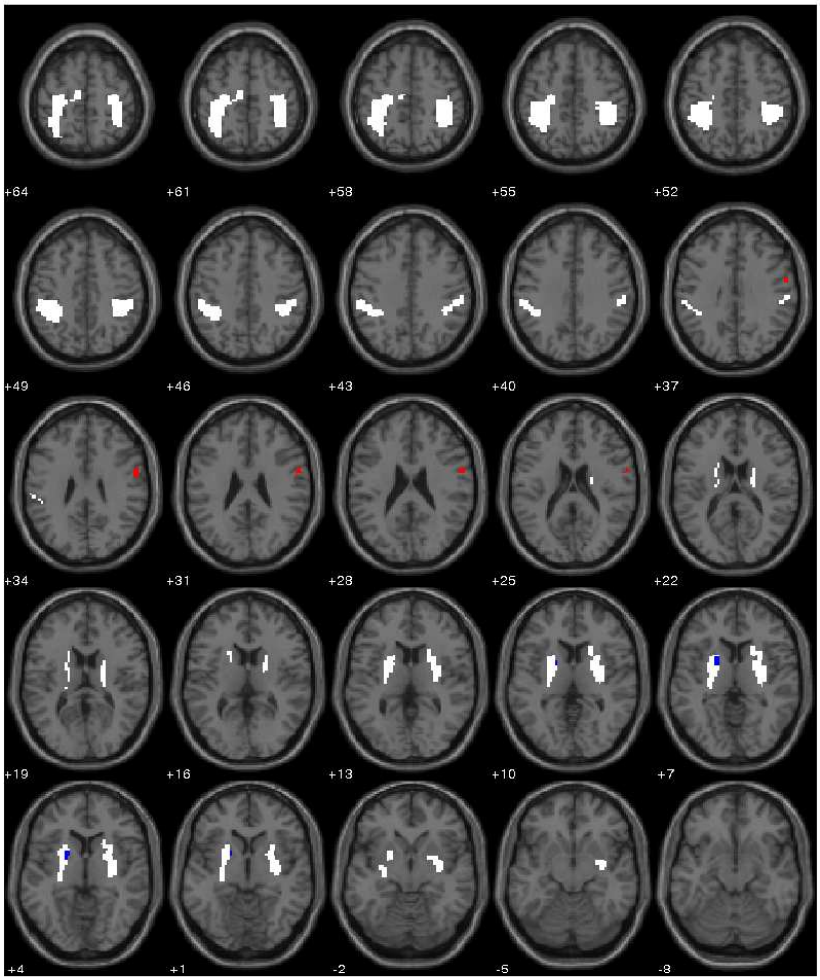


Figure B.9: The cluster pair with the third most significant difference between Baseline and Right. The white regions mark the ROIs and the red and blue mark the clusters.

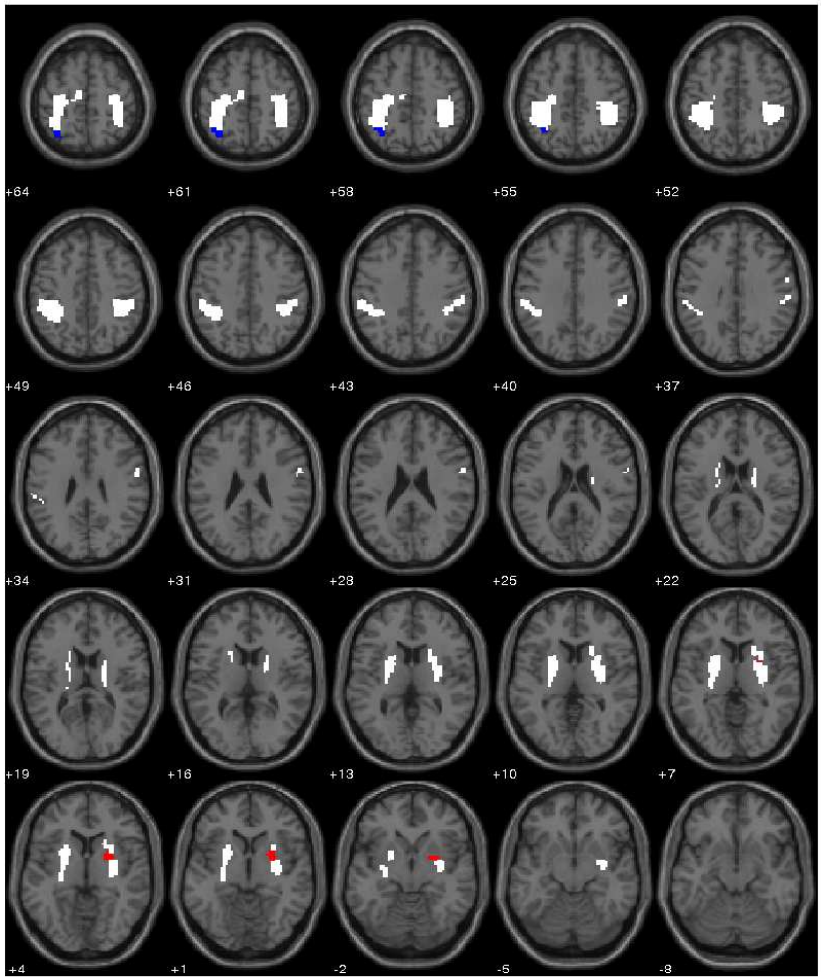


Figure B.10: The cluster pair with the most significant difference between Left and Right. The white regions mark the ROIs and the red and blue mark the clusters.

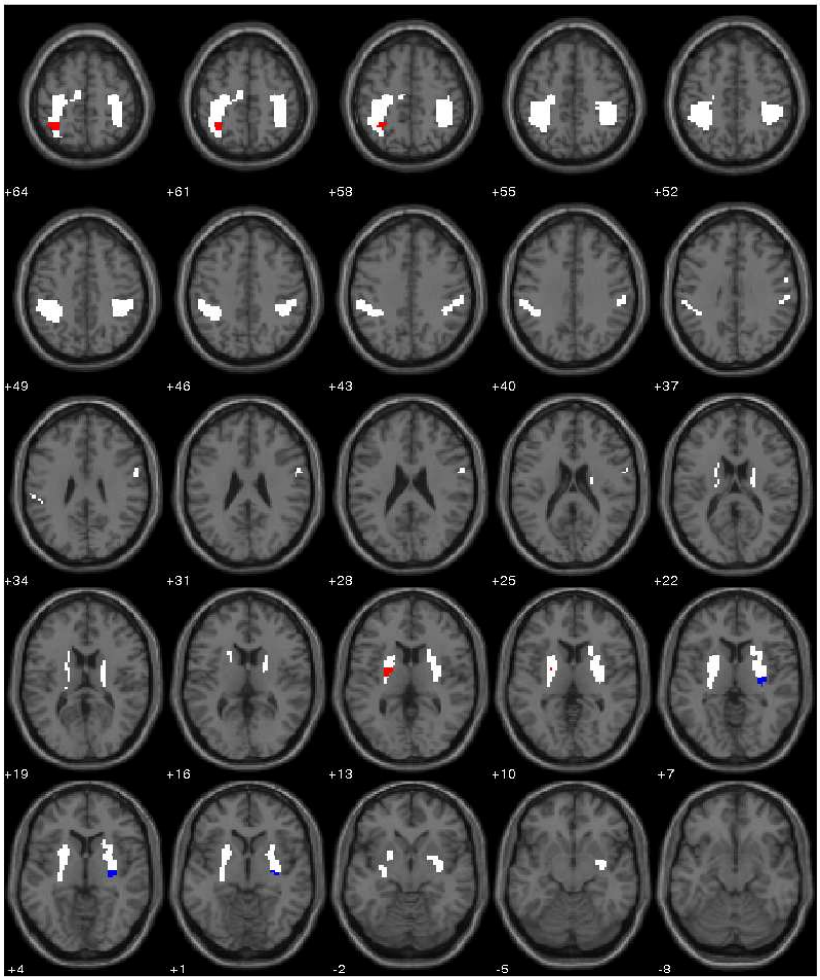


Figure B.11: The cluster pair with the second most significant difference between Left and Right. The white regions mark the ROIs and the red and blue mark the clusters.

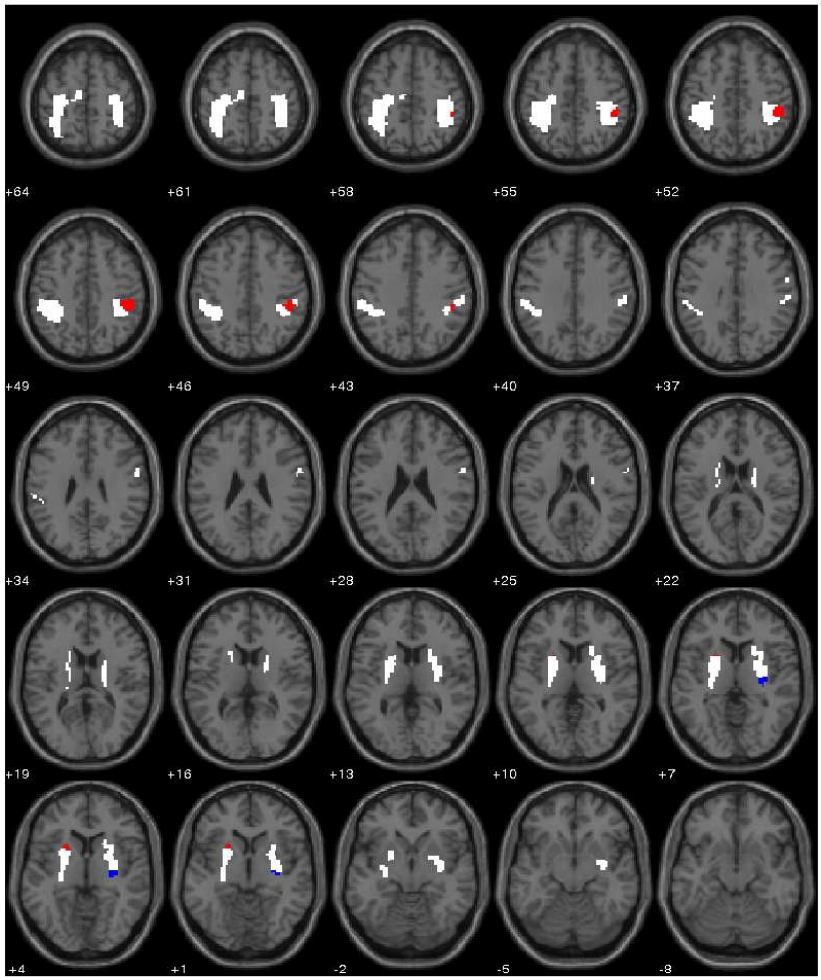


Figure B.12: The cluster pair with the third most significant difference between Left and Right. The white regions mark the ROIs and the red and blue mark the clusters.

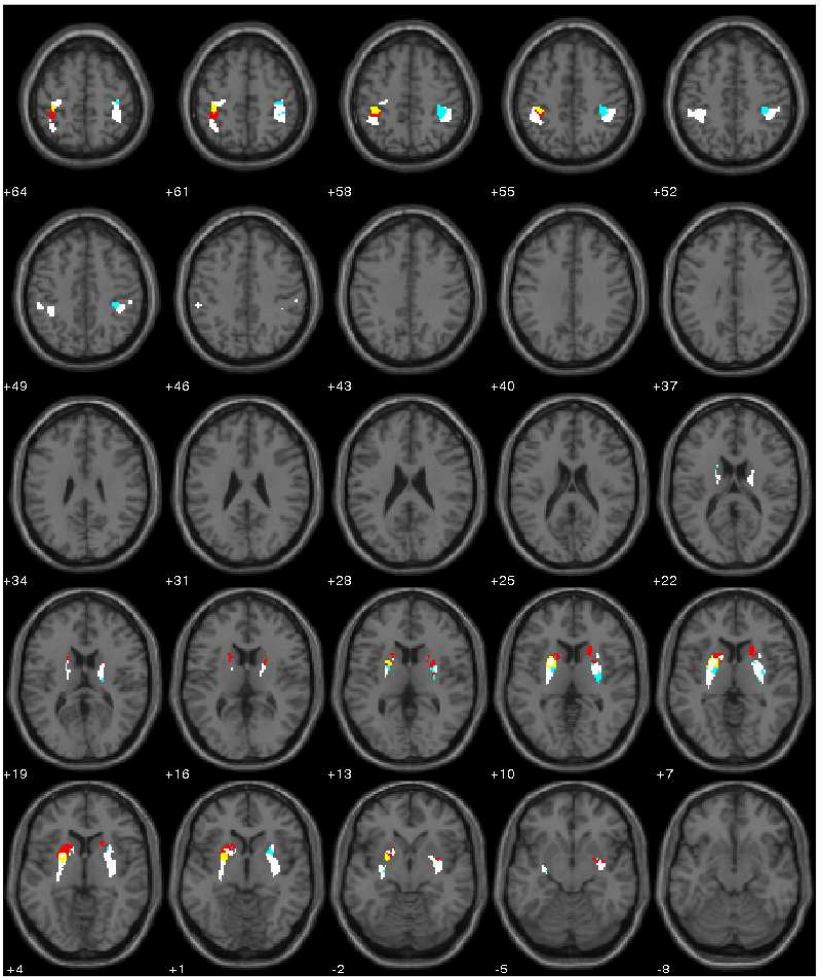


Figure B.13: The three biggest clusters found in the state Baseline. The order of the clusters regarding their size in descending order is red, cyan, and yellow. The white regions mark the ROIs.

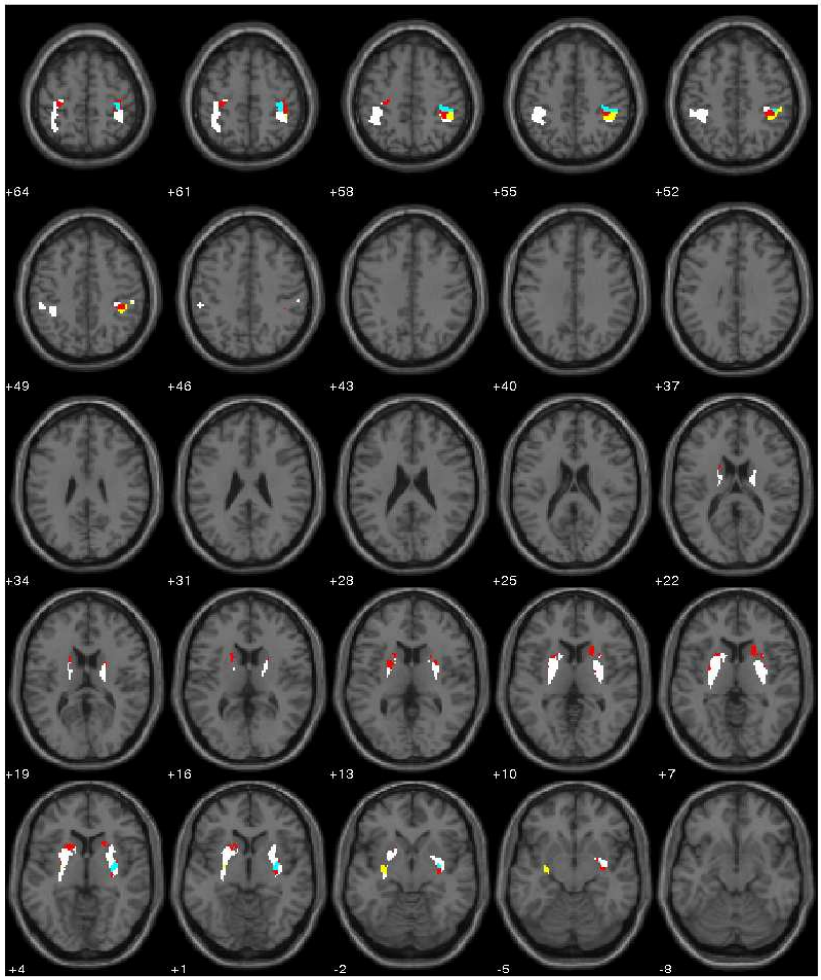


Figure B.14: The three biggest clusters found in the state Left. The order of the clusters regarding their size in descending order is red, cyan, and yellow. The white regions mark the ROIs.

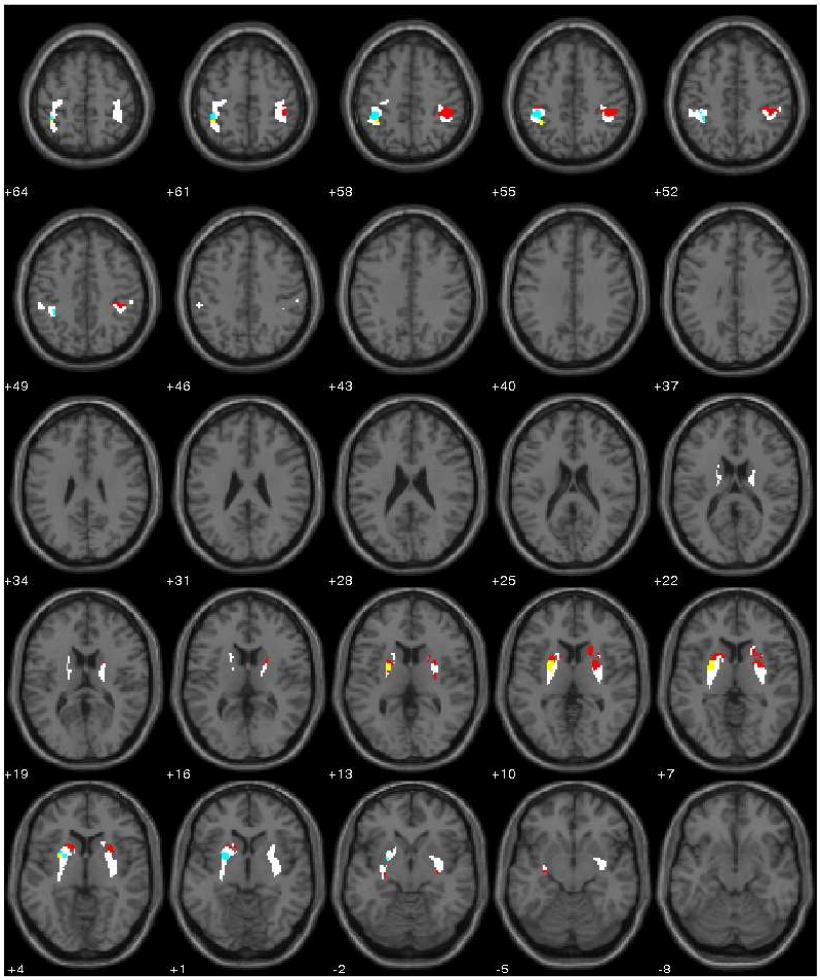


Figure B.15: The three biggest clusters found in the state Right. The order of the clusters regarding their size in descending order is red, cyan, and yellow. The white regions mark the ROIs.

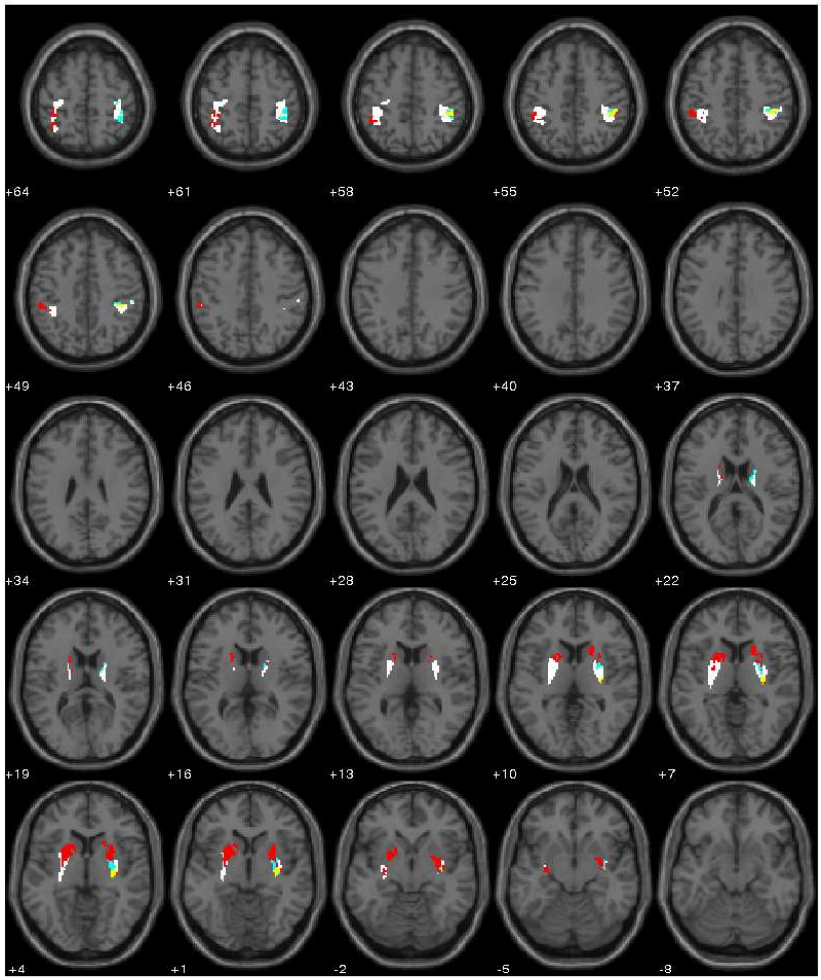


Figure B.16: The three biggest clusters found in the DWI graphs. The order of the clusters regarding their size in descending order is red, cyan, and yellow. The white regions mark the ROIs.

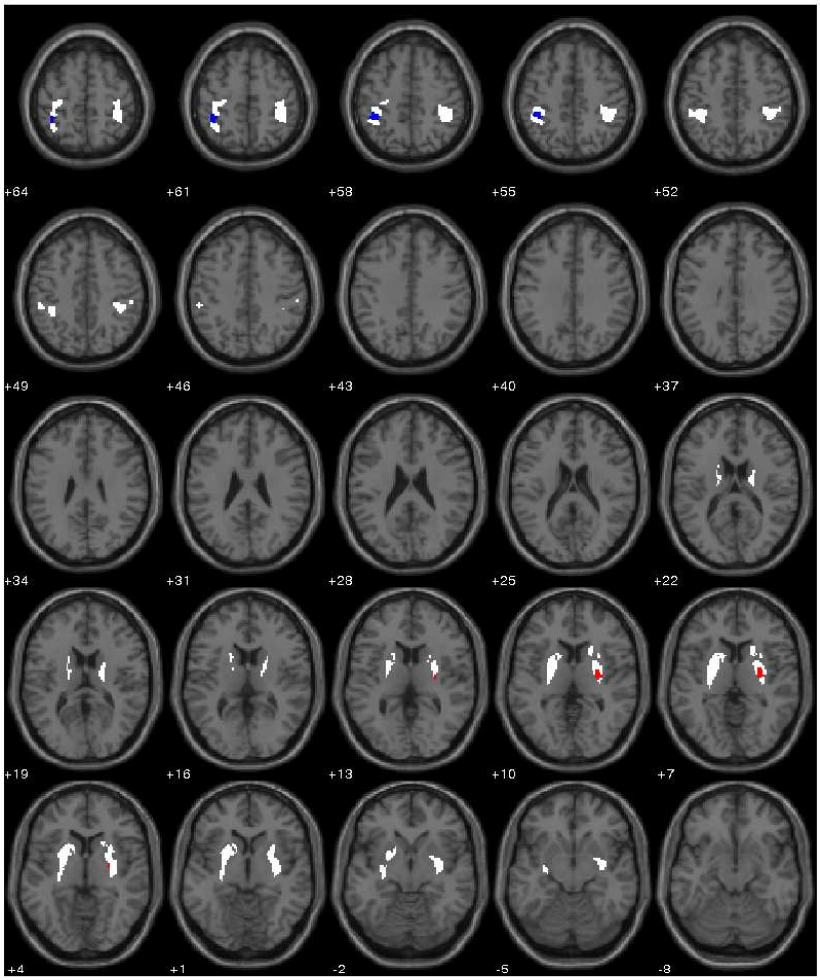


Figure B.17: The cluster pair with the most significant difference between Baseline and Right. The white regions mark the ROIs and the red and blue mark the clusters.

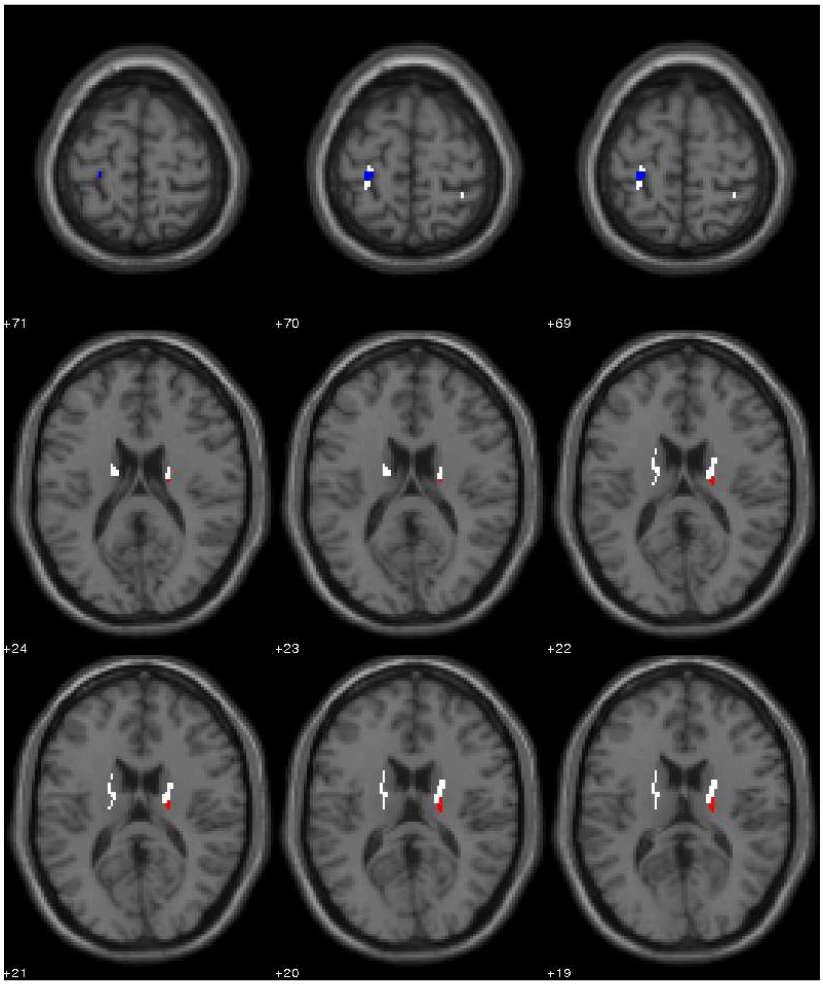


Figure B.18: The cluster pair with the second most significant difference between Baseline and Right. The white regions mark the ROIs and the red and blue mark the clusters.

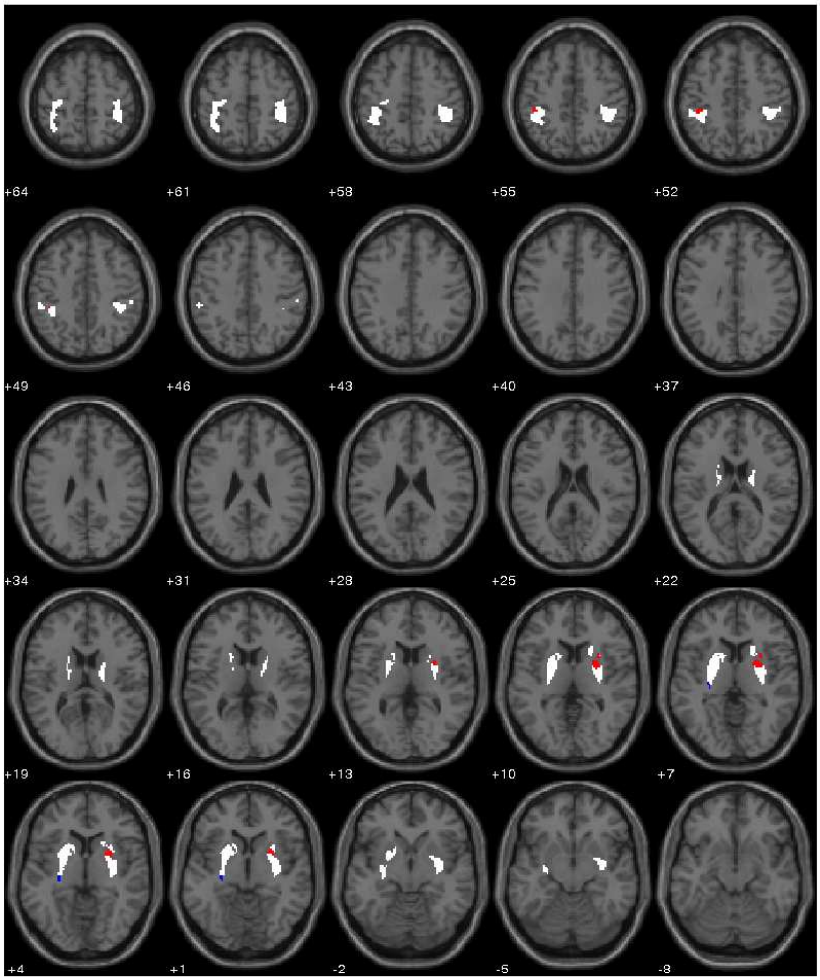


Figure B.19: The cluster pair with the most significant difference between Left and Right. The white regions mark the ROIs and the red and blue mark the clusters.

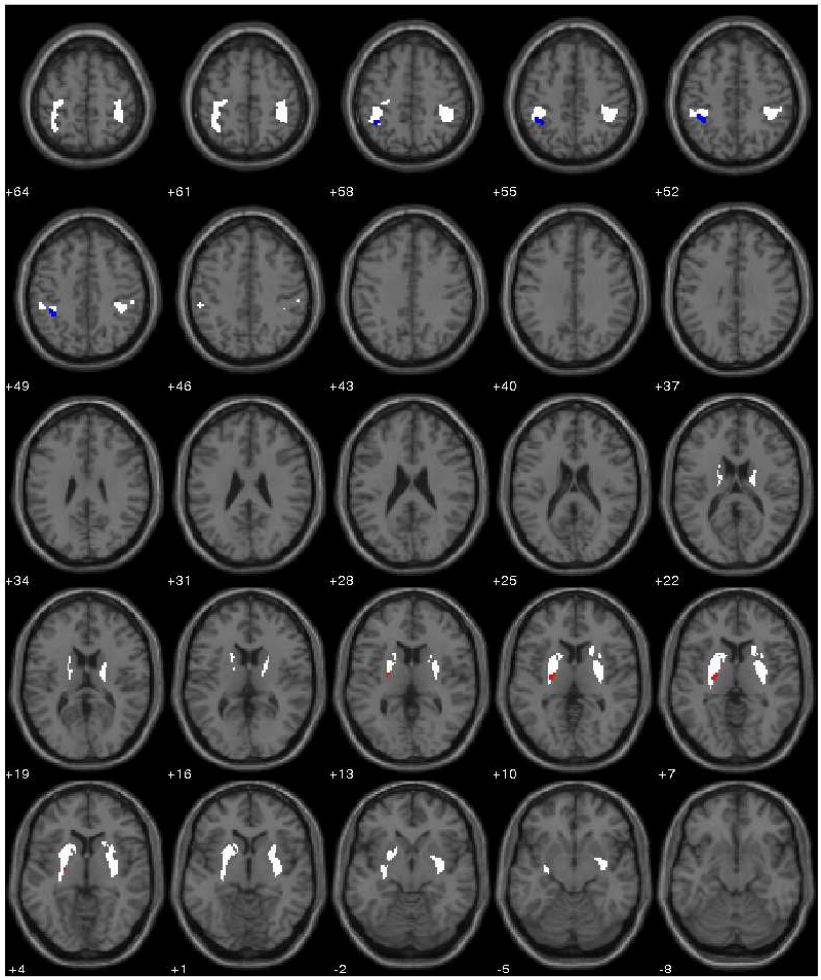


Figure B.20: The cluster pair with the most significant difference between Baseline and DWI. The white regions mark the ROIs and the red and blue mark the clusters.

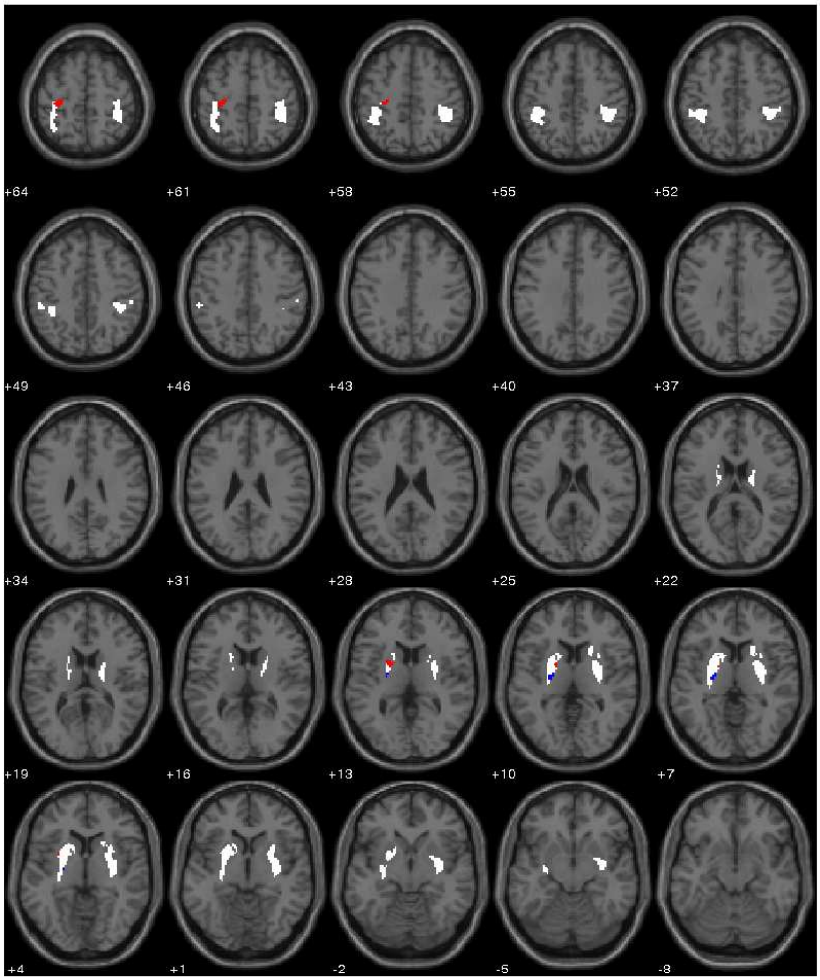


Figure B.21: The cluster pair with the second most significant difference between Baseline and DWI. The white regions mark the ROIs and the red and blue mark the clusters.

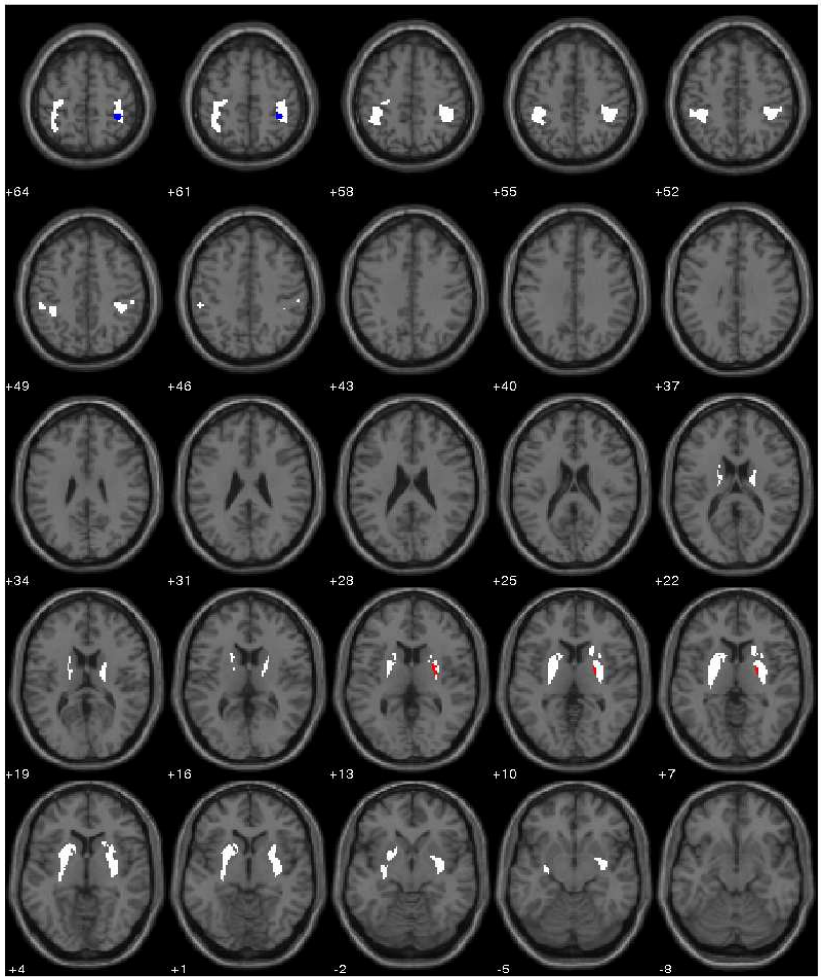


Figure B.22: The cluster pair with the third most significant difference between Baseline and DWI. The white regions mark the ROIs and the red and blue mark the clusters.

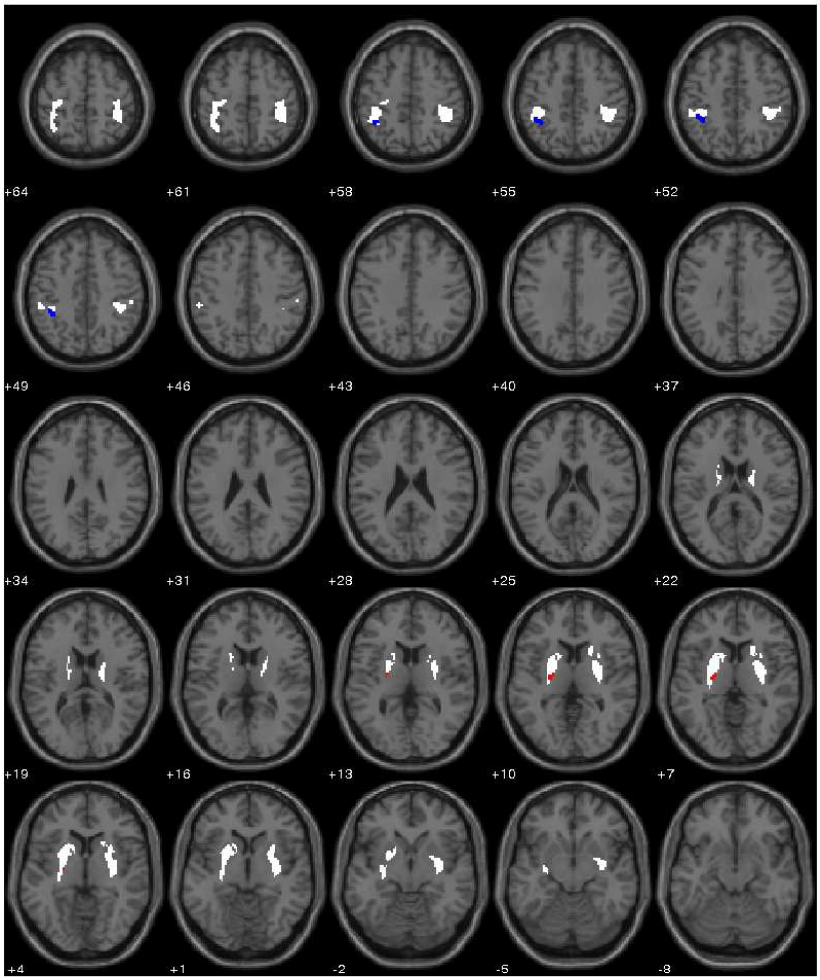


Figure B.23: The cluster pair with the most significant difference between Left and DWI and between Right and DWI. The white regions mark the ROIs and the red and blue mark the clusters.

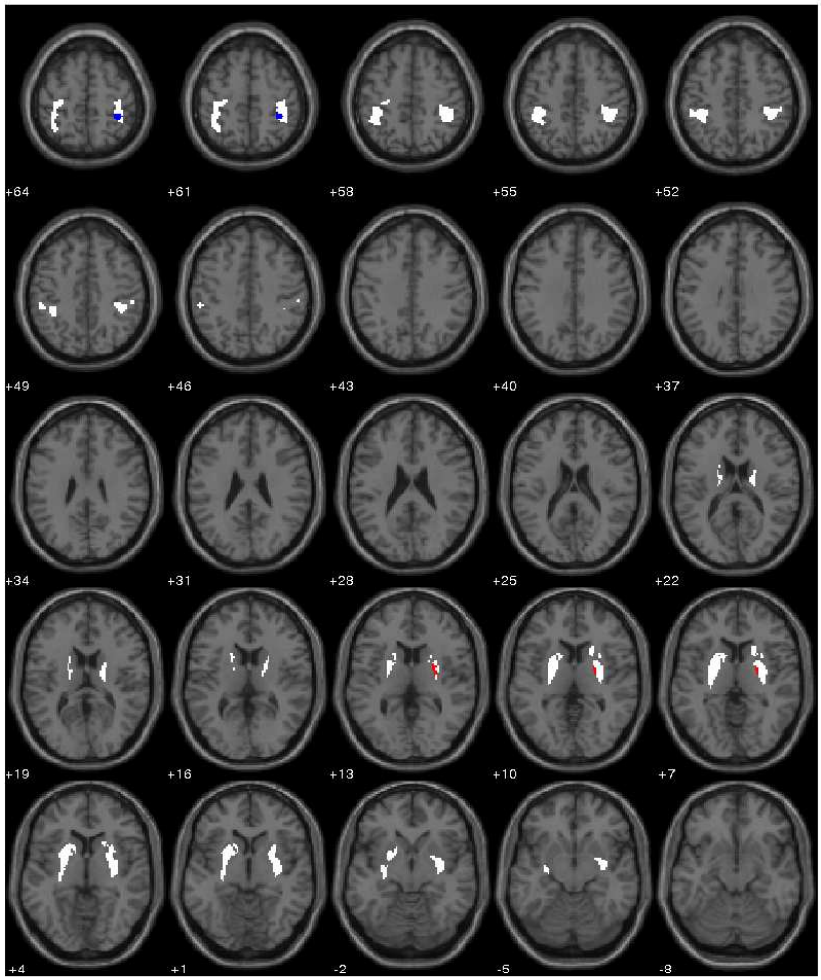


Figure B.24: The cluster pair with the second most significant difference between Left and DWI and between Right and DWI. The white regions mark the ROIs and the red and blue mark the clusters.

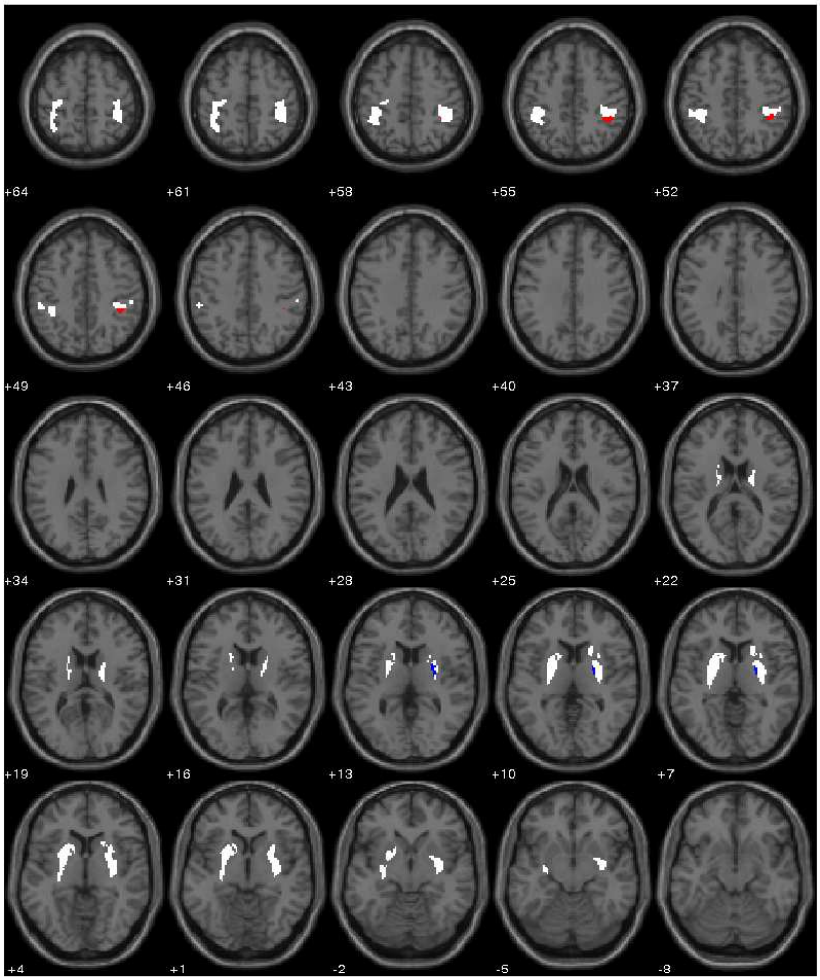


Figure B.25: The cluster pair with the third most significant difference between Left and DWI and between Right and DWI. The white regions mark the ROIs and the red and blue mark the clusters.

Appendix C

Matlab Functions

An outline of the Matlab scripts used in the project:

1. run_processing.m
    - Pre-processing of fMRI data
    - Author: Karen Sandø Ambrosen
2. run_model.m
    - Model specification and estimation for each subject
    - Author: Karen Sandø Ambrosen
3. run_grpmodel.m
    - Model specification and estimation for the group of subjects
    - Author: Karen Sandø Ambrosen
4. timeseries_extraction.m
    - Generates fMRI graphs
    - Calls roi_stat_extraction.m
    - Author: Karen Sandø Ambrosen

5. roi_stat_extraction.m
    - Extracts the timeseries from the fMRI data
    - Author: Kasper Winther Andersen
6. run_bet.m
    - BET brain extraction
    - Author: Karen Sandø Ambrosen
7. run_registration.m
    - Estimates the transformations between spaces: diffusion, structural, and standard
    - Author: Karen Sandø Ambrosen
8. run_probtrackx2.m
    - Performs probabilistic tratography
    - Author: Karen Sandø Ambrosen
9. threshold_DWI.m
    - Generates the DWI graphs
    - Author: Karen Sandø Ambrosen
10. func_analData.m
    - Infers all graphs simultaneously
    - Calls IRMUnipartite.m
    - Author: Karen Sandø Ambrosen
11. func_analData_splits.m
    - Splits the graphs in two sets and inferring the sets separately. Calculates the MI and the NMI between splits.
    - Calls IRMUnipartite.m, calcMI.m, and calcNMI.m
    - Author: Karen Sandø Ambrosen
12. func_analData_states.m
    - Infers the states separately
    - Calls IRMUnipartite.m
    - Author: Karen Sandø Ambrosen
13. func_analData_states_splits.m

- Splits the graphs from each state in two sets and inferring the sets separately. Calculates the MI, the NMI and the test log likelihood between splits.
- Calls IRMUnipartite.m, calcMI.m, and calcNMI.m
- Author: Karen Sandø Ambrosen

14. generateGraphCRPUnipartite.m
    - Generates unipartite graphs according to the Infinite Relational Model
    - Author: Morten Mørup
15. IRMUnipartite.m
    - Non-parametric clustering of undirected graphs based on collapsed Gibbs sampling with split-merge (IRM)
    - Author: Morten Mørup
16. calcMI.m
    - Calculates mutual information
    - Author: Morten Mørup
17. calcMI.m
    - Calculates normalized mutual information
    - Calls calcMI.m
    - Author: Morten Mørup
18. BayesianCommunityDetection.m
    - Non-parametric clustering of undirected graphs based on collapsed Gibbs sampling with split-merge (BCD)
    - Author: Morten Mørup
19. generateBCDGraph.m
    - Generates binary unipartite graphs according to the Bayesian Community Detection (BCD) generative model
    - Author: Morten Mørup
20. createValidationData.m
    - Defines entries in graphs to be treated as missing in the inference
    - Author: Morten Mørup
21. permutationTest.m

- Performs the permutation tests
- Calls calcNMI.m
- Author: Karen Sandø Ambrosen

22. clustertest_pairedttest.m

    - Finds the cluster pairs with most significant differences between the states, using a paired t-test.
    - Author: Karen Sandø Ambrosen

23. classify.m

    - Generates graphs using the estimated assignment matrix and the estimated link probability matrix, and calculates the test log likelihood for the different states.
    - Author: Karen Sandø Ambrosen

24. plot_segmentations.m

    - Superimposes the clusters on an anatomical brain template
    - Author: Kasper Winther Andersen

Bibliography

[1] ehumanbiofield. *https://ehumanbiofield.wikispaces.com/Neuron+Structure+NLL*. Last checked 26-01-2013.

[2] Fdt/userguide. *http://fsl.fmrib.ox.ac.uk/fsl/fslwiki/FDT/UserGuide*. Last checked 08-01-2013.

[3] Graph theory picture gallery: A planar graph. *http://www.online-utility.org/image/gallery.jsp?title=Graph+theory*. Last checked 29-01-2013.

[4] physicsarchives.com. *http://physicsarchives.com/index.php/courses/999*. Last checked 26-01-2013.

[5] K.W. Andersen, M. Mørup, H. Siebner, K.H. Madsen, and L.K. Hansen. Identifying Modular Relations In Complex Brain Networks. In *Machine Learning for Signal Processing (MLSP), 2012 {IEEE} International Workshop on*, pages 1–6, 2012.

[6] C.J. Anderson, S. Wasserman, and K. Faust. Building stochastic blockmodels. *Social Networks*, 14(1992):137–161, 1992.

[7] R. Bammer. Basic principles of diffusion-weighted imaging. *European Journal of Radiology*, 45(3):169–184, 2003.

[8] A. Barabasi and R. Albert. Emergence of scaling in random networks. *science*, 286(5439):509–512, 1999.

[9] P.J. Basser, J. Mattiello, and D. LeBihan. MR Diffusion Tensor Spectroscopy and Imaging. *Biophysical journal*, 66(1):259–267, 1994.

[10] F.D. Bowman, L. Zhang, G. Derado, and S. Chen. Determining functional connectivity using fMRI data with diffusion-based anatomical weighting. *NeuroImage*, 62(3):1769–1779, 2012.

[11] D. Chen and X. Xu. A Novel Approach for Finding Clusters from Complex Networks. *Journal of Software*, 6(1):132–139, 2011.

[12] E. F. Codd. Derivability, redundancy and consistency of relations stored in large data banks. *IBM Research Report, San Jose, California*, RJ599, 1969.

[13] E.F. Codd. A relational model of data for large shared data banks. *Communications of the ACM*, 13(6):377–387, 1970.

[14] T.M. Cover and J.A. Thomas. Chapter 2 Entropy , Relative Entropy and Mutual Information. In *Elements of Information Theory*, chapter 2, pages 12–49. John Wiley & Sons, Inc., New York, 1991.

[15] A.C. Evans, D.L. Collins, SR Mills, ED Brown, RL Kelly, and T.M. Peters. 3d statistical neuroanatomical models from 305 mri volumes. In *Nuclear Science Symposium and Medical Imaging Conference, 1993., 1993 IEEE Conference Record.*, pages 1813–1817, 1993.

[16] Wellcome Trust Centre for Neuroimaging. Spm. *http://www.fil.ion.ucl.ac.uk/spm*. Last checked 01-02-2013.

[17] S. Fortunato. Community detection in graphs. *Physics Reports*, 486(3-5):75–174, 2010.

[18] P.T. Fox and M.E. Raichle. Focal physiological uncoupling of cerebral blood flow and oxidative metabolism during somatosensory stimulation in human subjects. *Proceedings of the National Academy of Sciences of the United States of America*, 83(4):1140–1144, 1986.

[19] A. Goldenberg, A.X. Zheng, S.E. Fienberg, and E.M. Airoldi. A survey of statistical network models. *Foundations and Trends in Machine Learning*, 2(2):129–233, 2010.

[20] R.I. Goldman, J.M. Stern, J. Engel, and M.S. Cohen. Acquiring simultaneous EEG and functional MRI. *Clinical neurophysiology*, 111(11):1974–1980, 2000.

[21] N. Gulbahce and S. Lehmann. The art of community detection. *BioEssays*, 30(10):934–938, 2008.

[22] S.N. Haber. The primate basal ganglia: parallel and integrative networks. *Journal of Chemical Neuroanatomy*, 26(4):317–330, 2003.

[23] Y.O. Halchenko, S.J. Hanson, and B.A. Pearlmutter. Chapter 8 Multimodal Integration: fMRI , MRI , EEG , MEG. In *Advandced Image Processing in Magnetic Resonance Imaging*, volume 265, pages 223–265. 2005.

[24] L.G. Hanson. Introduction to Magnetic Resonance Imaging Techniques. *http://www.drcmr.dk/phd2012/MRI_physics_Hanson_notes.pdf.* Last checked 26-01-2013.

[25] T. Hesterberg, D.S. Moore, S. Monaghan, A. Clipson, and R. Epstein. Bootstrap methods and permutation tests. *Introduction to the Practice of Statistics*, 5:1–70, 2005.

[26] J.M. Hofman and C.H. Wiggins. A Bayesian Approach to Network Modularity. *Physical review letters*, 100(25):258701, 2008.

[27] P.W. Holland, K.B. Laskey, and S. Leinhardt. Stochastic blockmodels: First steps. *Social Networks*, 5(2):109–137, 1983.

[28] J.R. Ives, S. Warach, F. Schmitt, R.R. Edelman, and D.L. Schomer. Monitoring the patient's EEG during echo planar MRI. *Electroencephalography and clinical neurophysiology*, 87(6):417–420, 1993.

[29] S. Jain and R.M. Neal. A Split-Merge Markov Chain Monte Carlo Procedure for the Dirichlet Process Mixture Model. *Journal of Computational and Graphical Statistics*, 13(1):158–182, 2004.

[30] M. Jenkinson, C.F. Beckmann, T.E. Behrens, M.W. Woolrich, and S.M. Smith. Fsl. *Neuroimage*, 62:782–790, 2012.

[31] C. Kemp, J.B. Tenenbaum, T.L. Griffiths, T. Yamada, and N. Ueda. Learning systems of concepts with an infinite relational model. *Proceedings of the national conference on artificial intelligence*, 21(1):381–388, 2006.

[32] S. Lehéricy, E. Bardinet, L. Tremblay, P.F. Van de Moortele, J.B. Pochon, D. Dormont, Dae-Shik Kim, Jerome Yelnik, and Kamil Ugurbil. Motor control in basal ganglia circuits using fMRI and brain atlas approaches. *Cerebral cortex (New York, N.Y. : 1991)*, 16(2):149–61, 2006.

[33] P. Lieberman. Foxp2 and human cognition. *Cell*, 137(5):800–802, 2009.

[34] P.M. Matthews and P. Jezzard. Functional magnetic resonance imaging. *Journal of neurology, neurosurgery, and psychiatry*, 75(1):6–12, 2004.

[35] S. Mori and J. Zhang. Principles of diffusion tensor imaging and its applications to basic neuroscience research. *Neuron*, 51(5):527–539, 2006.

[36] M. Mørup, K.H. Madsen, A.M. Dogonowski, H. Siebner, and L.K. Hansen. Infinite relational modeling of functional connectivity in resting state fMRI. *Neural Information Processing Systems 23*, 2010.

[37] M. Mørup and M.N. Schmidt. Bayesian community detection. *Neural computation*, 24(9):2434–2456, 2012.

[38] K. Nowicki and T.A.B. Snijders. Estimation and prediction for stochastic blockstructures. *Journal of the American Statistical Association*, 96(455):1077–1087, 2001.

[39] R.C. Oldfield. The assessment and analysis of handedness: the edinburgh inventory. *Neuropsychologia*, 9(1):97–113, 1971.

[40] G.J.M. Parker, H.A. Haroon, and C.A.M. Wheeler-Kingshott. A framework for a streamline-based probabilistic index of connectivity (PICo) using a structural interpretation of MRI diffusion measurements. *Journal of magnetic resonance imaging*, 18(2):242–254, 2003.

[41] C. Pierpaoli and P.J. Basser. Toward a quantitative assessment of diffusion anisotropy. *Magnetic Resonance in Medicine*, 36(6):893–906, 2005.

[42] E. Rykhlevskaia, G. Gratton, and M. Fabiani. Combining structural and functional neuroimaging data for studying brain connectivity: a review. *Psychophysiology*, 45(2):173–187, 2007.

[43] S. M. Smith. Fast robust automated brain extraction. *Human Brain Mapping*, 17(3):143–155, 2002.

[44] S.M. Smith, M. Jenkinson, M.W. Woolrich, C.F. Beckmann, T.E.J. Behrens, H. Johansen-Berg, P.R. Bannister, M. De Luca, I. Drobnjak, D.E. Flitney, R. Niazy, J. Saunders, J. Vickers, Y. Zhang, N. De Stefano, J.M. Brady, and P.M. Matthews. Advances in functional and structural mr image analysis and implementation as fsl. *Neuroimage*, 23(S1):208–219, 2004.

[45] 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.

[46] S.C. Strother, J. Anderson, L.K. Hansen, U. Kjems, R. Kustra, J. Sidtis, S. Frutiger, S. Muley, S. LaConte, and D. Rottenberg. The quantitative evaluation of functional neuroimaging experiments: the NPAIRS data analysis framework. *NeuroImage*, 15(4):747–771, 2002.

[47] N. Tzourio-Mazoyer, B. Landeau, and D. Papathanassiou. Automated anatomical labeling of activations in spm using a macroscopic anatomical parcellation of the mni mri single-subject brain. *Neuroimage*, 15:273–289, 2002.

[48] The Analysis Group FMRIB Oxford UK. Fmrib software library. *http://fsl.fmrib.ox.ac.uk/fsl/fslwiki/*. Last checked 03-02-2013.

[49] D.J. Werring, C.A. Clark, G.J. Barker, D.H. Miller, G.J. Parker, M.J. Brammer, E.T. Bullmore, V.P. Giampietro, and A.J. Thompson. The structural and functional mechanisms of motor recovery: complementary use of diffusion tensor and functional magnetic resonance imaging in a traumatic injury of the internal capsule. *Journal of neurology, neurosurgery, and psychiatry*, 65(6):863–869, 1998.

[50] D.J. Werring, C.A. Clark, G.J.M. Parker, D.H. Miller, A.J. Thompson, and G.J. Barker. A direct demonstration of both structure and function in the visual system: combining diffusion tensor imaging with functional magnetic resonance imaging. *Neuroimage*, 9(3):352–361, 1999.

[51] M.W. Woolrich, S. Jbabdi, B. Patenaude, M. Chappell, S. Makni, T. Behrens, C. Beckmann, M. Jenkinson, and S.M. Smith. Bayesian analysis of neuroimaging data in fsl. *Neuroimage*, 45:173–186, 2009.

[52] Z. Xu, V. Tresp, K. Yu, and H.P. Kriegel. Infinite hidden relational models. *In Proceedings of the 22nd Inter- national Conference on Uncertainity in Artificial Intelligence*, 2006.

[53] W.W. Zachary. An information flow model for conflict and fission in small groups. *Journal of anthropological research*, pages 452–473, 1977.