

RADBOD UNIVERSITY NIJMEGEN

BACHELOR THESIS IN ARTIFICIAL INTELLIGENCE

Connectivity-Based Parcellation of the Brain using the Infinite Relational Model and Bayesian Community Detection

Author:

Lucy van Oostveen

Supervisors:

dr. Marcel A. J. van Gerven

Max Hinne, MSc

August 6, 2013



Abstract

A main interest in cognitive neuroscience is to understand how the brain is segregated into regions that subserve particular functions. There are different ways of approaching this, one of them is to cluster together brain regions that show similar structural connectivity patterns. The resulting clusters are taken to represent functionally distinct areas of the brain based on the assumption that structure and function are intimately related. In this thesis, structural connectivity data of the brain was used to compare two Bayesian clustering algorithms: the Infinite Relational model and Bayesian Community Detection. The maximum a posteriori estimates of twenty subjects, required by simulated annealing, were used to compare both models. It was found that Bayesian Community Detection produces significantly more clusters than the Infinite Relational Model while at the same time performing just as well as the Infinite Relational Model in terms of reproducibility.

I want to thank my friends and supervisors for the patience, help and motivation during the writing of this thesis. Special thanks go to Nikki for helping me with the finishing touch.

Contents

Introduction	5
Methods	6
The Infinite Relational Model	8
Bayesian Community Detection	10
Inference	11
MAP estimate	12
Measures	12
Data	14
Results	15
Number of Clusters	15
Adjusted Mutual Information	15
Number of nodes per cluster	16
Clusters	16
Discussion	20
MAP estimate	20
Adaption of BCD	20
Improvements and future research	21
References	22
Appendix	23
Number of Clusters	23
IRM clustering of Subject 7	23
BCD clustering of Subject 7	24
IRM clustering of Subject 14	25
BCD clustering of Subject 14	26

Introduction

For over a 100 years, researchers within neuroscience have been investigating how the brain is segregated into regions that have the same function. There are different methods to segregate the brain into regions. To parcellate individual cortices for functional studies, Sural/Gyral landmarks have long been used (Fischl, Van Der Kouwe, et al., 2004). To localise brain functions Atlas-based approaches can also be used, although they are limited by the high level of inter-subject variability (Fischl, Rajendran, et al., 2008). One of the approaches is to use the structure of the brain. The idea behind this approach is that the function of a brain region is constrained by the connections it has (Passingham, Stephan, & Kötter, 2002) and the assumption is made that structure and function are intimately related (Honey, Thivierge, & Sporns, 2010). Structural connectivity refers to the actual anatomical connections between brain regions. Functional connectivity refers to covarying patterns of activity between brain regions. In this thesis we focus on the use of structural connectivity data to infer a parcellation of the brain into different regions. Both type of data are used to parcellate the brain (Jbabdi, Woolrich, & Behrens, 2009; Mars et al., 2012; Beckmann, Johansen-Berg, & Rushworth, 2009). Because the process of parcellation has a lot of similarities with clustering, clustering methods can be used to parcellate the brain. Clustering is a process of grouping similar objects together. There exist various algorithms that can cluster data. Mørup et al. used the Infinite Relational Model (IRM) (Kemp, Tenenbaum, Griffiths, Yamada, & Ueda, 2006) for this in the context of functional connectivity analysis (Mørup, Madsen, Dogonowski, Siebner, & Hansen, 2010). The IRM is a model that uses relational data for clustering. Semantic knowledge exists of relations. This means that the IRM can be used to cluster semantic data. Recently another cluster method was developed, called Bayesian Community Detection (BCD). This method uses a slightly different definition of a cluster as compared to the IRM. In this thesis, both models will be used to estimate whole-brain parcellations using structural connectivity data. With IRM, clusters have a lot of connections within a cluster and few connections between clusters. With BCD, clusters have comparatively few connections between clusters than connections within cluster link. Due to the more reasonable formulation of a cluster, we expect the BCD model to lead to more reasonable parcellations. Therefore, in this thesis the following research question will be answered:

Does Bayesian Community Detection lead to more reasonable estimates of whole-brain parcellations compared to the Infinite Relational model?

Methods

In this thesis two different clustering algorithms were compared. There are a lot different approaches to model this process. The approach that fits best depends on the type of input data. In this thesis relational data was used as input. Relational data gives information about how the nodes are connected. A typical way to represent this type of data is using an adjacency matrix, which is a square matrix where every row and column represent a node from the data. Herein indicates a 1 on position (i, j) a connection between node i and j , a 0 indicates no connection between the two nodes (Figure 1).

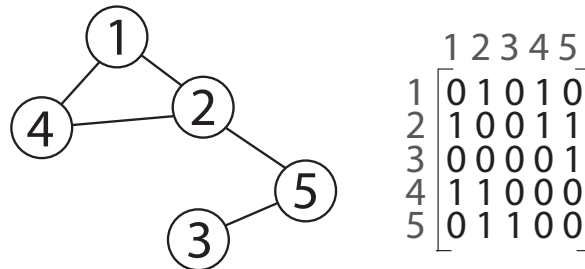


Figure 1: On the left is a simple network, on the right the corresponding adjacency matrix. The matrix is symmetric because connections are assumed to be undirected. A connection between two nodes i and j is indicated by a 1 on (i, j) and (j, i) since the matrix is symmetric.

Graphical Models

The IRM and BCD are both Bayesian clustering methods. Bayesian clustering methods can be represented as graphical models which capture the independence structure between random variables. The possible value of a random variable is influenced by some random process. The number of children of a random person in the street, the number thrown with a dice or the height of a random building in your city are examples of random variables. In order to illustrate the concept of a graphical model, we start with a Bayesian network defined for discrete variables.

The wet grass example

A classic example is the Bayesian network of wet grass. In Figure 2 the network that belongs to this problem is shown. This Bayesian model describes whether the grass in the backyard is wet or dry. The variable *wetgrass* depends on two other random variables: *sprinkler* and *rain*, because both have an effect on the wetness of the grass. Rain depends on the clouds, therefore *rain* depends on *cloudy*. The sprinkler depends on the clouds because it has a sensor that reacts to the amount of sunlight, therefore *sprinkler* depends on *cloudy*. A Bayesian model also specifies how variables influence each other. Next to the node of wet grass, is the probability table of the node. As can be seen, the probability that the grass is wet when it is not raining and the sprinkler is off is much smaller than when it is raining and the sprinkler is on.

Continuous random variables

In the above example all the random variables are discrete, but random variables can also be continuous. For example the temperature outside which can have an infinite number of possible values. To describe these type of variables continuous probability distributions are used. A continuous probability distribution defines the probability that a variable has a specific value. A common example of a continuous probability distribution is the normal (Gaussian) distribution. The Gaussian distribution is a distribution that has two parameters: the mean (m) and the standard deviation (σ). In Figure 3 different normal distributions are shown. On the horizontal axis you can read the value of the variable and on the vertical axis the corresponding probability that the variable has this specific value. The shape of the function depends on the parameters of the distribution. In general, graphical models can be defined over discrete and continuous random variables. When continuous random variables are used, the probability tables are replaced by continuous probability distributions.

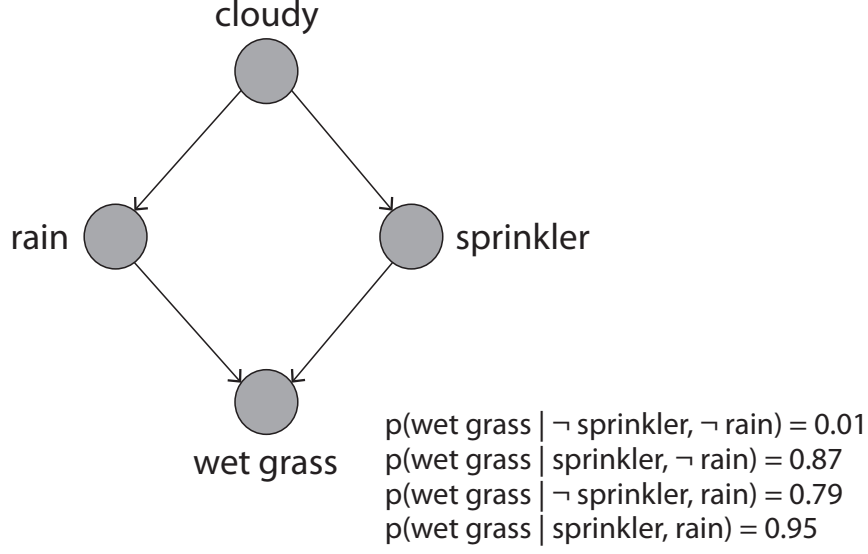


Figure 2: A classical Bayesian network example (Adapted from (Russell et al., 1995)).

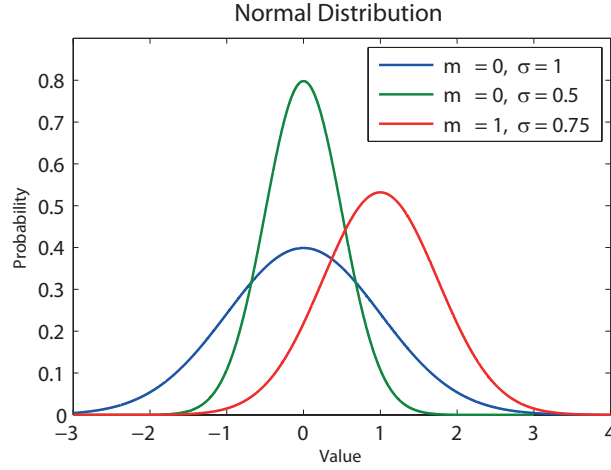


Figure 3: Three different normal distributions. The shape is determined by the mean (m) and the standard deviation (σ).

The hair length example

Temperature is a continuous random variable. One of the things that is influenced by the temperature is the average hair length of humans. When the temperature rises, more people will cut their hair because it is warmer, so the average hair length will be shorter. In Figure 4 the model that belongs to this problem is shown. The probability distribution that is associated with the temperature (T) has two parameters, m and σ . These parameters are called hyper-parameters of the model. The distribution of T is called the prior distribution. T influences the hair length (H) by influencing the parameter of the probability of the distribution that is associated with H . Naturally, there are other factors influencing the hair length. In this simple example, these factors are captured in the two extra parameters of H , c and s .

Bayesian clustering methods

With Bayesian clustering methods as the IRM and BCD, a graphical model is formed to describe the relation between the data and the clustering. In such network the clustering and input data are random

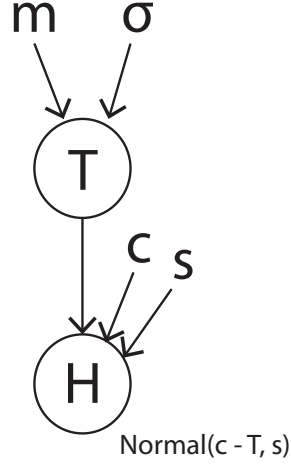


Figure 4: The hair length example. The temperature (T) influences the hair length (H). Therefore, the value of H depends on the value of T . c and s are other factors influencing the hair length.

variables in the network. The idea behind this is that the connections between nodes in the data depend on the clustering of the data.

The Infinite Relational Model

The IRM consist of three random matrices: Z , η and A . Z represents the clustering of all the nodes, η the probabilities of links between and within clusters and A the links between individual nodes. The relation between the random matrices is shown in Figure 5.

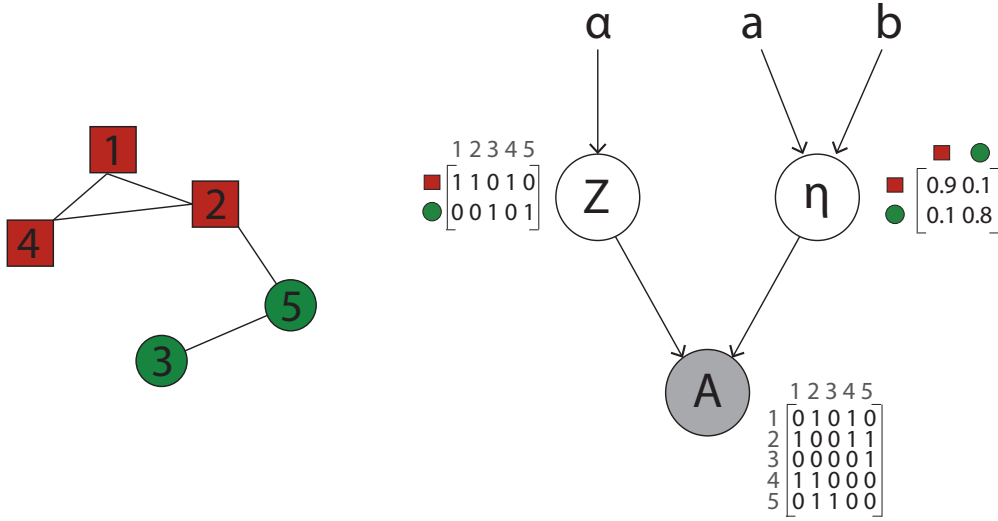


Figure 5: On the left an example of a network is shown, on the right the corresponding Infinite Relational Model.

Clustering of the nodes

In the IRM, Z is a matrix where the rows represent clusters and columns the nodes. Because every node belongs to exactly one cluster, every column has exactly one 1, indicating that this node belongs to a specific cluster. The rest of the column consists of zeros. The probability distribution of Z is induced by a process called the Chinese Restaurant Process (CRP):

$$Z \sim \text{CRP}(\alpha)$$

The CRP (Pitman, 2002) is a process that builds a clustering from the ground up, starting with adding the first node to the first cluster and adding nodes to clusters until all the nodes belong to a cluster. The CRP has one parameter, α , which influences the number of clusters that are generated. The process can be explained as follows: Every node is seen as a customer in a Chinese restaurant. This restaurant has an infinite number of tables (clusters) and every table has an infinite number of seats (the possibility of a node to belong to this cluster). Every time a customer walks into the restaurant, they are added to an already existing table or to a new table with a probability proportional to α . Customers tend to sit at more popular tables, making those even more popular. The CRP is an exchangeable process, which means that the order in which the nodes are processed does not matter.

Link probabilities

The matrix η defines the probability of a link between two nodes based on the clusters the nodes belong to. The value η_{ij} is the probability of a link between a node from cluster i and cluster j . In this thesis undirectional data is used, therefore the probability η_{ij} is equal to η_{ji} and η is a symmetrical square matrix. The probability distribution of η can be described by a Beta distribution. The values of a Beta distribution are restricted between 0 and 1. A Beta distribution has 2 parameters, a and b . These two parameters influence the shape of the distribution. Different parameters are used for within-cluster links and between-cluster links. The reason for this is to favour a higher within-cluster link density and a lower between-cluster link density. In this thesis, the values from Mørup et al. (Mørup et al., 2010) were used:

$$\eta_{ij} \sim \begin{cases} \text{Beta}(5,1) & \text{if } i = j \\ \text{Beta}(1,5) & \text{if } i \neq j \end{cases}$$

As can be seen in Figure 6 the probability that nodes within the same cluster are connected is much higher than the probability that there is a connection between nodes of a different cluster.

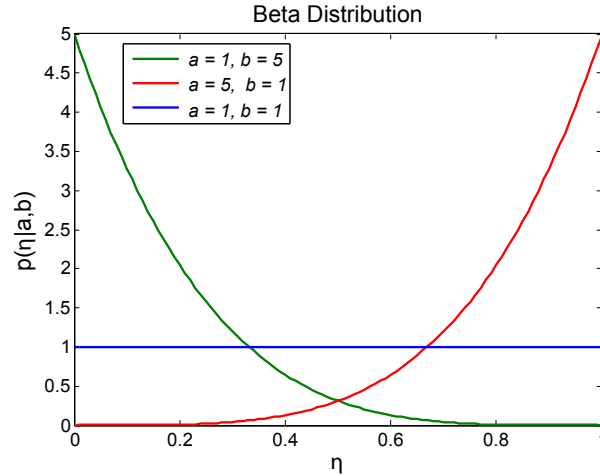


Figure 6: Beta distribution with different parameters.

Links between nodes

A is a square adjacency matrix which represents the relations between individual nodes. A depends on Z and η . The probability distribution of A is described by a Bernoulli distribution, a discrete probability distribution named after Jacob Bernoulli. The only values in this distribution are 0 and 1. The Bernoulli distribution has one parameter, p , which defines whether the value will be 0 or 1. The probability of 1 is p and the probability of 0 is $1 - p$ (Figure 7). In the IRM A is defined as follows:

$$A_{ij} \sim \text{Bernoulli}(p) \quad \text{where } p = (Z_{i_c}^T) \eta (Z_{j_c})$$

here, Z_{i_c} is the i th column of Z . As can be seen in Figure 8, p is the probability of a link between the two clusters that i and j belong to.

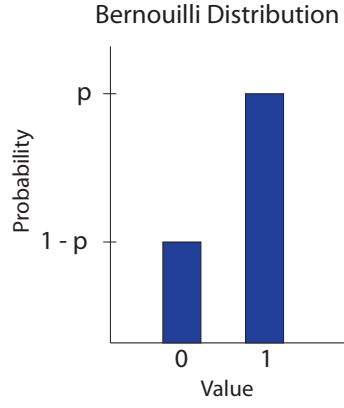


Figure 7: The Bernoulli distribution, p determines the probability of 1.

$$\begin{array}{c}
 \begin{array}{cc}
 & Z \\
 \begin{array}{c} \text{red square} \\ \text{green circle} \end{array} & \begin{bmatrix} \text{red box } i & \text{blue box } j \\ 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \end{bmatrix}
 \end{array}
 \quad
 \begin{array}{cc}
 & \eta \\
 \begin{array}{c} \text{red square} \\ \text{green circle} \end{array} & \begin{bmatrix} \text{red square} & \text{green circle} \\ 0.9 & 0.1 \\ 0.1 & 0.8 \end{bmatrix}
 \end{array}
 \end{array}$$

$$p = \underbrace{\begin{bmatrix} 1 & 0 \end{bmatrix}}_{Z_{ic}^T} \underbrace{\begin{bmatrix} 0.9 & 0.1 \\ 0.1 & 0.8 \end{bmatrix}}_{\eta} \underbrace{\begin{bmatrix} 0 \\ 1 \end{bmatrix}}_{Z_{jc}} = 0.1$$

Figure 8: The composition of p . Here, η is the probability that there is a connection between two clusters. Z determines the cluster every node belongs to. By transposing the column of the first node (i) and multiplying this with η and the column of the second node (j), p becomes the value in η that corresponds with the probability of a connection between two nodes from the cluster of i and the cluster of j .

Bayesian Community Detection

BCD is very similar to the IRM. The difference between the two model is the definition of a cluster that is used. Mørup and Schmidt give the following definition of a cluster in their paper:

”The organization of vertices in clusters, with many edges joining vertices of the same cluster and comparatively few edges joining vertices of different clusters.” (Mørup & Schmidt, 2012).

The graphical model used in BCD consists of four random matrices: Z , γ , η and A . In Figure 9 the relationship between these four random matrices is shown. The definitions of Z and A are the same as in the IRM.

Within-cluster link probability

The matrix η has, just as η in the IRM, a different distribution for the diagonal as for the rest of the matrix. The within-cluster link probability (the diagonal of the matrix) is distributed as a Beta distribution. In this thesis the same values as in the research of Mørup et al. (Mørup & Schmidt, 2012) were used:

$$\eta_{ii} \sim \text{Beta}(1, 1)$$

This caused the probability to be equally distributed (See Figure 6).

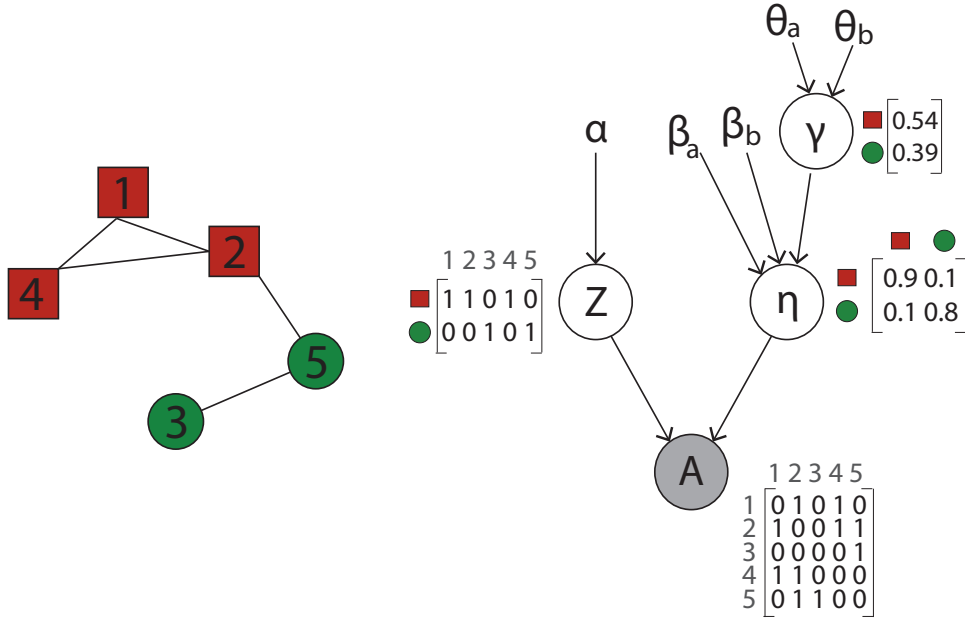


Figure 9: On the left an example of a network is shown, on the right the corresponding Bayesian Community Detection model.

Cluster gap

The matrix γ represents the cluster gap. The cluster gap is a value between 0 and 1 and forces the model to satisfy its definition of a cluster. When the cluster gap is multiplied by the within-cluster link probability, the maximum allowable between-cluster link probability is defined. The cluster gap is associated with a Beta distribution. Mørup et al. (Mørup & Schmidt, 2012) set both values to 1, in this thesis the same values were used:

$$\gamma_i \sim \text{Beta}(1, 1)$$

Between-cluster link probability

The between-cluster link probabilities are the values in η that are not on the diagonal. These values are the probabilities of links between nodes from different clusters. Because in the definition of a cluster it is specified that clusters have comparatively few edges between each other, the distribution of the between-cluster link probabilities is constrained:

$$\eta_{ij} \sim \text{BetaInc}(1, 1, x_{ij}) \quad \text{where } x_{ij} = \min[\gamma_i \eta_{ii}, \gamma_j \eta_{jj}]$$

here $\text{BetaInc}(a, b, x)$ is constrained to the interval $[0, x]$ and defined as follows:

$$p(\theta) = \frac{1}{B_x(a, b)} \theta^{a-1} (1 - \theta)^{b-1}$$

Where $B_x(a, b)$ is the constrained Beta function. As can be seen in Figure 10 the between-cluster link probability never exceeds the within-cluster link probability multiplied with the cluster gap.

Inference

When using a Bayesian model, particular variables may be observed. In the case of this thesis, the value of A was observed. What we wanted to know was the most probable value of Z . This process is called inference. To find Z the Matlab toolbox from Mørup et al. (Mørup & Schmidt, 2011) was used. In this toolbox a Gibbs sampling scheme combined with a split-merge sampling for the cluster assignment Z is implemented (Jain & Neal, 2004). This is the same sampling scheme as used by Mørup et al. (Mørup et al., 2010). Gibbs sampling is a Markov Chain Monte Carlo (MCMC) method. A Markov Chain is a chain of states where the next state is only influenced by the current state and not the previous

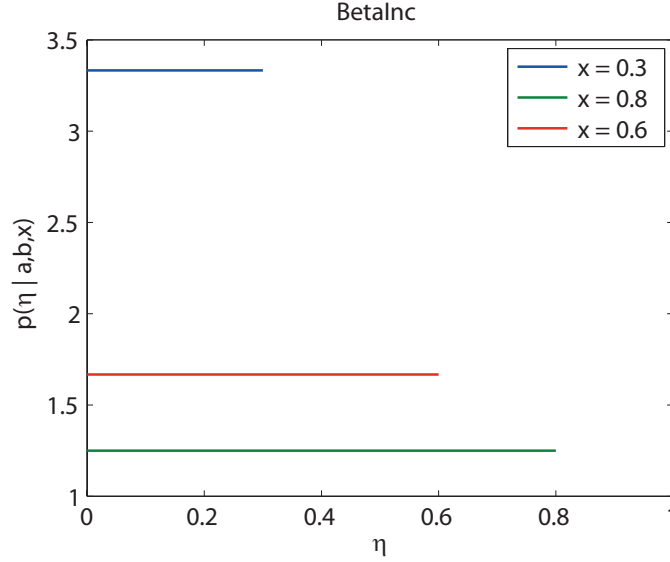


Figure 10: BetaInc distribution. Here, a and b are both set to 1, the value of x differs.

states. Monte Carlo sampling is used to estimate the distribution of a random variable by taking a large number of samples from this distribution. The goal of Gibbs sampling is to compose a Markov Chain that converges to the corresponding posterior probability distribution. Because the chain converges to the distribution, the first part of the chain is discarded from the distribution. This is called the burn-in. The split-merge sampling splits and merges clusters with a certain probability. This is combined with the gibbs sampling scheme to prevent the scheme from getting stuck in local optima.

MAP estimate

The result of MCMC clustering is the distribution of Z . Because it is difficult to draw conclusions from the comparison of two distributions, we wanted to compare the maximum a posteriori (MAP) estimate of Z . The MAP estimate is the sample that has the highest probability based on the model and the data. Because MCMC is an estimation of the distribution and not the distribution itself, taking the sample with the highest probability from the resulting distribution of Z is not necessarily the MAP estimate.

Simulated Annealing

To find the MAP estimate we used simulated annealing (SA) (Kirkpatrick, Jr., & Vecchi, 1983), an optimisation technique originating from physics. With Monte Carlo sampling a lot of samples from the distribution are drawn. SA influences the probability that a specific sample is drawn. SA is based on a temperature (T) and a decay factor (r). The temperature starts high, causing the probability that a random sample is drawn increases. In this stage a lot of samples are drawn, even when they do not have a high probability. As the temperature decreases, the probability that a sample with a lower probability than the current sample is accepted also decreases. Now only samples with very high probabilities are added to the distribution of Z . Using this algorithm the distribution slowly converges to just one sample, the MAP estimate. Simulated annealing was added to the IRM and BCD algorithms. In both algorithms the variables were initialised as follows: $T = 10$, $r = \frac{1}{2}^{\frac{10}{M}}$ where M is the maximum number of iterations. Every iteration the temperature is updated using the following linear update rule:

$$T = Tr.$$

Figure 11 shows that with these settings the temperature slowly approximates zero.

Measures

In this thesis, we tried to find which model has the most reasonable estimation of the whole-brain parcellations. When comparing two cluster algorithms, the best way to determine which one is the most

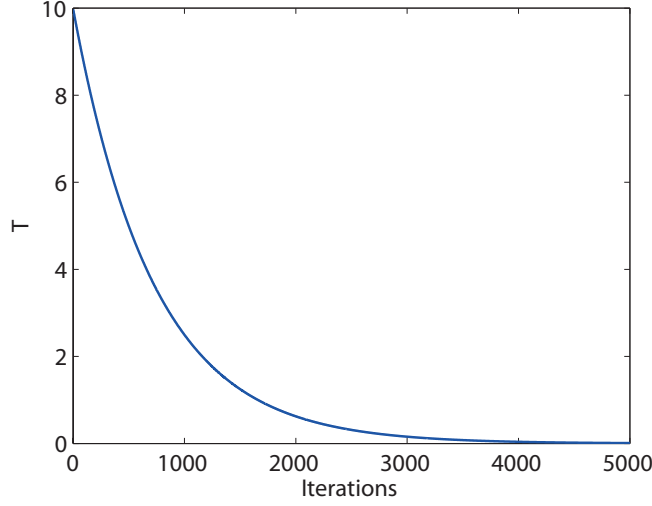


Figure 11: The T used to find the MAP estimate.

reasonable is by comparing both outcomes to the actual clustering. Unfortunately the actual clustering of the brain is not known. Therefore we needed a way to determine what is reasonable. In this thesis, we looked at size and reproducibility. The used measures are listed below.

Number of clusters

The number of clusters (NoC) is a straightforward measure. It counts the number of clusters a clustering consists of.

Number of nodes per cluster

The number of nodes per cluster is a straightforward measure. It counts the number of nodes every cluster consists of and measures the size of the clusters. This measure shows how the nodes are distributed over the clusters.

Adjusted Mutual Information

A good way of looking at a model is to measure how robust the model is. In this case robustness is measured in terms of reproducibility, which means that the model can deal with small differences in the input data (for example measurement errors) and still can produce a constant clustering. Reproducibility can tell us something about how certain the model is about the clustering. To test reproducibility, a measure is needed that can compare different clusterings. For this mutual information (MI) (Cover & Thomas, 2012) can be used. MI is a measure used to express how much information two clusterings share. It can tell us how much knowing one of the clusters reduces the uncertainty of the other. With two clusterings U (partitioned in non-overlapping subsets $\{U_1, U_2, \dots, U_R\}$) and V (partitioned in non-overlapping subsets $\{V_1, V_2, \dots, V_C\}$) MI is defined as follows:

$$I(U, V) = \sum_{i=1}^R \sum_{j=1}^C \frac{n_{ij}}{N} \log \frac{n_{ij}/N}{a_i b_j / N^2}$$

with N the total number of objects, n_{ij} the number of objects that are similar in U_i and U_j , a_i the number of objects in U_i and b_j the number of objects in V_j .

It is difficult to interpret this measure when there are clusterings that differ a lot in size. This is because the size of the cluster influences the probability that two clusterings are the same. When a clustering consists out of two clusters, the probability that two nodes are in the same cluster is a lot higher than when the clustering consists out of ten clusters. To solve this problem, the Adjusted Mutual

Information (AMI) (Vinh, Epps, & Bailey, 2010b) is used. This measure corrects the MI for chance, it takes the probability that two nodes are in the same cluster into account. The AMI is defined as follows:

$$AMI(U, V) = \frac{I(U, V) - E\{I(U, V)\}}{\max\{H(U), H(V)\} - E\{I(U, V)\}}$$

where $H(U)$ is the entropy of U , defined as follows:

$$H(U) = - \sum_{i=1}^R \frac{a_i}{N} \log \frac{a_i}{N}$$

and $E\{I(U, V)\}$ is the expected value for the mutual information (Vinh, Epps, & Bailey, 2009), defined as:

$$E\{I(U, V)\} = \sum_{i=1}^R \sum_{j=1}^C \sum_{n_{ij}=\max(a_i+b_j)-N}^{\min(a_i, b_j)} \frac{n_{ij}}{N} \log \frac{N n_{ij}}{a_i b_j} \frac{a_i! b_j! (N - a_i)! (N - b_j)!}{N! n_{ij}! (a_i - n_{ij})! (b_j - n_{ij})! (N - a_i - b_j + n_{ij})!}$$

To calculate the Adjusted Mutual Information a Matlab toolbox (Vinh, Epps, & Bailey, 2010a) was used.

Data

Twenty healthy subjects were scanned and diffusion-weighted images were obtained using a Siemens Magnetom Trio 3T System at the Donders Centre for Cognitive Neuroimaging, Radboud University Nijmegen, The Netherlands. Per voxel (a 3D pixel) 5000 streamlines were drawn. Streamlines with a sharp angle or shorter than 2 mm were discarded. Then the connections were generated. When there was a streamline from a to b and also from b to a, a and b were assumed to be connected. This constraint was added to eliminate false positives. After connecting the voxels, we grouped them within 116 AAL regions. AAL (Automated Anatomical Labelling) is a digital atlas of the human brain, developed by a French research group (Tzourio-Mazoyer et al., 2002). There is a connection between two AAL regions when there is at least one voxel in one region that is connected with a voxel in another region. The resulting adjacency matrices of all the twenty subjects were used as input to our Bayesian clustering algorithms.

Results

The adjacency matrix A for every subject was processed by the Matlab toolbox. We used 5000 iterations and $\alpha = \log N$, where N is the number of iterations. This resulted into forty clusterings, two per subject. On these data clustering measures were computed.

Number of Clusters

For every subject, the number of clusters were counted. As can be seen in Figure 12, the IRM generated less clusters than the BCD. A paired samples t-test indicates a statistically reliable difference between the number of clusters of the IRM ($M = 7.8$, $SE = 0.186$) and BCD ($M = 11.75$, $SE = 0.323$), $t(19) = 13.841$, $p \approx 0$, $\alpha = 0.05$. A full overview of the number of clusters per subject can be found in the appendix.

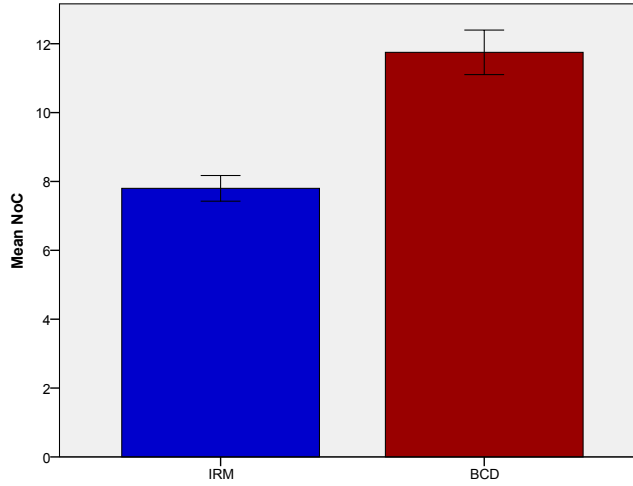


Figure 12: The means (IRM = 7.8, BCD = 11.75) of the number of clusters of both models. The error bars represent 2 times the standard error of the mean (IRM = 0.186, BCD = 0.323).

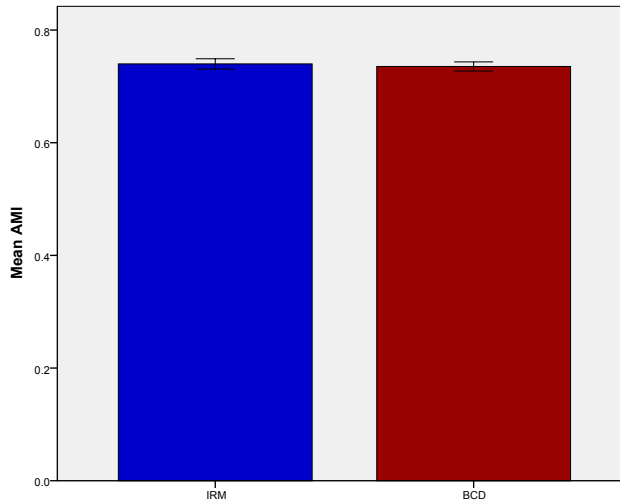


Figure 13: The means (IRM = 0.7399, BCD = 0.7353) of the adjusted mutual information of both models. The error bars represent 2 times the standard error of the mean (IRM = 0.00456, BCD = 0.00407).

Adjusted Mutual Information

The AMI was calculated for every combination of the twenty subjects. In Figure 13 the value of every combination of both models is plotted. As can be seen, the AMI of the models was almost the same.

A paired samples t-test failed to indicate a statistically reliable difference between the AMI of the IRM ($M = 0.7399$, $SE = 0.00456$) and the AMI of BCD ($M = 0.7353$, $SE = 0.00407$), $t(189) = 0.782$, $p = 0.435$, $\alpha = 0.05$.

Number of nodes per cluster

For every cluster produced by the IRM and BCD, the number of nodes per cluster were counted (See Figure 14). As can be seen, both the IRM and BCD produce a lot of clusters that consist of one node. BCD produced a lot of smaller clusters from five to fifteen nodes, while the IRM produced more clusters of a size between the twenty and twenty-five nodes. This was expected since BCD produced more clusters per subject, therefore the clusters will be smaller on average because the number of data points are the same. Noticeably, the IRM produced a lot of very large clusters, from 40 tot 45 nodes, while the biggest node the BCD produced consists of twenty-eight nodes.

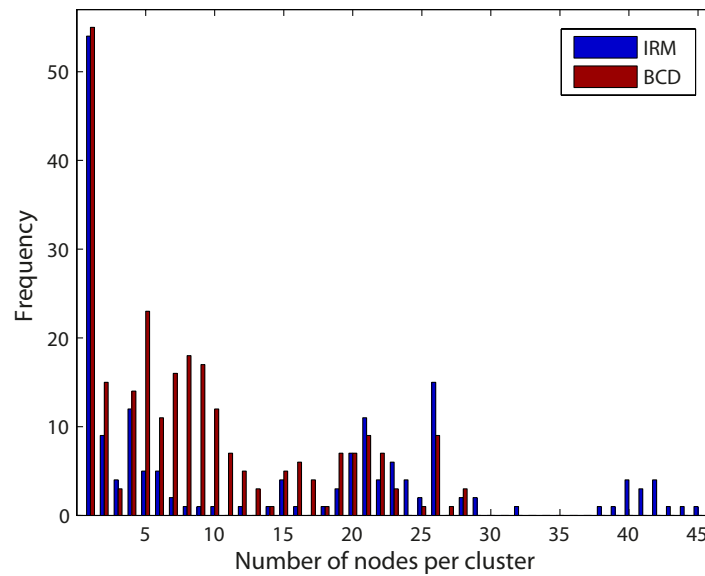


Figure 14: The number of nodes per cluster for both models. For every cluster that is generated, the number of nodes is counted and plotted above.

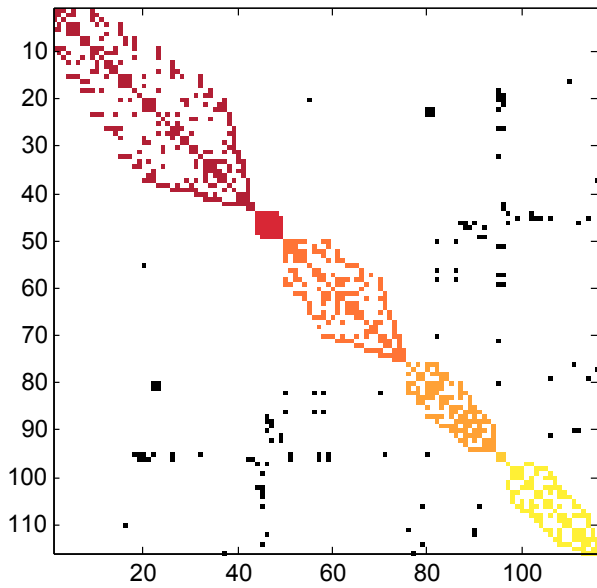
Clusters

Although the actual clustering of the brain is not known and therefore can not be used to judge the quality of the produced clusters, some things can be said about the clusters that are produced by the two algorithms. Because the two models mainly differ on the number of clusters, we decided to look at the subject with the least number of clusters with IRM (sparse clustering) and the subject with the most number of clusters with BCD (dense clustering).

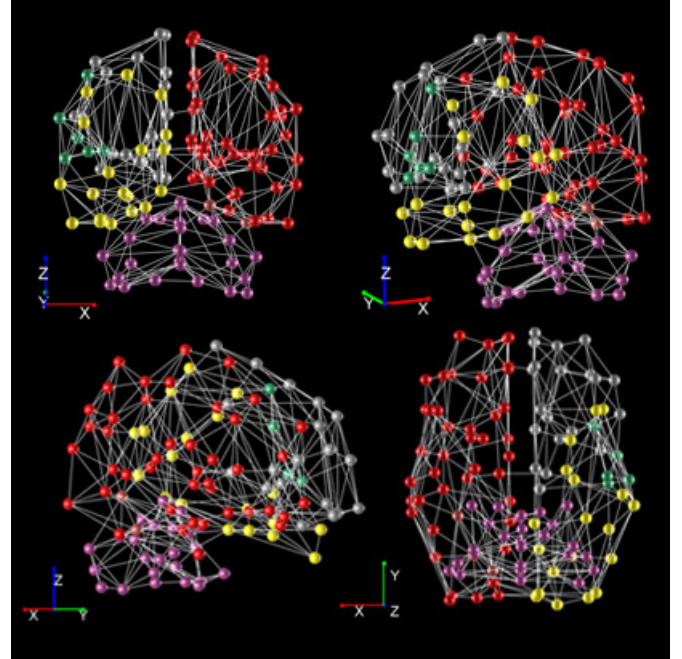
Sparse clustering results

The clustering with the least number of clusters formed by the IRM is Subject 7. Figure 15 shows the IRM's clustering that belongs to this subject. In Figure 15a can be seen that most connections are within a cluster. Some clusters have more connections than others. Figure 15b shows that the IRM was able to split the left hemisphere from the right hemisphere. The cerebellum is also clustered together. When we look at the size and locations of the clusters, it stands out that the red cluster almost completely covers the right hemisphere, while the left hemisphere is split up into multiple clusters.

In Figure 16 BCD's clustering of Subject 7 is shown. This clustering consists of more clusters than IRM's clustering. As can be seen in Figure 16a, more connections exist within clusters than between clusters. Also, clusters are more equal of size. Figure 16b shows that BCD was, just as the IRM, able to split the left and the right hemisphere. Also the cerebellum is clustered together. When comparing the



(a)



(b)

Figure 15: IRM's clustering of Subject 7. In (a) a sorted adjacency matrix of the clustering can be seen. Black squares indicate connections between nodes from different clusters, coloured squares indicate connections between nodes from the same cluster. Every cluster has a new colour. The nodes are sorted per cluster, so the order of the nodes differs from other sorted adjacency matrices. In (b) four different views of the 3D model of the clustering are plotted. For the 3D model, the VisualConnectome toolbox (Dai & He, 2011) was used.

clusters that are produced by the two algorithms, it is striking that the clusters of BCD are almost sub clusters of the IRM. When the clusters of BCD are combined, they almost form the clustering of IRM. An overview of the location of the clustered nodes for both models is given in the appendix.

Dense clustering results

BCD produces for Subjects 14, 16 and 19 a clustering that consists out of 14 nodes. We decided to look at Subject 14. In Figure 17 the clustering of BCD is shown. With this subject there are some clusters that have more links with nodes outside the cluster than within the cluster; for example Cluster 8 (the central green cluster, see Figure 17a). This clustering also contains a lot of very small clusters. In Figure 17b can be seen that BCD was able to separate the hemispheres. The cerebellum is not clustered into one cluster, but divided into multiple clusters.

In Figure 18 IRM's clustering of Subject 14 is shown. As can be seen, this clustering has less clusters than the clustering produced by BCD. Figure 18a shows that some clusters have more connections with nodes from other clusters than with nodes within the same cluster, just like the clustering BCD produced. Also, The IRM was able to split the left and the right hemisphere. The cerebellum is clustered into one cluster. Noticeably, the blue cluster almost covers the complete right hemisphere (See Figure 18b). In the appendix a full overview of the regions of the nodes per cluster is given. As can be seen in the appendix, a lot of clusters consist out of the same nodes. Some of the clusters of BCD are sub clusters of the clusters of IRM, although the differences are more pronounced compared to the results of Subject 7.

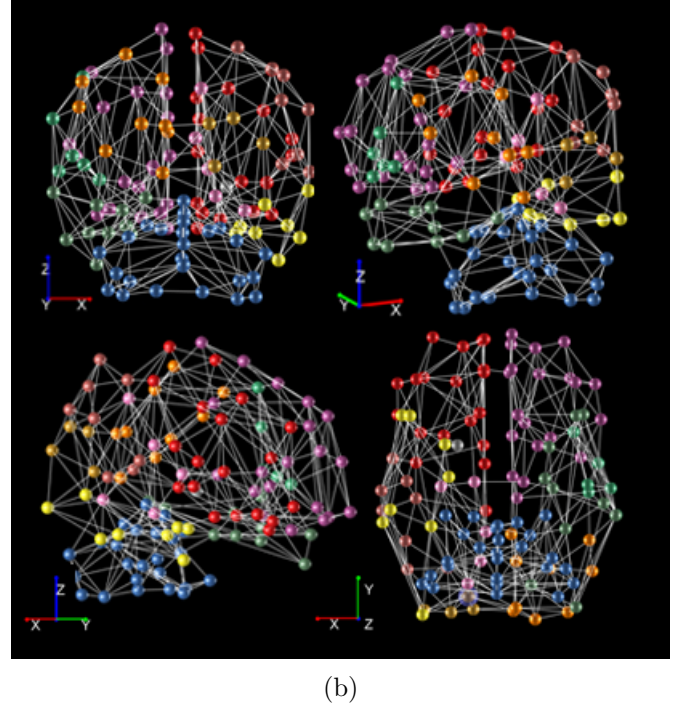
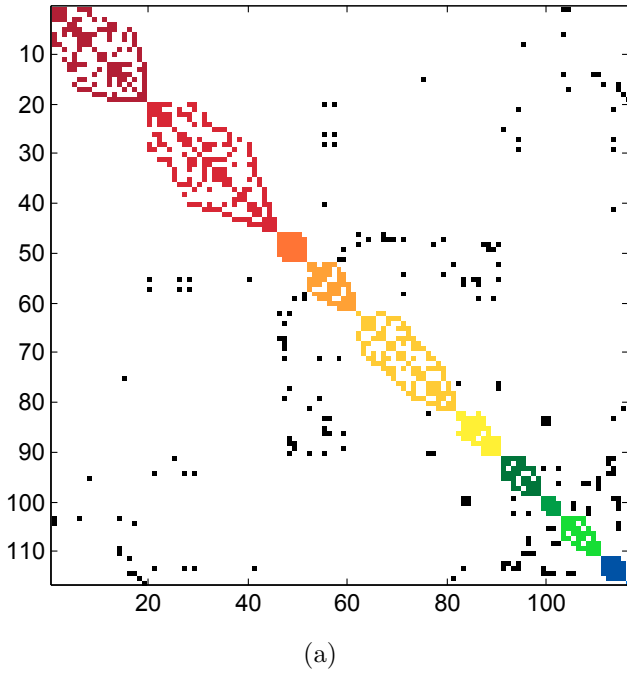


Figure 16: BCD's clustering of Subject 7. In (a) a sorted adjacency matrix of the clustering can be seen. Black squares indicate connections between nodes from different clusters, coloured squares indicate connections between nodes from the same cluster. Every cluster has a new colour. The nodes are sorted per cluster, so the order of the nodes differs from other sorted adjacency matrices. In (b) four different views of the 3D model of the clustering are plotted. For the 3D model, the VisualConnectome toolbox (Dai & He, 2011) was used.

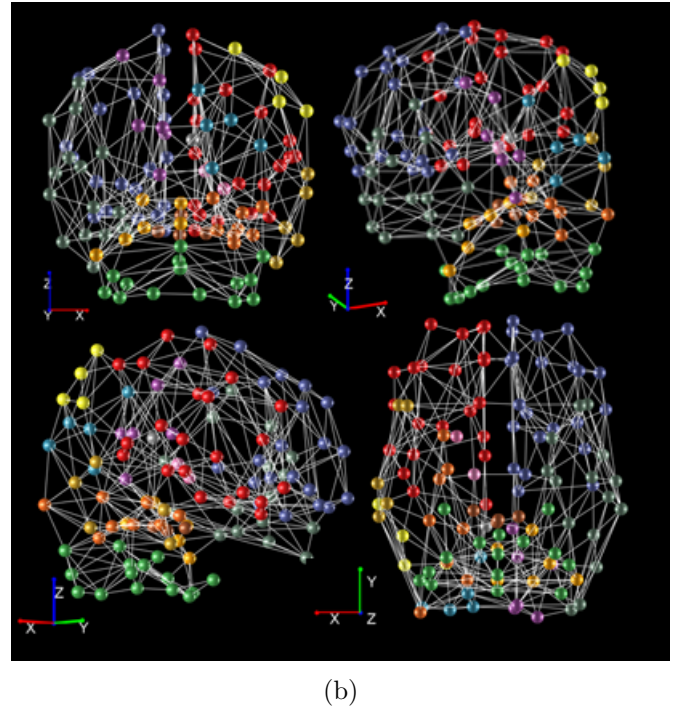
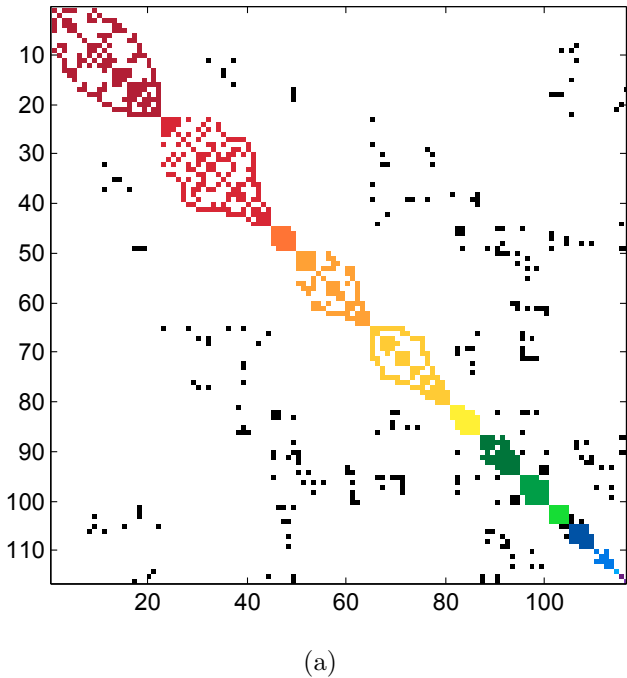


Figure 17: BCD's clustering of Subject 14. In (a) a sorted adjacency matrix of the clustering can be seen. Black squares indicate connections between nodes from different clusters, coloured squares indicate connections between nodes from the same cluster. Every cluster has a new colour. The nodes are sorted per cluster, so the order of the nodes differs from other sorted adjacency matrices. In (b) four different views of the 3D model of the clustering are plotted. For the 3D model, the VisualConnectome toolbox (Dai & He, 2011) was used.

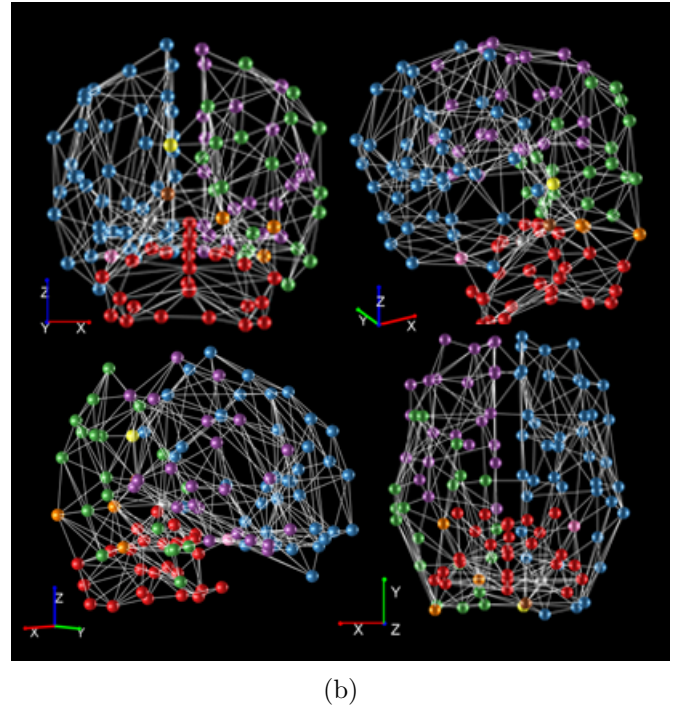
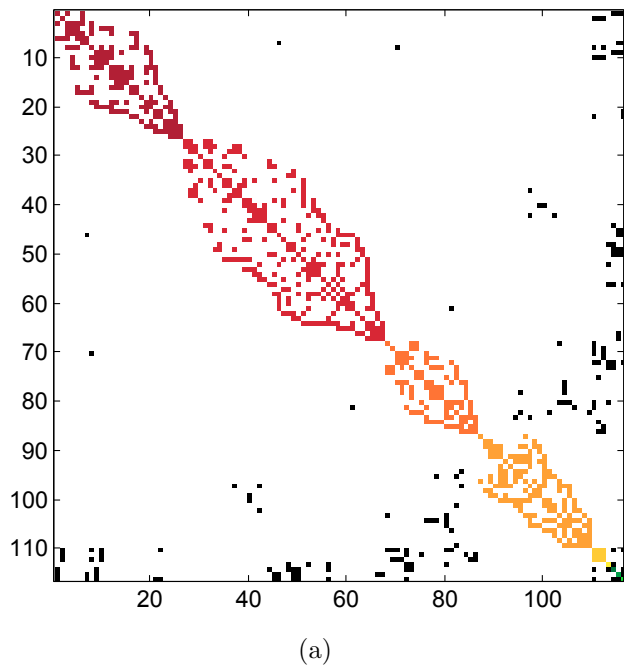


Figure 18: IRM's clustering of Subject 14. In (a) a sorted adjacency matrix of the clustering can be seen. Black squares indicate connections between nodes from different clusters, coloured squares indicate connections between nodes from the same cluster. Every cluster has a new colour. The nodes are sorted per cluster, so the order of the nodes differs from other sorted adjacency matrices. In (b) four different views of the 3D model of the clustering are plotted. For the 3D model, the VisualConnectome toolbox (Dai & He, 2011) was used.

Discussion

In this thesis, two Bayesian clustering methods were tested and compared to answer the following research question:

Does Bayesian Community Detection lead to more reasonable estimates of whole-brain parcellations compared to the Infinite Relational model?

To answer this question, the structural connectivity data of twenty subjects was clustered by both models. Simulated annealing was implemented to find the MAP estimate. Because the actual clustering of the brain is unknown, several measures were used to compare the two models. It was found that the number of clusters produced by BCD is significantly higher than the number of clusters produced by the IRM. The adjusted mutual index, used to measure reproducibility, is the same in both models. The IRM produced clusters that contain more nodes than BCD. When examining the location of the clusters, both models were able to distinguish the left and right hemisphere. In most cases the cerebellum was detected as a single. Some clusters of BCD were sub clusters of the IRM, which means that the same nodes were clustered together. In some cases a cluster of the IRM could be exactly reproduced by combining multiple clusters of BCD.

The main difference between the two models is the number of clusters and the size of the clusters. In both subjects discussed the IRM produced a cluster that almost completely covered one hemisphere. This makes the clustering of the IRM very generic. Also, when the clusters of IRM are combined, they almost form the clustering of the IRM. Thus the clusterings of BCD are more specific and IRM and BCD cluster the same nodes together. Because of this, BCD leads to more reasonable estimates of the whole brain parcellation compared to the IRM. In the follow, the main findings will be discussed into more detail.

MAP estimate

As a caveat, after running the same data multiple times, not the exact same MAP estimates were found. We observed only minor differences between the found MAP estimates. Therefore we assumed that these matrices were at least very close to the actual MAP estimate, but it could also be possible that there were multiple local optima found instead of the MAP estimate. This could have influenced the test results. To solve this, a lot more samples could be drawn and a slower cooling schedule could be used.

Adaption of BCD

In the result of Subject 14, the clustering of BCD contained clusters that had more between-cluster links than within-cluster links. IRM's clustering also had this type of clusters. With IRM, this is not a very strange observation. The probabilities of the links were determined by a Beta distribution. Although the parameters differed between within- and between-cluster links, it was still possible for the between cluster link probability to become higher than the within-cluster link probability (See Figure 6). With BCD, this was not the case. The definition of a cluster is as follows:

"The organization of vertices in clusters, with many edges joining vertices of the same cluster and comparatively few edges joining vertices of different clusters." (Mørup & Schmidt, 2012).

This definition is reflected in the between-cluster link probability. This probability is defined by a Beta distribution constrained by the cluster gap multiplied by the within-cluster link probability. Since the cluster gap is a value between 0 and 1, the between-cluster link probability will always be smaller than the within-cluster link probability. We felt the observation that clusters have more between-cluster links than within-cluster links was contradicting with the definition of a cluster given in the paper. Of course, η only consists of probabilities. The lower probability does not mean that it is not possible, just less likely. However, when we considered the definition and the model, we realised that the definition of a cluster could be understood in two different ways:

1. When considering one node, the probability that there is a connection between this node and a node from the same cluster is comparatively higher than the probability that there is connection between this node and a node from another cluster.

2. When considering one node, the probability that there is a connection between this node and a node from the same cluster is comparatively higher than the probability this node has a connection with *any* node from another cluster.

The difference is in the words *a* and *any*. The first definition is used in BCD. But when following this interpretation some strange effects can occur. To illustrate this effect, we use the following example: A data set consists of a lot of clusters and one cluster consists of very few nodes. When following the first interpretation of the definition, a node of this small cluster will have a higher probability to be connected to a node outside of his cluster than to a node within its cluster. Although the probability that a node is connected with a specific node within the same cluster is higher than with a specific node from another cluster, there are much more nodes outside the cluster than inside the cluster. This causes the probability of a connection outside the cluster to become higher than the probability of a connection inside the cluster. Following the second definition could reduce this effect. To follow this interpretation some adjustments should be made to the BCD. Instead of constraining every value of η , the sum of the values of the rows and columns (except the diagonal) should be constrained. Further research is needed to investigate whether this adjustment improves the performance of the model.

Improvements and future research

Various other analysis can be performed in order to further investigate the differences between the two models. Firstly, the models can be tested on more subjects. Another thing that can be done is to test the reproducibility on another level. AAL regions are fairly big and assumed to be almost the same between different humans. Differences between the connections of AAL regions will therefore probably be measurement errors. But when looking at a more detailed level, brains do differ from each other. A more fine-grained representation of the brain can be used to test how the models perform in such setting. Finally, although the actual clustering of the brain is unknown, more of the brain is known than used in this thesis. This knowledge can be used to judge the quality of the clusters produced by the two algorithms.

Although a lot more research can be done, this thesis provides a clear overview of the models and can be used to make a weighted decision between the two.

References

- Beckmann, M., Johansen-Berg, H., & Rushworth, M. F. (2009). Connectivity-Based Parcellation of Human Cingulate Cortex and Its Relation to Functional Specialization. *The Journal of Neuroscience*, 29(4), 1175.
- Cover, T. M., & Thomas, J. A. (2012). *Elements of Information Theory*. John Wiley & Sons.
- Dai, D., & He, H. (2011). *VisualConnectome: Toolbox for brain network visualization and analysis*. Available from <http://code.google.com/p/visualconnectome/>
- Fischl, B., Rajendran, N., et al. (2008). Cortical Folding Patterns and Predicting Cytoarchitecture. *Cerebral Cortex*, 18(8), 1973.
- Fischl, B., Van Der Kouwe, A., et al. (2004). Automatically Parcellating the Human Cerebral Cortex. *Cerebral Cortex*, 14(1), 11.
- Honey, C. J., Thivierge, J.-P., & Sporns, O. (2010). Can structure predict function in the human brain? *NeuroImage*, 52(3), 766.
- Jain, S., & Neal, R. M. (2004). A Split-Merge Markov Chain Monte Carlo Procedure for the Dirichlet Process Mixture Model. *Journal of Computational and Graphical Statistics*, 13(1), 158.
- Jbabdi, S., Woolrich, M. W., & Behrens, T. E. J. (2009). Multiple-subjects connectivity-based parcellation using hierarchical Dirichlet process mixture models. *NeuroImage*, 44(2), 373.
- Kemp, C., Tenenbaum, J. B., Griffiths, T. L., Yamada, T., & Ueda, N. (2006). Learning Systems of Concepts with an Infinite Relational Model. In *Proceedings of the national conference on artificial intelligence* (Vol. 21, p. 381).
- Kirkpatrick, S., Jr., D. G., & Vecchi, M. P. (1983). Optimization by Simulated Annealing. *Science*, 220(4598), 671.
- Mars, R. B., Sallet, J., Schüffelgen, U., Jbabdi, S., Toni, I., & Rushworth, M. F. (2012). Connectivity-Based Subdivisions of the Human Right “Temporoparietal Junction Area”: Evidence for Different Areas Participating in Different Cortical Networks. *Cerebral Cortex*, 22(8), 1894.
- Mørup, M., Madsen, K. H., Dogonowski, A. M., Siebner, H., & Hansen, L. K. (2010). Infinite Relational Modeling of Functional Connectivity in Resting State fMRI. *Neural Information Processing Systems*, 23, 1750.
- Mørup, M., & Schmidt, M. N. (2011). *Matlab code for Bayesian Community Detection*. Available from http://www2.imm.dtu.dk/pubdb/views/publication_details.php?id=6147
- Mørup, M., & Schmidt, M. N. (2012). Bayesian Community Detection. *Neural Computation*, 24(9), 2434.
- Passingham, R. E., Stephan, K. E., & Kötter, R. (2002). The Anatomical Basis of Functional Localization in the Cortex. *Nature Reviews Neuroscience*, 3(8), 606.
- Pitman, J. (2002). *Combinatorial Stochastic Processes* (Vol. 1875; Tech. Rep.).
- Russell, S. J., Norvig, P., Canny, J. F., Malik, J. M., & Edwards, D. D. (1995). *Artificial Intelligence: a Modern Approach* (Vol. 74). Prentice hall Englewood Cliffs.
- Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., et al. (2002). Automated Anatomical Labeling of Activations in SPM using a Macroscopic Anatomical Parcellation of the MNI MRI Single-Subject Brain. *NeuroImage*, 15(1), 273.
- Vinh, N. X., Epps, J., & Bailey, J. (2009). Information Theoretic Measures for Clusterings Comparison: Is a Correction for Chance Necessary? In *Proceedings of the 26th Annual International Conference on Machine Learning* (p. 1073).
- Vinh, N. X., Epps, J., & Bailey, J. (2010a). *Code for computing the Adjusted Mutual Information (AMI) in Matlab*. Available from <https://sites.google.com/site/vinhnguyenx/software>
- Vinh, N. X., Epps, J., & Bailey, J. (2010b). Information Theoretic Measures for Clusterings Comparison: Variants, Properties, Normalization and Correction for Chance. *The Journal of Machine Learning Research*, 9999, 2837.

Appendix

Number of Clusters

Subject	IRM	BCD
1	7	11
2	8	11
3	8	13
4	8	11
5	9	11
6	7	11
7	6	11
8	7	11
9	8	10
10	8	13
11	9	13
12	8	12
13	8	12
14	9	14
15	7	9
16	8	14
17	7	13
18	8	10
19	9	14
20	7	11

IRM clustering of Subject 7

Cluster	Regions
1	Frontal_Sup_L Frontal_Mid_L Frontal_Inf_Oper_L Frontal_Inf_Orb_L Supp_Motor_Area_L Frontal_Sup_Medial_L Rectus_L Cingulum_Ant_L Cingu- lum_Post_L ParaHippocampal_L Calcarine_L Lingual_L Occipital_Mid_L Fusiform_L Parietal_Sup_L SupraMarginal_L Precuneus_L Caudate_L Pallidum_L Heschl_L Temporal_Pole_Sup_L Temporal_Pole_Mid_L Precentral_R Frontal_Mid_Orb_R Frontal_Inf_Tri_R Rolandic_Oper_R Frontal_Mid_Orb_R Insula_R Cingulum_Mid_R Hippocampus_R Amygdala_R Cuneus_R Occipital_Sup_R Occipital_Inf_R Postcen- tral_R Parietal_Inf_R Angular_R Paracentral_Lobule_R Putamen_R Thalamus_R Temporal_Sup_R Temporal_Mid_R Temporal_Inf_R
2	Cingulum_Mid_L Postcentral_L Frontal_Sup_Medial_R Cingulum_Post_R Pre- cuneus_R Caudate_R
3	Cerebelum_Crus1_L Cerebelum_Crus1_R Cerebelum_Crus2_L Cerebelum_Crus2_R Cerebelum_3_L Cerebelum_3_R Cerebelum_4.5_L Cerebelum_4.5_R Cerebelum_6_L Cerebelum_6_R Cerebelum_7b_L Cerebelum_7b_R Cerebelum_8_L Cerebelum_8_R Cerebelum_9_L Cerebelum_9_R Cerebelum_10_L Cerebelum_10_R Vermis_1.2 Ver- mis_3 Vermis_4.5 Vermis_6 Vermis_7 Vermis_8 Vermis_9 Vermis_10
4	Paracentral_Lobule_L Putamen_L Thalamus_L Temporal_Sup_L Tempo- ral_Mid_L Temporal_Inf_L Frontal_Sup_R Frontal_Mid_R Frontal_Inf_Oper_R Frontal_Inf_Orb_R Supp_Motor_Area_R Rectus_R Cingulum_Ant_R ParaHip- pocampal_R Calcarine_R Pallidum_R Heschl_R Temporal_Pole_Sup_R Tempo- ral_Pole_Mid_R
5	Frontal_Sup_Orb_R Olfactory_R
6	Precentral_L Frontal_Sup_Orb_L Frontal_Mid_Orb_L Frontal_Inf_Tri_L Rolandic_Oper_L Olfactory_L Frontal_Mid_Orb_L Insula_L Hippocampus_L Amygdala_L Cuneus_L Occipital_Sup_L Occipital_Inf_L Parietal_Inf_L Angular_L Lingual_R Occipital_Mid_R Fusiform_R Parietal_Sup_R SupraMarginal_R

BCD clustering of Subject 7

Cluster	Regions
1	Frontal_Sup_L Frontal_Mid_L Frontal_Inf_Oper_L Frontal_Inf_Orb_L Supp_Motor_Area_L Frontal_Sup_Medial_L Rectus_L Cingulum_Ant_L ParaHippocampal_L Calcarine_L Lingual_L Occipital_Mid_L Fusiform_L Parietal_Sup_L SupraMarginal_L Precuneus_L Occipital_Sup_R Occipital_Inf_R Postcentral_R
2	Cerebelum_Crus1_L Cerebelum_Crus1_R Cerebelum_Crus2_L Cerebelum_Crus2_R Cerebelum_3_L Cerebelum_3_R Cerebelum_4_5_L Cerebelum_4_5_R Cerebelum_6_L Cerebelum_6_R Cerebelum_7b_L Cerebelum_7b_R Cerebelum_8_L Cerebelum_8_R Cerebelum_9_L Cerebelum_9_R Cerebelum_10_L Cerebelum_10_R Vermis_1_2 Ver- mis_3 Vermis_4_5 Vermis_6 Vermis_7 Vermis_8 Vermis_9 Vermis_10
3	Cingulum_Mid_L Postcentral_L Frontal_Sup_Medial_R Cingulum_Post_R Pre- cuneus_R Caudate_R
4	Putamen_L Thalamus_L Temporal_Sup_L Frontal_Sup_R Frontal_Inf_Orb_R Supp_Motor_Area_R Pallidum_R Heschl_R Temporal_Pole_Sup_R Tempo- ral_Pole_Mid_R
5	Precentral_L Frontal_Sup_Orb_L Frontal_Mid_Orb_L Frontal_Inf_Tri_L Rolandic_Oper_L Olfactory_L Frontal_Mid_Orb_L Insula_L Hippocampus_L Amygdala_L Cuneus_L Occipital_Sup_L Occipital_Inf_L Parietal_Inf_L Angular_L Lingual_R Occipital_Mid_R Fusiform_R Parietal_Sup_R SupraMarginal_R
6	Paracentral_Lobule_L Temporal_Mid_L Temporal_Inf_L Frontal_Mid_R Frontal_Inf_Oper_R Rectus_R Cingulum_Ant_R ParaHippocampal_R Calcarine_R
7	Heschl_L Temporal_Pole_Sup_L Rolandic_Oper_R Olfactory_R Thalamus_R Tempo- ral_Sup_R Temporal_Mid_R Temporal_Inf_R
8	Temporal_Pole_Mid_L Precentral_R Frontal_Mid_Orb_R Frontal_Inf_Tri_R
9	Cingulum_Post_L Frontal_Mid_Orb_R Insula_R Cingulum_Mid_R Hippocampus_R Amygdala_R Paracentral_Lobule_R Putamen_R
10	Caudate_L Pallidum_L Frontal_Sup_Orb_R Cuneus_R Angular_R
11	Parietal_Inf_R

IRM clustering of Subject 14

Cluster	Regions
1	Cerebelum_Crus1_L Cerebelum_Crus1_R Cerebelum_Crus2_L Cerebelum_Crus2_R Cerebelum_3_L Cerebelum_3_R Cerebelum_4_5_L Cerebelum_4_5_R Cerebelum_6_L Cerebelum_6_R Cerebelum_7b_L Cerebelum_7b_R Cerebelum_8_L Cerebelum_8_R Cerebelum_9_L Cerebelum_9_R Cerebelum_10_L Cerebelum_10_R Vermis_1_2 Ver- mis_3 Vermis_4_5 Vermis_6 Vermis_7 Vermis_8 Vermis_9 Vermis_10
2	Precentral_L Frontal_Sup_Orb_L Frontal_Mid_Orb_L Frontal_Inf_Tri_L Rolandic_Oper_L Olfactory_L Frontal_Mid_Orb_L Insula_L Cingulum_Mid_L Hippocampus_L Amygdala_L Cuneus_L Occipital_Sup_L Occipital_Inf_L Post- central_L Parietal_Inf_L Angular_L Paracentral_Lobule_L Putamen_L Thalamus_L Temporal_Sup_L Frontal_Mid_R Frontal_Inf_Oper_R Frontal_Inf_Orb_R Frontal_Sup_Medial_R Rectus_R Cingulum_Ant_R Cingulum_Post_R ParaHip- pocampal_R Calcarine_R Lingual_R Occipital_Mid_R Fusiform_R Parietal_Sup_R SupraMarginal_R Precuneus_R Caudate_R Pallidum_R Heschl_R Temporal_Pole_Sup_R Temporal_Pole_Mid_R
3	Caudate_L Pallidum_L Heschl_L Temporal_Pole_Sup_L Temporal_Pole_Mid_L Pre- central_R Frontal_Mid_Orb_R Frontal_Inf_Tri_R Insula_R Cingulum_Mid_R Hip- pocampus_R Amygdala_R Cuneus_R Angular_R Putamen_R Thalamus_R Temporal_Sup_R Temporal_Mid_R Temporal_Inf_R
4	Frontal_Sup_L Frontal_Mid_L Frontal_Inf_Oper_L Frontal_Inf_Orb_L Supp_Motor_Area_L Frontal_Sup_Medial_L Rectus_L Cingulum_Ant_L Cingu- lum_Post_L ParaHippocampal_L Calcarine_L Lingual_L Occipital_Mid_L Fusiform_L Parietal_Sup_L SupraMarginal_L Precuneus_L Frontal_Mid_Orb_R Occipital_Sup_R Occipital_Inf_R Postcentral_R Parietal_Inf_R Paracentral_Lobule_R
5	Frontal_Sup_Orb_R Rolandic_Oper_R Olfactory_R
6	Temporal_Inf_L
7	Temporal_Mid_L
8	Supp_Motor_Area_R
9	Frontal_Sup_R

BCD clustering of Subject 14

Cluster	Regions
1	Frontal_Sup_L Frontal_Mid_L Frontal_Inf_Oper_L Frontal_Inf_Orb_L Supp_Motor_Area_L Frontal_Sup_Medial_L Rectus_L Cingulum_Ant_L Cingu- lum_Post_L ParaHippocampal_L Calcarine_L Lingual_L Occipital_Mid_L Fusiform_L Parietal_Sup_L SupraMarginal_L Precuneus_L Frontal_Mid_Orb_R Occipital_Sup_R Occipital_Inf_R Postcentral_R Paracentral_Lobule_R
2	Precentral_L Frontal_Sup_Orb_L Frontal_Mid_Orb_L Frontal_Inf_Tri_L Rolandic_Oper_L Olfactory_L Frontal_Mid_Orb_L Insula_L Hippocampus_L Amygdala_L Cuneus_L Occipital_Sup_L Occipital_Inf_L Postcentral_L Parietal_Inf_L Angular_L Frontal_Sup_Medial_R Lingual_R Occipital_Mid_R Fusiform_R Pari- etal_Sup_R SupraMarginal_R
3	Temporal_Pole_Mid_L Precentral_R Frontal_Mid_Orb_R Frontal_Inf_Tri_R Cuneus_R
4	Cerebelum_Crus1_R Cerebelum_Crus2_L Cerebelum_Crus2_R Cerebelum_7b_L Cere- belum_7b_R Cerebelum_8_L Cerebelum_8_R Cerebelum_9_L Cerebelum_9_R Cere- belum_10_L Cerebelum_10_R Vermis_7 Vermis_8 Vermis_9 Vermis_10
5	Cingulum_Mid_L Putamen_L Thalamus_L Temporal_Sup_L Frontal_Inf_Oper_R Frontal_Inf_Orb_R Supp_Motor_Area_R Cingulum_Ant_R Cingulum_Post_R ParaHippocampal_R Precuneus_R Caudate_R Pallidum_R Heschl_R Tempo- ral_Pole_Sup_R Temporal_Pole_Mid_R
6	Paracentral_Lobule_L Temporal_Mid_L Temporal_Inf_L Frontal_Mid_R Rectus_R Calcarine_R
7	Pallidum_L Heschl_L Temporal_Pole_Sup_L Frontal_Sup_Orb_R Rolandic_Oper_R Olfactory_R Cerebelum_4.5_R Cerebelum_6_R
8	Frontal_Sup_R Cerebelum_Crus1_L Cerebelum_4.5_L Cerebelum_6_L Vermis_4.5 Vermis_6
9	Insula_R Cingulum_Mid_R Hippocampus_R Amygdala_R
10	Putamen_R Thalamus_R Temporal_Sup_R Temporal_Mid_R Temporal_Inf_R
11	Cerebelum_3_L Cerebelum_3_R Vermis_1.2 Vermis_3
12	Parietal_Inf_R
13	Angular_R
14	Caudate_L