Non-parametric Bayesian graph models reveal community structure in resting state fMRI

Kasper Winther Andersen[a,b,*], Kristoffer H. Madsen[b], Hartwig Roman Siebner[b,c,d], Mikkel N. Schmidt[a], Morten Mørup[a], Lars Kai Hansen[a]

[a] *Department of Applied Mathematics and Computer Science, Technical University of Denmark, Matematiktorvet, Bygning 303 B, 2800 Kgs. Lyngby, Denmark*
[b] *Danish Research Centre for Magnetic Resonance, Centre for Functional and Diagnostic Imaging and Research, Copenhagen University Hospital Hvidovre, Kettegaard Alle 30, 2650 Hvidovre, Denmark*
[c] *Department of Clinical Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, Blegdamsvej 3B, 2200 København N, Denmark*
[d] *Department of Neurology, Copenhagen University Hospital Bispebjerg, Bispebjerg Bakke 23, 2400 København NV, Denmark*

Abstract

Modeling of resting state functional magnetic resonance imaging (rs-fMRI) data using network models is of increasing interest. It is often desirable to group nodes into clusters to interpret the communication patterns between nodes. In this study we consider three different nonparametric Bayesian models for node clustering in complex networks. In particular, we test their ability to predict unseen data and their ability to reproduce clustering across datasets. The three generative models considered are the Infinite Relational

*Corresponding author
Kasper Winther Andersen, kasperwj@drcmr.dk, Danish Research Centre for Magnetic Resonance, Department 714, Centre for Functional and Diagnostic Imaging and Research, Copenhagen University Hospital Hvidovre, Kettegaard Alle 30, DK-2650 Hvidovre, Denmark, Tel: (+45) 3862 2852

Email addresses: kasperwj@drcmr.dk (Kasper Winther Andersen), stoffer@drcmr.dk (Kristoffer H. Madsen), hartwig.siebner@drcmr.dk (Hartwig Roman Siebner), mns@dtu.dk (Mikkel N. Schmidt), mm@dtu.dk (Morten Mørup), lkai@dtu.dk (Lars Kai Hansen)

Model (IRM), Bayesian Community Detection (BCD), and the Infinite Diagonal Model (IDM). The models define probabilities of generating links within and between clusters and the difference between the models lie in the restrictions they impose upon the between-cluster link probabilities. IRM is the most flexible model with no restrictions on the probabilities of links between clusters. BCD restricts the between-cluster link probabilities to be strictly lower than within-cluster link probabilities to conform to the community structure typically seen in social networks. IDM only models a single between-cluster link probability, which can be interpreted as a background noise probability. These probabilistic models are compared against three other approaches for node clustering, namely Infomap, Louvain modularity, and hierarchical clustering. Using 3 different datasets comprising healthy volunteers' rs-fMRI we found that the BCD model was in general the most predictive and reproducible model. This suggests that rs-fMRI data exhibits community structure and furthermore points to the significance of modeling heterogeneous between-cluster link probabilities.
Keywords: complex network, graph theory, infinite relational model, Bayesian community detection, resting state fMRI

Highlights

1. Three nonparametric Bayesian models for node clustering are used to model rs-fMRI.
2. Models' predictability and reproducibility are extensively evaluated using resampling.
3. The community structure model shows better predictability and repro-

ducibility.

4. This finding suggests that rs-fMRI graphs exhibit community structure.

5. Modeling between-cluster link probabilities adds important information.

1. Introduction

Analysis of resting state functional magnetic resonance imaging (rs-fMRI) has emerged as a powerful research tool to study whole-brain functional connectivity. Since rs-fMRI provides information about intrinsic fluctuations in functional connectivity within and among brain networks, many conventional analysis schemes applied in task-related fMRI studies are irrelevant. Hence, a number of new techniques have been developed based on identification of stable spatio-temporal multivariate structure in the brain wide set of blood oxygen level dependent (BOLD) time series.

Using correlation methods or flexible multivariate techniques like independent component analysis (ICA) it has been shown that the BOLD signals of distant brain regions are coordinated suggesting interaction as they form so-called resting-state networks. The number and precise definition of these networks are controversial but several networks are broadly accepted, including the default mode network, motor network, visual network, fronto-parietal, dorsal attention network (Damoiseaux et al., 2006). In addition to signals reflecting neuronal activity, the BOLD signal may be contaminated by physiological noise stemming from respiratory and cardiac cycles and head motion (Birn et al., 2006; Power et al., 2014).

Complex network analysis is a very active research field (Barabási, 2003)

that has already found application in neuroimaging and in modeling resting state connectivity (Bullmore and Bassett, 2011; Sporns, 2011). The basic object is the 'network graph'. When applied to neuroimage analysis the network graph is formed by brain regions represented as nodes. Nodes are connected by a link if brain regions are co-activated above a certain threshold. In rs-fMRI co-activation is often measured simply by calculating correlation between time series.

Network structure can be studied at many levels, from local motifs to global features like scale free link distributions signifying long-range coordination (van den Heuvel et al., 2008). Likewise, dense connections between high degree nodes is referred to as 'rich club organization' (van den Heuvel and Sporns, 2011). At the intermediate level we may identify clusters of highly linked nodes, i.e., high within-cluster link density and low link density to nodes in other clusters. By analogy to social networks such groups are referred to as *communities*. The presence of community structure in a network can be quantified by the global *modularity* index (Newman, 2006). Modularity can also be used to identify communities, i.e., by clustering nodes such that the modularity index is maximized (Newman, 2006; Lehmann and Hansen, 2007). Bassett et al. (2011) showed that 'flexibility', a measure for the number of cluster-assignment changes for nodes in a modularity optimized node-partition across time, is predictive for the amount of learning in a motor task in a subsequent session. Stevens et al. (2012) showed that modularity predicts visual working memory capacity, and Meunier et al. (2009) found that modularity is reduced during normal aging. Likewise, evidence is emerging that global modularity can be used as a bio-marker. For in-

stance patients with childhood-onset schizophrenia have reduced modularity of their resting state networks (Alexander-Bloch et al., 2010). However, focusing on modularity as the single summary of a complex network may be overly simplistic as the modularity measure does not account for variability in the inter-linking relations between functional clusters. Hence, modularity driven clustering might not reveal all salient aspects of community structure in a network. Indeed, modularity has been criticized for its lack of flexibility as a measure of community structure (Fortunato and Barthélemy, 2007).

A better understanding of this important mid-level structure in brain networks requires methods that can capture more informative representations of community structure. For this we turn to a family of expressive generative network models. Relational Models are statistical generalizations of graph clustering that consider not only the within-cluster density but also take the specific relations between clusters into consideration. The Infinite Relational Model (IRM) (Kemp et al., 2006; Xu et al., 2006) is a non-parametric generalization of the stochastic block model (Nowicki and Snijders, 2001), for inference of such generalized group structure in complex networks. As the IRM representation considers both linking within and between groups, a highly inter-linked group of nodes could in fact be clustered in different groups if they link in different ways to other clusters, i.e., the IRM can infer more general group structures beyond the conventional community structure. An additional feature of the IRM type of model is that it conveniently allows for analysis of multi-graph networks, which for neuroimaging data could be graphs from multiple sessions or subjects. For multi subject analysis one could look for a common node clustering structure over subjects but allow

individual subject cluster linking densities (Mørup et al., 2010) or test the hypothesis that both clustering and link structure are shared between all subjects (Andersen et al., 2012b).

A constrained variant of the IRM representing the community structure of graphs in the sense of grouping highly connected node sets was proposed recently by Mørup and Schmidt (2012). The Bayesian Community Detection (BCD) scheme restricts the between-cluster link densities to be strictly lower than within-cluster link densities, thus constraining the more general IRM to conform with the notion of a community in a social network. Another constraint is introduced by the so-called Infinite Diagonal Model (IDM) (Mørup and Schmidt, 2012; Schmidt and Mørup, 2013). The IDM allows for differential within-cluster link densities but models only a single between-cluster density and as such the variability in the link densities between clusters is neglected when inferring the clustering structure. Since the between-cluster link density is shared across clusters, it can be thought of as a background-noise density.

It should be noted that certain metrical properties can be expected when basing the graph on simple time series correlation, thereby assuming stationarity. If a node A is highly correlated with node B, and B is highly correlated with C, then there is a lower limit on the correlation between nodes A and C which can be inferred by the triangle inequality (Zalesky et al., 2012). This bound will support the formation of community structure, as in social relations: 'Friends of friends are friends', however, we also note that by thresholding the correlation, the impact on the community structure of these geometrical constraints is non-trivial.

Spatial grouping of brain regions by similarity of BOLD time series as pursued in the present work can be seen as complementary to classical approaches to spatial grouping such as time series clustering (Goutte et al., 1999) and independent component analysis (ICA) (McKeown et al., 1998, 2003). Compared with conventional clustering, the relational modeling approach has the advantage that clusters are formed by considering the connectivity patterns both within and between clusters, and furthermore relational models avoid the formation of a group prototype, hence allow for more flexible group structures to be found (Kemp et al., 2006). The use of ICA is based on assumptions of independence either in spatial or temporal dimensions, which can be questioned in the resting state as it has been observed that components are negatively correlated in time and have extensive overlaps in space (Fox et al., 2005).

In this study, we apply the above-mentioned community detection schemes to rs-fMRI data acquired in three cohorts of healthy volunteers and investigate to which degree functional brain networks as measured by rs-fMRI exhibit community structure. The three Bayesian relational methods, i.e. IRM, BCD, and IDM, for inference of group structure in complex networks differ only in the way the link probabilities between clusters are modeled. The rich link structures of the relational models can be seen as a way of inferring functional integration at the inter-community level as discussed in (Hagmann et al., 2008; Sporns, 2013). We evaluate the performance of these models with respect to their ability to predict out-of-sample data (predictability) and the robustness of their clustering under re-sampling of data (reproducibility) using the NPAIRS split-half framework (Strother et al., 2002). The evaluation

is carried out on three datasets from different sites and the models are evaluated both within and between sites for several thresholds of the correlation matrices. In addition, we compare the three models with three other methods for grouping nodes into clusters, namely Infomap, Louvain modularity, and hierarchical clustering. The work in this paper builds on work presented in (Andersen et al., 2012b).

2. Methods

For generality we investigate three rs-fMRI datasets. One dataset acquired locally at the Danish Research Centre for Magnetic Resonance (Copenhagen) and two other rs-fMRI datasets publicly available in the FCON1000 database (Biswal et al., 2010) (viz., the 'Beijing' and the 'Leipzig' datasets).

2.1. Copenhagen data

The Copenhagen dataset included 30 healthy controls with no history of neurological or psychiatric disease. At the day of scanning all subjects were asked to refrain from caffeine, cigarettes or alcohol intake at least six hours prior to the scanning session. All subjects gave written informed consent prior to scanning and the study was approved by the local scientific ethics committee of Copenhagen and Frederiksberg Communities (protocol no. KF01 - 131/03 with addendum). The Edinburgh handedness inventory (Oldfield, 1971) revealed that all participants except two were right handed.

All MRI measurements were performed on a 3.0 Tesla Magnetom Trio scanner (Siemens, Erlangen, Germany). Each participant underwent an MRI session including a structural scan as well as a functional scan during rest both with full brain coverage. During the functional scan subjects were

instructed to rest with their eyes closed without falling asleep, and refrain from any voluntary motor or cognitive activity.

The first scan during each session was the rs-fMRI functional scan which consisted of a T2* weighted echo planar imaging (EPI) sequence with a repetition time of 2490 ms, echo time 30 ms and flip angle 90 degrees. Over a period of 20 minutes we collected 480 brain volumes each consisting of 42 axial slices with an isotropic resolution of 3 mm, field of view (FOV): 192x192 mm. During scanning we monitored the subjects cardiac and respiratory cycles using an infrared pulse oximeter and a pneumatic thoracic belt. The structural scan was based on a magnetization prepared rapid gradient echo (MPRAGE) sequence with the following parameters: Repetition time (TR) = 1550 ms, echo time (TE) = 3.04 ms, inversion time (IT) = 800 ms; 192 sagittal slices; 1 mm isotropic resolution; FOV = 256 mm; flip-angle = 9 degrees.

The functional images were preprocessed using statistical parametric mapping software (SPM8, Wellcome Trust Centre for Neuroimaging, `http://www.fil.ion.ucl.ac.uk/spm`) implemented in Matlab 7.9 (MathWorks, Massachusetts, USA). In order to allow stabilization of T1 equilibrium effects we discarded the first five volumes prior to analysis. The remaining 475 brain volumes were realigned to the time-series mean using a two-step procedure and then co-registered to the same-session T1-weighted MPRAGE scan by a 6-parameter rigid-body transformation. The T1-weighted scan was spatially normalized to the Montreal Neurological Institute (MNI) 305 standard template using the unified segmentation/normalisation procedure as implemented in SPM8 (Ashburner and Friston, 2005). Subsequently the same

normalisation parameters were used to normalise the EPI images.

Both hardware instability and unwanted physiological effects (such as cardiac pulsation and respiration) produce signal changes in fMRI time-series (Smith et al., 1999; Dagli et al., 1999; Glover et al., 2000; Lund et al., 2006). These signal changes may give rise to signal fluctuation resembling those typically observed in rs-fMRI data (Birn et al., 2006). In order to reduce these effects prior to extraction of time series for the regions of interest we applied comprehensive temporal filtering of cardiac, respiratory and motion related effects. The filter included a high-pass filter based on discrete cosine basis functions (cut-off frequency 1/128 Hz). Cardiac and respiratory cycles were modeled using Fourier expansions of the aliased cardiac (10 parameters) and respiratory (6 parameters) cycles as well as first order cardiac by respiration cycles interaction (4 parameters) effects (Glover et al., 2000). Residual motion effects (24 parameters) were modeled using a Taylor expansion of the estimated movement parameters including spin-history effects (Friston et al., 1996). Changes in the respiration volume over time has been demonstrated to produce signal changes resembling those observed in rs-fMRI (Birn et al., 2006). We model these changes by included 41 time delayed versions (time delay between 20 and -20 seconds in one second intervals) of the respiration volume. Finally the filter included individual time series from cerebrospinal fluid voxels and white matter voxels from both the right and left hemispheres. In total the linear filter included 108 regressors leaving 367 degrees of freedom for the data.

2.2. Beijing and Leipzig data

Two other datasets were used from the FCON1000 database (Biswal et al., 2010) (`http://fcon_1000.projects.nitrc.org`). See Appendix A for a list of subjects used. The Beijing dataset consists of 42 of the subjects from the Beijing Zang set. The dataset is recorded with 33 slices using TR=2000ms and with 225 brain volumes. The Leipzig dataset consists of 37 subjects (21 females), ages 20-42, TR=2300ms, 34 axial slices, and 195 brain volumes. For both datasets the first 5 volumes had already been discarded. Preprocessing was done in SPM8 and included realigning to time-series mean for motion correction and normalising to standard MNI space using the template EPI image included in SPM.

2.3. Graph construction

We extracted the mean signal in each of the 116 regions covered in the AAL database (Tzourio-Mazoyer et al., 2002) and constructed the correlation matrix for each subject. Since this matrix is symmetric only the upper diagonal is further considered. Each subject's correlation matrix was binarized at an individual level to maintain the highest d-percent correlations corresponding to having a graph link density at d-percent. After thresholding an adjacency matrix $\mathbf{A}$ is retrieved where $A_{i,j}$ is the (i,j)-th element of $\mathbf{A}$ and $A_{i,j} = 1$ if there is a link between nodes i and j and $A_{i,j} = 0$ otherwise. Since we model multiple subjects, $\mathbf{A}^{(n)}$ denotes the adjacency matrix corresponding to subject n.

2.4. The models

This section will provide an overview of the models considered in this paper. For a more in depth description please refer to (Schmidt and Mørup, 2013). The goal is to group nodes into non-overlapping clusters, such that a common node-clustering across subjects is retrieved. Let $\mathbf{z}$ be the vector of nodes assignments where z_i takes the integer value corresponding to the cluster index node i belongs to. The models used are all generative models meaning that given the model definition and the model parameters one can generate new graphs by drawing samples from the model. The models differ in the way they model the link probability between and within clusters. Let $\rho_{k,l}$ represent the link probability between clusters k and l. Since we here consider undirected graphs $\boldsymbol{\rho}$ is symmetric.

2.4.1. The Infinite Relational Model

In IRM link probabilities within and between clusters are modeled individually and without restrictions. As such the model allows for complex relations between clusters, and thus allows for flexible clustering of nodes.

Consider generating graphs from this model. The first step is to draw assignments of nodes into clusters, which is done using the Chinese Restaurant Process (CRP) (Aldous, 1985) using the hyper-parameter α. The CRP generates a cluster assignment, where α controls the number of clusters generated, where larger α will generate more clusters. Next, the link probabilities within and between clusters $\rho_{k,l}$ are generated from the symmetric Beta distribution using the hyper-parameter β. Finally, the cluster assignments and the link densities are used to generate links between nodes. This is done using the Bernoulli distribution, where the probability of a link (success) between

nodes i and j is determined by the clusters (z_i and z_j) the nodes belong to.

The generative model can be summarized as:

Infinite Relational Model	
Cluster assignments:	$\mathbf{z} \sim \text{CRP}(\alpha)$
Link probabilities:	$\rho_{k,l} \sim \text{Beta}(\beta, \beta)$
Links:	$A_{i,j}^{(n)} \sim \text{Bernoulli}(\rho_{z_i,z_j})$

In Appendix B.1 we derive the likelihood function for the IRM which is used in model inference.

2.4.2. Infinite Diagonal Model

The model termed Infinite Diagonal Model (IDM) (Mørup and Schmidt, 2012) is a special case of the IRM where link probabilities between clusters are constrained to be equal. As such, the IDM does not model the relation between clusters but has a constant background link probability. The only difference in terms of the model formulation is then

$$\rho_{k,l} = \begin{cases} \rho_{k,k} & \text{if } k = l \\ \rho_b & \text{otherwise.} \end{cases}$$

2.4.3. Bayesian Community Detection

A network community is defined as a group of nodes with more dense linking internally than externally. The Bayesian Community Detection (BCD) model proposed in (Mørup and Schmidt, 2012) enforces larger within-cluster link probabilities than between-cluster link probabilities. Like IRM, the cluster assignments are first generated using the CRP. A cluster-gap is then

drawn from a symmetric Beta distribution with hyperparameter v. The within-cluster link probabilities are then drawn for each cluster again using the Beta distribution. The between-cluster link probabilities are subsequently drawn using the incomplete Beta distribution $\text{BetaInc}(a, b, x)$ constrained to the interval $[0, x]$, with the density, $p(\theta) = \frac{1}{B_x(a,b)}\theta^{a-1}(1-\theta)^{b-1}$, where $B_x(a, b)$ is the incomplete beta function. Thus, a draw from the incomplete Beta distribution will return a value between $[0, x]$, which can then be used to control the maximal value the between-cluster link probability can take. By setting x to the cluster-gap times the within-cluster link probability, the between-cluster link probability between two clusters k and l can then at most be as high as the smaller of the two within-cluster link probabilities multiplied by the cluster gap. The lower the gap-value the higher difference in within and between-cluster link probability. Finally, links are drawn using the Bernoulli distribution just like the other models. The generative model for BCD can thus be summarized as:

Bayesian Community Detection

$$
\begin{aligned}
\text{Cluster assignments}: &\quad \mathbf{z} \sim \text{CRP}(\alpha) \\
\text{Cluster gap}: &\quad \gamma \sim \text{Beta}(v, v) \\
\text{Link probability}: &\quad \rho_{k,l} \sim \begin{cases} \text{Beta}(\beta, \beta) & \text{if } k = l \\ \text{BetaInc}(\beta, \beta, w_{k,l}) & \text{otherwise.} \end{cases} \\
&\quad \text{where } w_{k,l} = \min[\gamma\rho_{ll}, \gamma\rho_{kk}] \\
\text{Links}: &\quad A_{i,j}^{(n)} \sim \text{Bernoulli}(\rho_{z_i,z_j})
\end{aligned}
$$

2.5. Example 1

For illustration we generate a graph consisting of 50 nodes from each of the three models with $\alpha = 5, \beta = 1$. For the BCD model we set $v = 1$. Figure 1 shows the generated graphs. The plots are a combination of both the cluster assignment matrix, the adjacency matrix, and the link probability matrix. The adjacency matrix $\mathbf{A}$ is plotted, where links between nodes are indicated by small black dots. Cluster membership is indicated with the colors to the left and top of the adjacency matrix and the link probability matrix is indicated with gray shaded background. For IRM there are no restrictions in the link probability values, resulting in some between-cluster link probabilities being larger than within-cluster link probabilities. For the BCD model the between-cluster link probability between two clusters are restricted to be smaller than the within-cluster link probability times the gap. The gap was drawn from the Beta distribution and in this case the gap is $\gamma = 0.96$. For the IDM model all the between-cluster link probabilities are equal meaning that clusters are only defined in the way they link internally in the clusters.

2.6. IRM and IDM model inference

In the previous sections we defined the generative models, which allow one to generate data by sampling from the model. However, we are interested in inferring the model parameters given the data. By using the model definition the joint likelihood can be written and by using Bayes theorem an expression for the posterior distribution can be found. It is then possible to sample from this posterior distribution using Markov chain Monte Carlo sampling (MCMC) methods. For IRM and IDM the link probabilities can

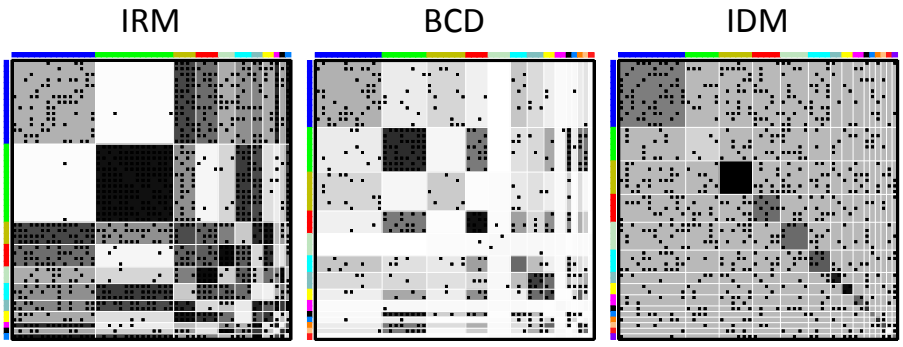

Figure 1: Example 1. Figure illustrating data drawn from each of the three models, IRM, BCD, and IDM respectively. Plots illustrates both the adjacency matrix $\mathbf{A}$ (links indicated by small black squared dots), cluster membership $\mathbf{z}$ as color codes to the left and top of the adjacency matrix, and link probability matrix $\boldsymbol{\rho}$ as gray shading of the matrix elements.

analytically be integrated out which means that we only have to sample over the node assignments. For that Gibbs sampling in combination with split-merge Metropolis-Hastings updates (Jain and Neal, 2004; Kemp et al., 2006; Mørup et al., 2010) is used. Below is a description of these two steps.

Gibbs sampling is a Markov-chain Monte Carlo sampling method. For each scan of the Gibbs sampler each node's cluster assignment is updated using the conditional distribution of that node's assignment given the assignments of the remaining nodes. For IRM the conditional distribution is derived in Appendix B.1 (equation B.2) and for IDM it is derived in Appendix B.2 (equation B.3).

Given the incremental nature of the Gibbs sampling algorithm it has difficulties escaping local maxima of the probability landscape. For instance it is hard to split a single cluster into two new clusters since this requires that nodes are moved one at a time from a cluster to the other cluster. To

overcome this we use a restricted split-merge sampling scheme (Jain and Neal, 2004), which potentially merges two existing clusters into one or split a single cluster into two clusters. At each step of the algorithm two nodes are selected at random with uniform probability. If the two selected nodes currently are assigned different clusters then an assignment is proposed where these two clusters are merged into one cluster. On the contrary, if these two selected nodes are currently assigned to the same cluster then a new assignment is proposed where all nodes assigned to this cluster are split into two separate clusters. The split-proposal is found using a restricted Gibbs-sampling procedure. First a launch state is found by allocating the two nodes to two different empty clusters as proposed in (Dahl, 2005). Then remaining nodes are in random order assigned to either of the two clusters based upon their conditional probability. This state is then referred to as the launch state. The launch state is refined by restricted Gibbs sampling steps where nodes from the two new clusters can be re-assigned either of the two clusters based on the conditional probability (equation B.2 and B.3). This procedure is restricted because only nodes from the cluster from which the nodes originally came from are re-assigned and they can only be assigned to either of the two new clusters. The proposed configuration is then sampled from the launch state. If this proposed state $\mathbf{z}^*$ in the Markov chain is accepted with the Metropolis-Hasting acceptance probability $a(\mathbf{z}^*, \mathbf{z})$ then this becomes the new state else the old state $\mathbf{z}$ is kept as the new state. The acceptance probability is given as $a(\mathbf{z}^*, \mathbf{z}) = \min\left[1, \frac{q(\mathbf{z}|\mathbf{z}^*)}{q(\mathbf{z}^*|\mathbf{z})}\frac{\pi(\mathbf{z}^*)}{\pi(\mathbf{z})}\right]$, where $\pi(\mathbf{z}) = P(\mathbf{A}|\mathbf{z})P(\mathbf{z})$ (please see Appendix B) and $q(\mathbf{z}^*|\mathbf{z})$ is the probability of transition from $\mathbf{z}$ to $\mathbf{z}^*$. For further detail about the split-merge sampling

please refer to Jain and Neal (2004).

2.7. BCD model inference

In IRM and IDM we are able to marginalize link probabilities ($\boldsymbol{\rho}$) out. This is not the case in BCD because between-cluster link probabilities are dependent of the within-cluster link probabilities. However, the vast majority of the parameters, namely the between-cluster link probabilities, can be integrated out (Appendix B.3). The remaining parameters $\mathbf{z}, \dot{\boldsymbol{\rho}}$, and γ are sampled using MCMC, where $\dot{\boldsymbol{\rho}}$ refer to the within-cluster link probabilities (the diagonal of $\boldsymbol{\rho}$). The within-cluster link probabilities and cluster gaps are sampled with Metropolis-Hastings. The cluster assignments $\mathbf{z}$ are like the IRM sampled with Gibbs sampling and split-merge moves, however new possible values for the within link probabilities and cluster gaps are first drawn from their prior. In Appendix B.3 we derive the conditional distributions used in the sampling. For further information please see (Mørup and Schmidt, 2012).

2.8. Example 2

We illustrate differences in cluster assignments and link probabilities inferred by each of the three models. We generate a synthetic graph with 40 nodes, 10 nodes in each of four clusters. The example is designed such that *cluster1* and *cluster2* share the same within and between-cluster link probabilities, however only *cluster2* is connected with *cluster3*. *Cluster3* and *cluster4* have low within-cluster probabilities but high between-cluster link probability. *Cluster3* and *cluster4* are not connected to *cluster1* and *cluster2*. The first row in Figure 2 show the true assignment vector ($\mathbf{z}$) coded as

a 1-of-n matrix and the true link probabilities. The next rows show the assignments and link probabilities inferred by the IRM, BCD, and IDM models respectively. Except for a single node IRM finds the correct grouping structure. BCD assigns the first two clusters correctly and mislabels the same node as IRM, but BCD has difficulties with the remaining nodes because the true model has higher between-cluster than within-cluster link probabilities. Since IDM does not model the between-cluster link probabilities, it groups the first two clusters together and the next two clusters together.

2.9. NPAIRS Evaluation Criteria

To evaluate the performance of the models, we used the NPAIRS split-half evaluation framework (Strother et al., 2002). Under this framework the set of subjects were split into two half-splits (S1 and S2) and models were inferred on each half-split enabling us to estimate the predictability and reproducibility of the models. The models' predictability was evaluated using test log likelihood. The node assignment and link probabilities from the sample with the highest value of the posterior distribution were used to calculate the test log likelihood of the other split. The test log likelihood was calculated for both splits (with the other split as training data) and the average test log likelihood was calculated and used as the predictability measure. The test log likelihood for split S2 (using the model parameters inferred using split S1) was calculated by

$$\log P(\mathbf{A}^{\mathrm{S2},(1)}, ..., \mathbf{A}^{\mathrm{S2},(N)} | \boldsymbol{\rho}, \mathbf{z}) =$$
$$\frac{1}{N} \sum_{n=1}^{N} \sum_{j>i} \Big[A_{i,j}^{\mathrm{S2},(n)} \log(\rho_{z_i,z_j}) + (1 - A_{i,j}^{\mathrm{S2},(n)}) \log(1 - \rho_{z_i,z_j}) \Big]. \qquad (1)$$

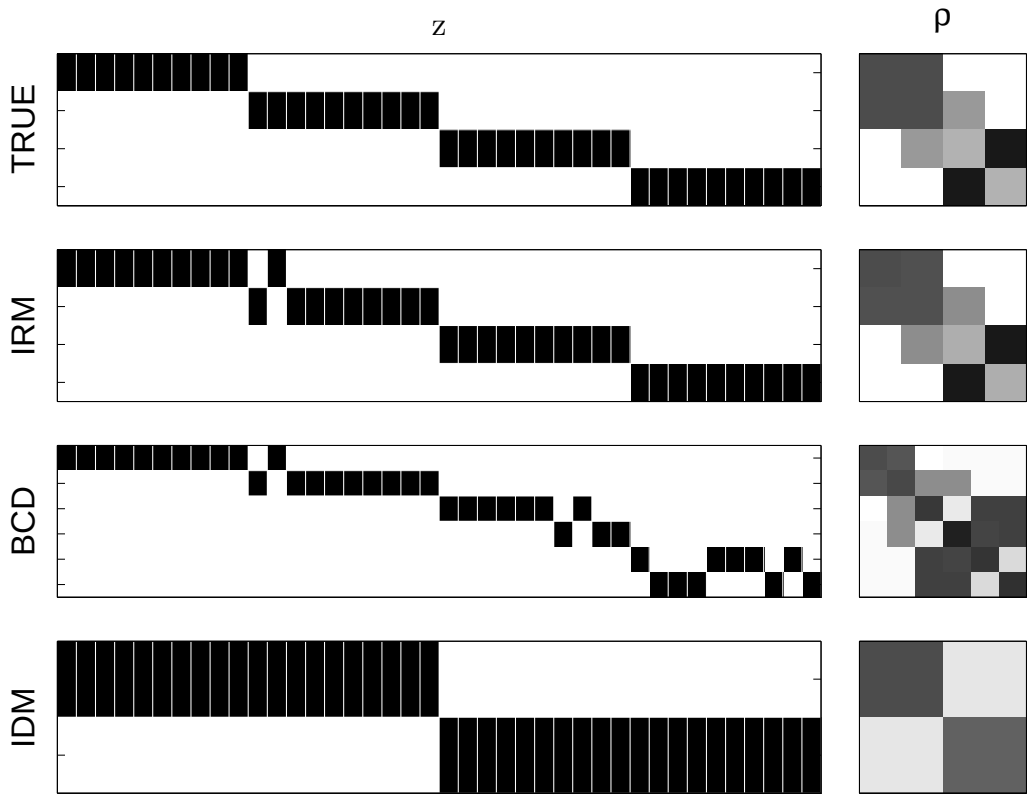

Figure 2: Example 2. First row show the true assignments coded as a 1-of-n matrix and the true link probabilities. The next rows show the structure and link probabilities inferred by IRM, BCD, and IDM respectively.

We measured the reproducibility of the models using normalized mutual information between assignment matrices ($\mathbf{z}^{S1}$ and $\mathbf{z}^{S2}$) of the sample with the highest value of the posterior distribution inferred using the two different splits.

$$\mathrm{NMI} = \frac{2\mathrm{MI}(\mathbf{z}^{S1}, \mathbf{z}^{S2})}{\mathrm{MI}(\mathbf{z}^{S1}, \mathbf{z}^{S1}) + \mathrm{MI}(\mathbf{z}^{S2}, \mathbf{z}^{S2})}, \tag{2}$$

where

$$\mathrm{MI}(\mathbf{z}^{\mathrm{S1}}, \mathbf{z}^{\mathrm{S2}}) = \sum_{k=1}^{D1} \sum_{l=1}^{D2} p(z^{\mathrm{S1}} = k, z^{\mathrm{S2}} = l) \log \left(\frac{p(z^{\mathrm{S1}} = k, z^{\mathrm{S2}} = l)}{p(z^{\mathrm{S1}} = k) p(z^{\mathrm{S2}} = l)} \right), \quad (3)$$

where $D1$ and $D2$ are the number of clusters inferred using S1 and S2, respectively.

The model used in e.g. Mørup et al. (2010) used individual subject link probabilities, that is, each subject was modeled with her own link probability matrix while sharing the node assignments $\mathbf{z}$. This allows for subject variability in the communication between clusters and can be used to test for differences in subject populations. However, here we are interested in the models' predictive abilities, that is, how well can a model and its parameters learned from a sub-group of subjects predict the graphs from another group of subjects. Therefore we do not model individual subject link densities but constrain $\boldsymbol{\rho}$ to be common across subjects. The derivation of the models (Appendix B.1) reveals that this amounts to simply summing the adjacency matrices across subjects $\sum_n \mathbf{A}^{(n)} = \mathbf{A}^{\mathrm{tot}}$. This means that under this restricted model definition inference of the latent variables of the model does not scale with the number of graphs (subjects) and therefore our model formulation allows for analysis of large numbers of subjects.

2.10. Experiments

The initial assignment of nodes to clusters might affect the final clustering, so we did some initial experiments with different number (1, 20, 50, or 116) of initial clusters (data not shown). For IRM and BCD, the similarity between different choices were in general very high (mean NMI > 0.95) with a tendency of generating more clusters when initializing all nodes in its own

cluster. For the IDM the initialization had a greater impact on the final clustering. Initialization in few (1 or 20) clusters showed greater variability in the final clustering whereas initialization in 50 or 116 clusters showed more stable final clustering. Thus to compromise, in all the experiments described below, nodes were randomly assigned to one of 50 clusters and the sampler ran for 500 iterations. The sample with the highest value of the posterior distribution was then used as representative for a given run. In all experiments $\alpha = \beta = 1$.

2.10.1. Estimated clusters

To inspect the clustering of the different models, the inference procedure was launched 10 times for each model using the graph link density $d = 8\%$ and the sample with the overall highest value of the posterior distribution across the 10 runs was visualized. The reproducibility measured as the mean NMI between the samples with the highest value of the posterior distribution for each run was calculated. As the inference is stochastic, this measures the methods' ability to reproduce clusterings for different restarts. Likewise, the clustering similarity between the methods was also estimated by calculating the mean NMI between each pair of the 10 solutions found.

In addition, we investigated the impact on different choices of graph link density. For each of the densities $d = 2\%, 4\%, 8\%, 16\%, 32\%$ we launched the inference 10 times for each model and estimated the mean NMI between densities within each model. The clustering with the highest value of the posterior distribution for each density was visualized for the BCD model. These experiments used all the subjects from the Copenhagen dataset.

2.10.2. Predictability and reproducibility

We asked how well the clusterings reproduce between datasets and how well the models predict new data. To this end, we evaluated the models using the NPAIRS framework. Subjects were randomly split into two equally sized groups and model inference was conducted on each split. The highest posterior distribution sample was identified for the two splits and NMI between clusterings was calculated as a measure of the models' reproducibility. Using the estimated link probability matrix and assignment from the sample with the highest value of the posterior distribution of one split, the test log likelihood for the other split was calculated as a measure of the models' predictability. This was done for 100 different half-splits of the Copenhagen dataset using 8% graph link density.

2.10.3. Predictability and reproducibility for various link densities

For further evaluation of the methods the analysis were repeated within each of the three datasets as well as between the datasets for graph link densities of $d = 2\%, 4\%, 8\%, 16\%, 32\%$. For analysis done within each individual dataset the subjects were randomly split in half. For the between dataset analysis, inference was done within each dataset and NMI and test log likelihood was calculated between datasets. For each link density the log likelihood ratio was calculated as the log likelihood of a random Erdős-Rényi model having the considered link density divided by the log likelihood of the inferred model. This makes the predictability measure more comparable between link densities, however, we note that the log likelihood cannot directly be compared for different link densities as the data itself changes when changing the link densities.

We compare the Baysian methods with two of the best community detection algorithms (Fortunato, 2010) as well as with a simple method based on hierarchical clustering. The first method is Infomap (Rosvall and Bergstrom, 2008) using the C++ implementation available from `http://www.mapequation.org/`. Infomap has previously been used for fMRI networks, see e.g. (Power et al., 2011). The second method is the so-called Louvain method (Blondel et al., 2008) as implemented in the Brain Connectivity Toolbox (`https://sites.google.com/site/bctnet/`) (Rubinov and Sporns, 2010). This method is based on modularity optimization. The third method is the agglomerative hierarchical clustering based on average linkage using the Matlab function 'linkage'. For this method we formed clusters by thresholding the hierarchical tree at the distance 0.9.

To obtain a single clustering across a group of subjects we ran the methods on the summed adjacency matrix across the subjects in each half-split. This summed adjacency matrix is also used for inference in the probabilistic models (as noted in section 2.9), which therefore allows for a comparison between methods. To compare the results from these three methods we treat the clustering found as it was produced by the IRM and thus calculate the link probabilities between clusters as it was done for the IRM model. This allows us to calculate the predictability for unseen data as described in Section 2.9. In addition to the predictability and reproducibility we also evaluate the modularity index for all methods. The modularity index is given as $Q = \mathrm{Tr}(\mathbf{Z}\mathbf{B}\mathbf{Z}^\top)/2m$ where $\mathbf{Z}$ is 1-of-D encoding matrix of the link assignment vector $\mathbf{z}$ and m is the number of links in the graph. $\mathbf{B} = \mathbf{A} - (\mathbf{k}\mathbf{k}^\top)/2m$ is the modularity matrix where $\mathbf{k}$ is the vector of node degrees.

3. Results

3.1. Estimated clusters

We thresholded the graphs to maintain the top 8% correlations. The threshold correspond to a mean (std) p-value across subjects of $4.75*10^{-5}$ ($1.80*10^{-4}$). The reproducibility between solutions found with different restarts was measured as the NMI between the sample with the highest value of the posterior distribution for each run. This was done within all three methods and between the methods and results are shown in table 1 along with the number of clusters estimated by each of the methods. For all three methods the clustering for different initializations showed a very high consistency as the NMI was greater than 0.96 for all methods. Also, the number of estimated clusters was very consistent within method, but showed a great between method variability where IRM estimated on average 35.7 clusters, BCD estimated 41.0 and IDM estimated only 18.8. For BCD the mean (std) gap parameter was estimated to 0.88 (0.02). The IRM and BCD clusterings were found to be very similar with a mean NMI of 0.94. The IDM clustering, however, was less similar to the other two methods with a mean NMI of 0.76 and 0.75 to IRM and BCD respectively.

In figure 3 the samples with the highest value of the posterior distribution across the 10 runs for each method are visualized. The first column shows the link probability matrix $\boldsymbol{\rho}$ which has been permuted such that clusters with the greatest overlap between methods are first. The labels for the clusters can be found in Appendix C. The matrix elements are color-coded in grey-scale according to the value of the link probabilities and the size of the matrix elements indicate the size of the clusters. The first 5 clusters were

Method	IRM	BCD	IDM	Mean (std) D
IRM	0.96 (0.01)	-	-	35.7 (1.25)
BCD	0.94 (0.01)	0.96 (0.02)	-	41.0 (2.05)
IDM	0.76 (0.02)	0.75 (0.01)	0.97 (0.02)	18.8 (1.14)

Table 1: The mean(std) of normalized mutual information (NMI) between the clustering of 10 runs within and between method along with the number of clusters (D) estimated with each of the three methods IRM, BCD, and IDM.

identical between the three methods. The next 12 clusters were identical between IRM and BCD while IDM had all these clusters in one large cluster. When looking at the link probabilities between these 12 clusters it is evident that there is a high link probability within and between these nodes, but subtle differences exist between the different clusters which caused the IRM and BCD to cluster them into separate clusters. Since IDM does not consider the between-cluster link probabilities these clusters were grouped together in the IDM method. The same was true for the next 6 clusters which were identical for the IRM and BCD and all lumped together in the IDM model since the link probabilities between these clusters were relatively high. The next three columns show the found clusters in posterior, lateral and superior views of the brain. The clusters are colored according to the colors shown next to the link probability matrices (and the labels given in Appendix C). Brain regions within clusters are connected with lines where line thickness indicates the average link density over subjects for the specific connection. This figure shows that the IRM and BCD clusterings were very similar. In general, these two methods produced clusters with relatively few nodes and grouped interhemispheric homologues areas together. IDM also grouped interhemispheric

homologues areas together, however, as just described this method does not consider specific relations to other brain areas, which resulted in larger and rather unspecific clusters. For instance the cluster colored in turquoise is a cluster made up of 34 nodes including nodes in frontal, occipital, parietal, and temporal lobes.

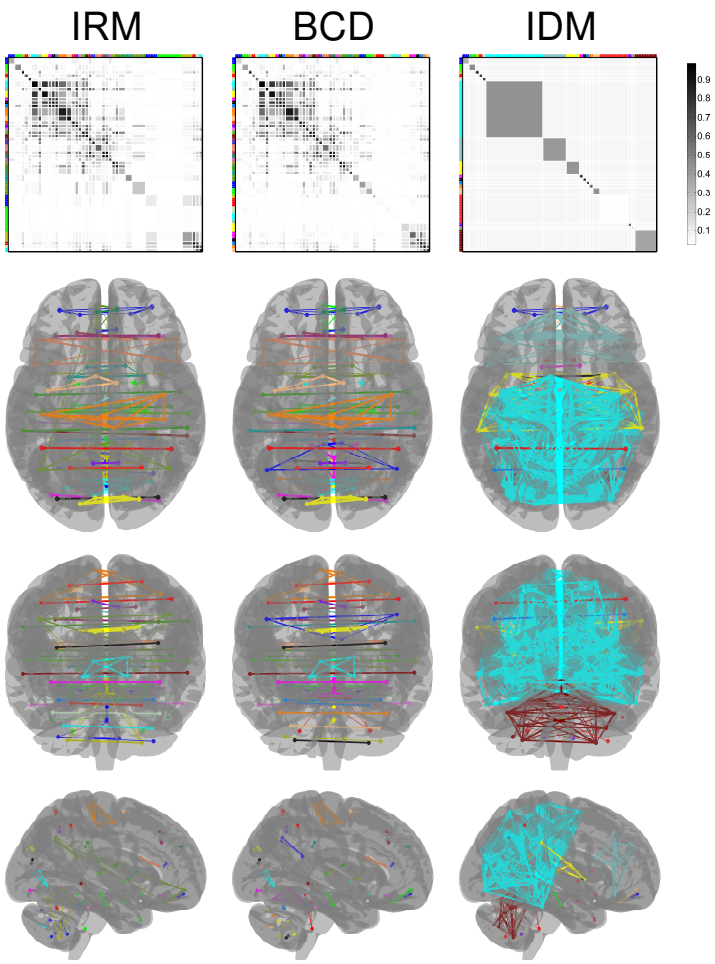

Figure 3: The extracted clusters using the three methods IRM, BCD, and IDM respectively. The first row shows the link probability matrices $\boldsymbol{\rho}$, which have been permuted such that the order of the clusters corresponds across methods. The matrix elements are color-coded according to the value of the link probabilities and the size of the matrix element indicates the size of the respective cluster. The colors next to the matrices correspond to different clusters. The next three rows show the clusters in three different views (superior, posterior, and lateral) of the brain. The clusters are color coded according to the colors next to the link probability matrices and node assignment for each node can be found in Appendix C with the same color as plotted here. Different brain regions within each cluster are connected with lines where the thickness of the line indicates the average link density across subjects for the specific connection.

In figure 4 we show an example cluster and its connectivity to other clusters. A cluster composed of left and right supplementary motor area and left precentral gyrus (A) was selected. This cluster was identical for IRM and BCD while results are not shown for IDM. The figure also displays the 4 clusters with highest between-cluster link probabilities to this cluster. These 4 clusters with highest link probabilities were: (B, $\rho_{A,B} = 0.732$) left and right postcentral gyrus, left and right paracentral lobule and right precentral gyrus; (C, $\rho_{A,C} = 0.714$) left and right middle cingulate gyrus; (D, $\rho_{A,D} = 0.516$) left and right superior frontal gyrus; (E, $\rho_{A,E} = 0.456$) left and right superior temporal gyrus. The line widths between clusters in the figure reflect between-cluster link probabilities, likewise the widths of the boxes reflect the within-cluster link probabilities.

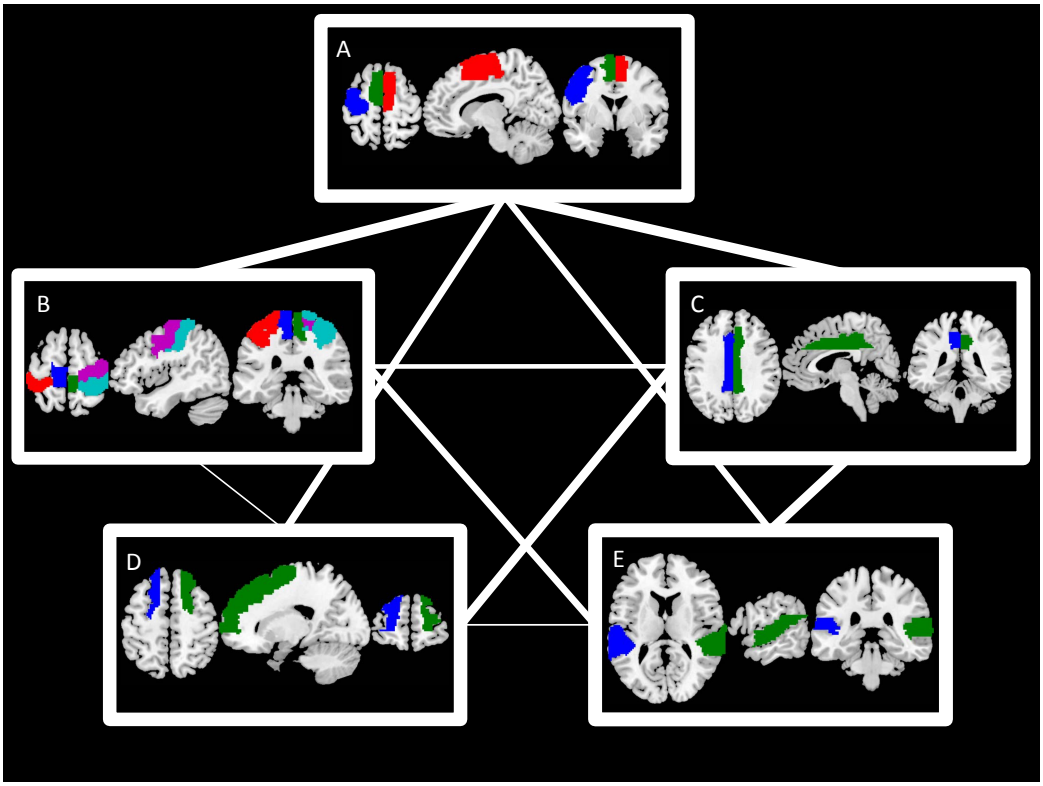

Figure 4: This figure shows a single cluster A composed of left and right supplementary motor area and left precentral gyrus. Also, the 4 clusters with highest between cluster link probability to this cluster are shown. These 4 clusters are (B, $\rho_{A,B} = 0.732$) left and right postcentral gyrus, left and right paracentral lobule and right precentral gyrus; (C, $\rho_{A,C} = 0.714$) left and right middle cingulate gyrus; (D, $\rho_{A,D} = 0.516$) left and right superior frontal gyrus; (E, $\rho_{A,E} = 0.456$) left and right superior temporal gyrus. The line widths between clusters reflect the link probabilities between clusters, likewise the widths of the boxes reflect the within-cluster link probabilities.

Figure 5 shows the clustering similarity both within the same link density and between different link densities for IRM, BCD, and IDM, respectively. IRM and BCD showed a similar pattern with increasing clustering similarity with increasing link density. When considering the link densities 8%, 16%,

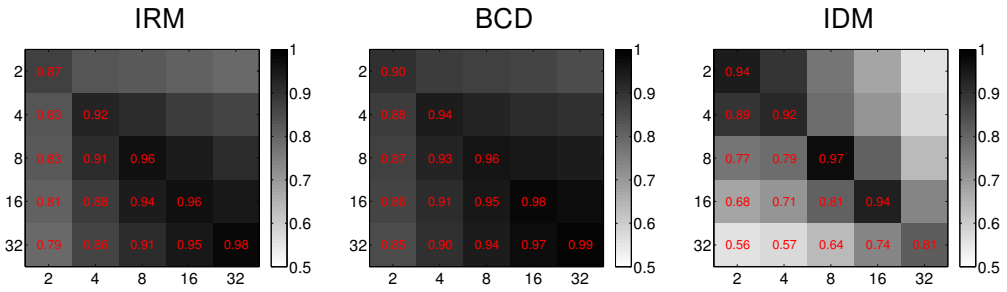

Figure 5: Mean NMI between clusterings found with different graph link densities for IRM, BCD, and IDM, respectively.

and 32% the mean NMI between the clusterings were all above 0.91, reflecting high similarity. Comparing the similarity between low and high link densities reveal lower similarity, for instance the mean NMI between 2% and 32% were 0.79 and 0.85 for IRM and BCD, respectively. BCD showed in general higher NMI values than IRM. For IDM the pattern was opposite, with decreasing similarity with increasing link density. Comparing clusterings between 2% and 32% for IDM revealed a relatively low NMI value of 0.56. The NMI values were in general lower for IDM than for IRM and BCD.

Figure 6 shows the solutions with the highest values of the posterior using BCD for different link densities. The tendency across different link densities is that clusters are very left-right symmetric.

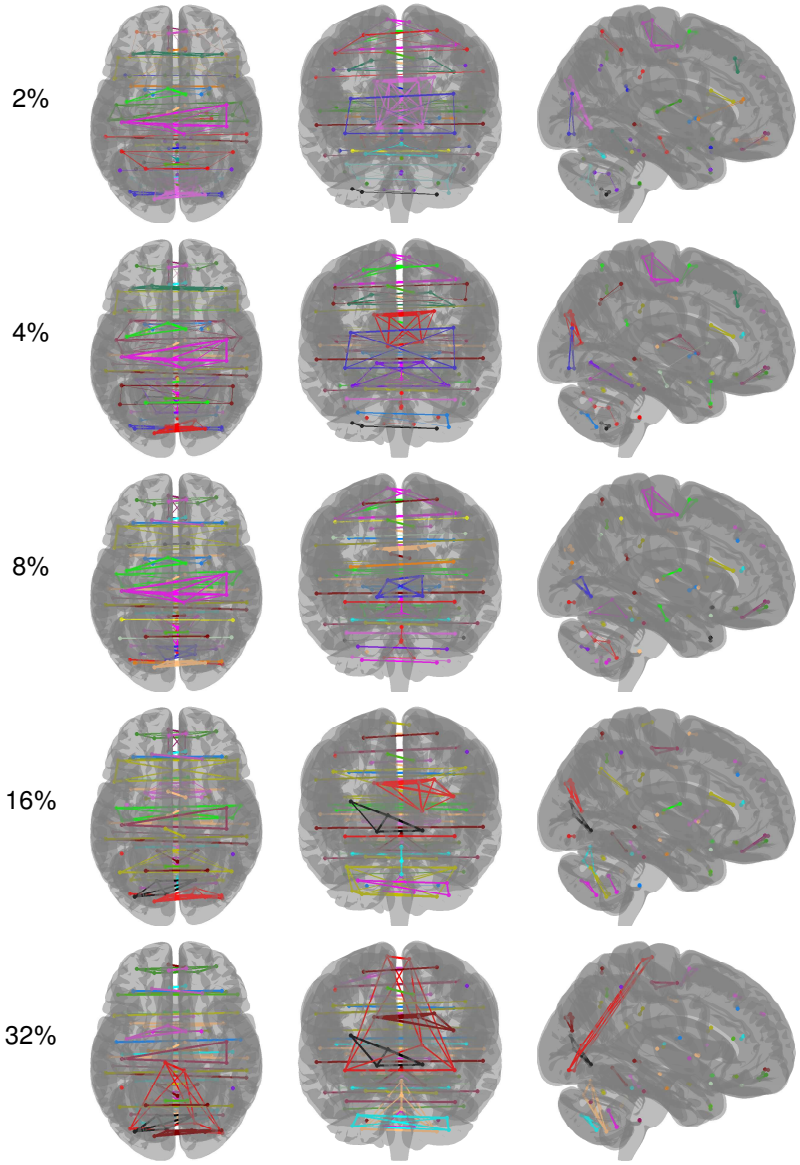

Figure 6: Clusterings found using BCD for different graph link densities.

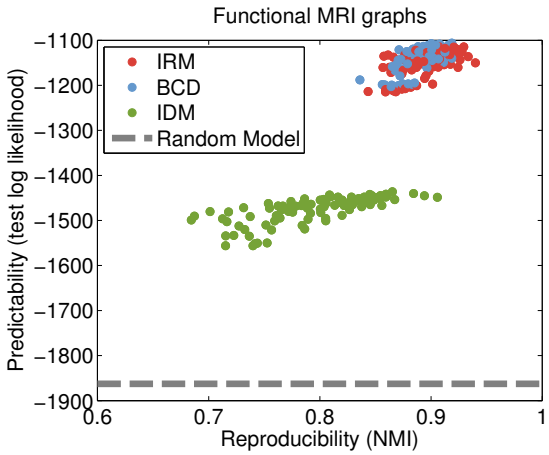

Figure 7: Reproducibility vs predictability plot for the three models using a link density of 8%. IDM and BCD showed better NMI and test log likelihood compared with IDM ($p < 0.0001$). BCD showed better predictability ($p = 0.023$) compared with IRM while IRM and BCD did not differ in reproducibility ($p = 0.15$).

3.2. Predictability and reproducibility

Figure 7 shows the PR scatter plot of the predictability versus reproducibility of the 3 methods using the NPAIRS split-half framework. Clearly, IRM and BCD performed better compared with IDM in both reproducibility and predictability as measured with NMI and test log likelihood ($p < 0.0001$, permutation test). IRM and BCD overlap, however, when testing for differences BCD showed slightly better predictability than IRM ($p = 0.023$) while the two methods did not differ in reproducibility ($p = 0.15$). On average IRM estimated 29.6 (std=0.83) clusters while BCD estimated 34.8 (0.88) and IDM estimated 17.7 (1.13). The number of clusters reported here was estimated on half-splits of the subject sample and are therefore different from the numbers reported in table 1 for models estimated on the whole sample.

3.3. Predictability and reproducibility for various link densities

Figure 8 shows the mean data and its standard error for the reproducibility, predictability, number of clusters, and modularity index within and between the three datasets when varying link density. This is shown for IRM, BCD, and IDM, as well as Infomap, Louvain modularity, and hierarchical clustering. The columns represent the different datasets (Copenhagen, Leipzig, Beijing, and between datasets respectively). Inspecting clustering reproducibility the general tendency was that BCD and IRM increased with increasing link densities while the other methods tended to decrease. For link densities greater than 4% IRM and BCD were superior to the other methods. BCD performed better or on par with IRM across all datasets and for all link densities. For high link densities Infomap produced only a single cluster causing both the nominator and denominator in the NMI calculation to be zero, thus these values are not shown.

BCD and IRM generally showed higher predictability compared with the other methods for all datasets and link densities. For the three within dataset analyses BCD performed better compared with IRM for low link densities, for higher link densities these two methods were on par. Interestingly, hierarchical clustering generally showed higher predictability and reproducibility than IDM, Louvain, and Infomap. Please note that the test log likelihood ratio cannot be compared directly between different link densities.

When inspecting the number of clusters estimated by the methods two patterns are observed. For IRM and BCD the number of clusters increased with increasing link density. The opposite is seen with the other methods where the number of clusters decreased with increasing link density.

Not surprisingly, the Louvain method had higher modularity values for all sets and link densities compared with the other methods. IRM and BCD had comparable modularity but was generally lower than the other methods. IDM had higher modularity compared with the two other probabilistic methods IRM and BCD.

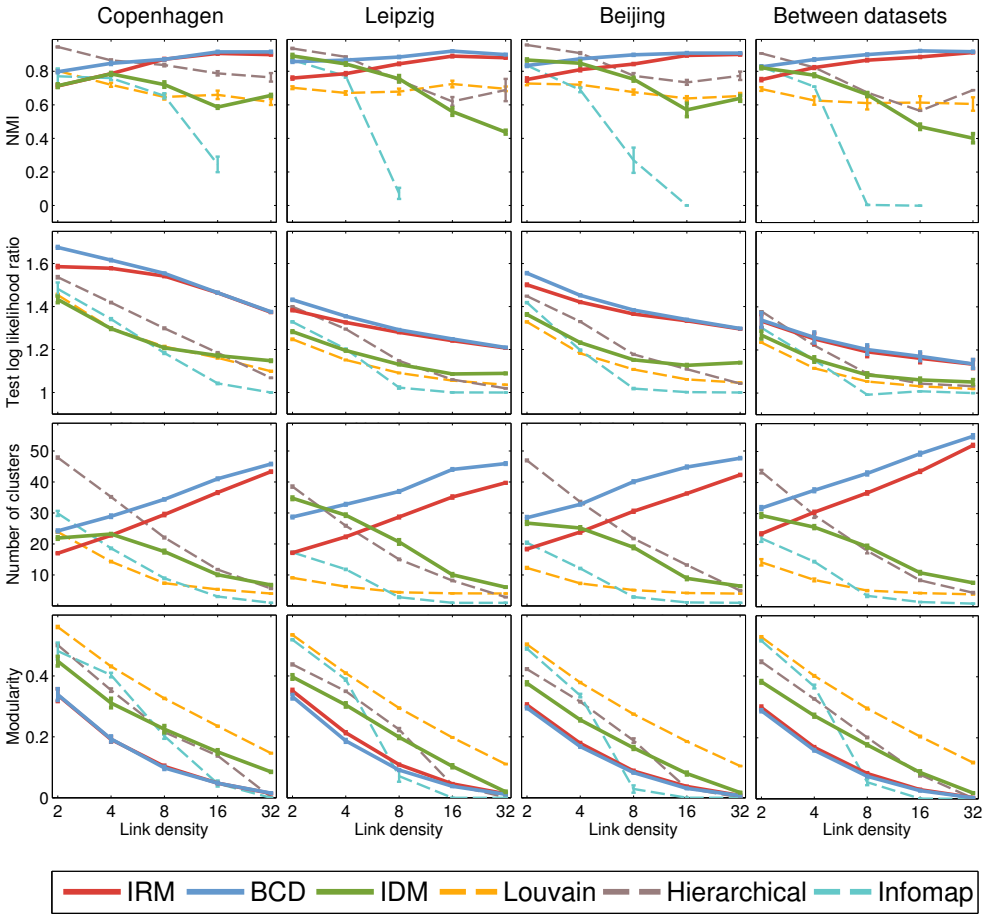

Figure 8: First row reproducibility (NMI), second row predictability (test log likelihood ratio), third row number of clusters, and fourth row modularity index as function of graph densities. Columns represent Copenhagen, Leipzig, Beijing, and between datasets respectively.

4. Discussion and conclusion

Our aim was to explore statistical models for finding structure in networks at the intermediate level. Accumulated evidence points to the importance of

community structure in brain networks, hence, we tested three statistical link models, which differed in terms of the different restrictions that were imposed on how nodes are clustered. The IRM is a very flexible representation for graph clustering, in which nodes can be grouped together without having a high link density among them. The BCD is a constrained version of the IRM that discards such group structures by insisting on higher within-cluster interaction, conforming with the notion of community structure. Finally, the IDM model is further constrained to ignore potential differences in the way nodes in a community interact with other communities, inspired by the methods aimed at identifying structure based on the global modularity concept. These probabilistic models were compared against three other non-probabilistic methods for finding community structure in networks; Infomap, Louvain, and hierarchical clustering.

The results show a remarkably difference between the two models (IRM and BCD), which models the between-cluster linking, and the other methods, which does not specifically take the between-cluster linking into account. IRM and BCD had generally higher reproducibility (except for low link densities) and predictability and showed an increasing number of estimated clusters with increasing link densities. On the contrary, these two models had lower modularity index as compared with the other methods. Modularity is often used for node clustering in brain networks, however, the results shown here indicate that a modularity optimized partition is neither the most reproducible nor the most predictable.

In general IRM and BCD clustered few nodes together corresponding to interhemispheric homologues areas. IRM and BCD model the between-

cluster link probabilities, which allows one to inspect how different clusters link to each other, an example of this is shown in figure 4. While a low number of nodes in a specific cluster might not reveal a lot of information in itself, important characteristics can be extracted and interpreted when considering the information available from the between-cluster link probabilities. In such a way, the between-cluster link probabilities can reveal how the different clusters are linked, indicating the communication pattern between clusters. In contrast to these two most expressive models, IDM does not model specific between-cluster link probabilities. This results in larger clusters with relatively high within-cluster link probabilities, which are formed since the model does not care about specific relations to other clusters. These clusters are generally coarser and less nuanced compared to IRM and BCD rendering cluster interpretation difficult. An example of this is the large turquoise cluster shown in figure 3, which is composed of nodes in frontal, occipital, parietal, and temporal lobes.

The clusterings produced by IRM and BCD were very similar with mean NMI between clusterings of 0.94 at 8% link density. The similarity between the representations of IRM and BCD indicates that the flexibility of IRM is not needed when modeling rs-fMRI data. Even though IRM is able to cluster nodes such that the clustering does not obey the community structure, we see that IRM in general does produce clusterings which are very similar to BCD. The difference between BCD and IRM was most pronounced for smaller link densities suggesting that despite the large similarity between IRM and BCD it helps having the community structure constraint on the clustering. This is most evident for smaller link densities where the graphs contain less

information about the network. The better performance of BCD adds to the evidence that coordinated activation in the resting state is community structured.

By invoking a non-parametric Bayesian approach, the three modeling schemes considered are less sensitive to conventional model specification issues such as determining the number of communities as the number of clusters is inferred during model inference. However, the models' hyper-parameters still need to be set. In this study a uniform distribution was used as prior for the link probabilities (obtained by setting $\beta = 1$) and the CRP hyper-parameter α was set to 1, however other strategies could be considered. For instance, given the Bayesian framework it would be straightforward to sample the hyper-parameters as part of the model inference (Kemp et al., 2006). Our analysis scheme is a population level model, as we enforced graphs (subjects) to share the model's link probability matrix. The choice of fixing the link probabilities across subjects was done partly for evaluation purposes as it allows us to estimate the test likelihood for unseen data and thus allows us to estimate the models' predictability. However, the assumption that all subjects have similar linking between clusters with no inter-subject variability is somehow sub-optimal. Extending the models to allow for individual subject link probabilities is straight-forward (Mørup et al., 2010; Andersen et al., 2012a). The results can be seen as compressed networks, where the new nodes are formed by the clusters and (weighted) links are given by the individual subject link probabilities. This enables test for group differences in link probabilities or correlating with behavioral or personality measures where specific between-cluster linking can be considered and enables a more specific conclu-

sion about how, e.g., different population groups differ in linking structure. Using models which allow for individual subject link probabilities generally result in fewer clusters than models which restrict link probabilities to be equal across subjects (data not shown). This can be attributed to individual subject link probabilities causing a more flexible model, thereby capturing some of the flexibility which was previously accounted for by additional clusters.

A number of studies have reported relevance of the conventional network modularity measure to important cognitive measures, such as short term memory capacity, reaction time etc. (Bassett et al., 2011; Stevens et al., 2012; Meunier et al., 2009). Our findings suggest that there is important structure in resting state networks beyond the global modularity. The rich link structures of the relational models can be seen as a way of inferring functional integration at the inter-community level as discussed in (Hagmann et al., 2008; Sporns, 2013). Hence, an interesting open question is how to convert the flexible representations of the IRM and BCD to summary statistics that can be used as bio-markers. Indeed, initial evidence for the relevance of the community level link density ($\boldsymbol{\rho}$) as a bio-marker for multiple sclerosis was presented in Mørup et al. (2010)

When constructing the graphs one have to make decisions on how to define nodes and links. In this paper we used the brain regions from the AAL atlas (Tzourio-Mazoyer et al., 2002) to define nodes, which enables comparison with a large body of existing literature as the AAL atlas is the most commonly used atlas in the fMRI brain network literature (Stanley et al., 2013). The AAL regions have, however, been criticised for not reflecting

homogeneous functional units (Craddock et al., 2012). For instance, some of the AAL regions are rather large and thus could include more functional sub-units, which, when averaged together, degrade the functional connectivity to other regions. Similarly, if a functional unit lies between two or more AAL regions they will be purely represented. The main purpose of this paper was to compare different models and not, per se, interpret the resulting clusterings. The AAL definition thus allow for a broad comparison of the results with other methods already reported using this atlas. However, using a higher resolution network, e.g. by invoking an initial parcellation of voxels into functionally coherent units (Craddock et al., 2012), it will be possible to make more interesting neurobiological interpretations of the resulting clusters and their interactions. Likewise, the measure used for forming links between nodes have a great impact on the network. Here we used Pearson correlation between nodes' time series. As described in the introduction, a high correlation between two nodes can be found simply if the two nodes are both highly correlated with a third node. The use of partial correlations or the inverse covariance matrix (Varoquaux and Craddock, 2013) can remove correlations mediated by a third node and thus remove this transiency effect. However, partial correlations are less stable and are therefore also less reproducible than simple correlations.

The models presented here use simple graphs, that is, unweighted and undirected graphs, which requires that the correlation matrices are thresholded at a certain level. The results show that the choice of threshold have an impact on the resulting clusters and that different methods is affected differently by increasing the threshold. However, for IRM and BCD the

consistency between the found clusterings higher thresholds ($8\%, 16\%, 32\%$) were all very high as revealed by mean NMI all above 0.91. The exact choice of threshold is somewhat arbitrary to the problem as there is no natural threshold for which nodes can be said to be functionally connected. In addition, large negative correlations between nodes could provide important information to the network as well. However, negative coupling between nodes are by nature different from positive couplings, and thus should be considered differently in the network. Choosing a high threshold (meaning low link density) will fragment the network, that is, parts of the network or even single nodes will be unconnected to the rest of the network. In fact, in the networks presented here a number of nodes has very low or zero node degrees and are thus disconnected to the remaining network. An example of such nodes is seen in figure 3, where the IRM have clustered together a group of nodes[1] with low linking to the rest of the network (the cluster is best seen as the bright diagonal element in the lower right corner of the link probability matrix in the top panel of figure 3). These nodes are similar in the sense that they have low node degrees and can thus be represented as a 'null'-cluster. A similar 'null'-cluster[2] is found with the IDM model. For this cluster the within-cluster link probability is actually lower than the shared

[1]Green cluster composed of Amygdala L+R, Temporal Pole Mid L+R, Olfactory L+R, Pallidum L+R, Vermis 1 2, Vermis 3, Vermis 10, Cerebelum 3 L+R, Cerebelum 10 L+R. See Figure Appendix C.

[2]Red cluster composed of Amygdala L+R, Cerebelum 3 L, Cerebelum 7b R, Cerebelum 10 L+R, Olfactory L+R, Pallidum L+R, Temporal Pole Mid L+R, Vermis 1 2, Vermis 3, Vermis 7, Vermis 9, Vermis 10.

between-cluster link probability. For the BCD model this null-cluster is split into several smaller clusters which have slightly higher within-cluster link probabilities than between-cluster link probabilities to other clusters, thus conforming to the definition of the community model. Given the somewhat arbitrary thresholding for binary graphs, an interesting future direction is to model dense weighted graphs. In fact, a modification of the IRM have already been proposed, which models dense weighted graphs. The normal Infinite Relational Model (nIRM) (Herlau et al., 2012) models weighted links with a normal distribution instead of the Bernoulli distribution for binary links. This means that the correlation between nodes are modelled directly without thresholding. The relational models have also been extended to model hierarchical grouping structure and to allow for nodes to be members of multiple clusters, please see (Schmidt and Mørup, 2013) for an overview.

In conclusion, we evaluated three different Bayesian models for finding structure in rs-fMRI graphs and compared them with 3 other methods for node clustering in complex networks. We showed that BCD performs best compared to IRM and IDM in terms of predictability and reproducibility. This suggests that (1) rs-fMRI data adhere to the community structure and (2) modeling specific between-cluster linking improves predictability and reproducibility.

Toolbox

A Matlab toolbox for performing the experiments conducted in this paper can be found at `https://brainconnectivity.compute.dtu.dk/`

Acknowledgement

This work is funded by a project grant from the Lundbeck Foundation to Hartwig Siebner (grant-nr R48 A4846). The Magnetom Trio MR scanner was donated by the Simon Spies Foundation.

5. References

References

Aldous, D.J., 1985. Exchangeability and related topics. École d'Été de Probabilités de Saint-Flour XIII - 1983 (Lecture Notes in Mathematics) 1117, 1–198.

Alexander-Bloch, A.F., Gogtay, N., Meunier, D., Birn, R., Clasen, L., Lalonde, F., Lenroot, R., Giedd, J., Bullmore, E.T., 2010. Disrupted modularity and local connectivity of brain functional networks in childhood-onset schizophrenia. Frontiers in systems neuroscience 4.

Andersen, K.W., Madsen, K.H., Siebner, H., Hansen, L.K., Mørup, M., 2012a. Identification of Functional Clusters in the Striatum Using Infinite Relational Modeling, in: Langs, G., Rish, I., Grosse-Wentrup, M., Murphy, B. (Eds.), Machine Learning and Interpretation in Neuroimaging. Springer Berlin Heidelberg. Lecture Notes in Computer Science, pp. 226–233.

Andersen, K.W., Mørup, M., Siebner, H., Madsen, K.H., Hansen, L.K., 2012b. Identifying modular relations in complex brain networks, in: 2012 IEEE International Workshop on Machine Learning for Signal Processing, IEEE. pp. 1–6.

Ashburner, J., Friston, K.J., 2005. Unified segmentation. NeuroImage 26, 839–51.

Barabási, A.L., 2003. Linked: how everything is connected to everything else and what it means for business, science, and everyday life. Plume Editors.

800 Bassett, D.S., Wymbs, N.F., Porter, M.a., Mucha, P.J., Carlson, J.M.,
801 Grafton, S.T., 2011. Dynamic reconfiguration of human brain networks
802 during learning. Proceedings of the National Academy of Sciences of the
803 United States of America 108, 7641–6.

804 Birn, R.M., Diamond, J.B., Smith, M.A., Bandettini, P.A., 2006. Sep-
805 arating respiratory-variation-related fluctuations from neuronal-activity-
806 related fluctuations in fMRI. NeuroImage 31, 1536–48.

807 Biswal, B.B., Mennes, M., Zuo, X.N., Gohel, S., Kelly, C., Smith, S.M., Beck-
808 mann, C.F., Adelstein, J.S., Buckner, R.L., Colcombe, S., Dogonowski,
809 A.M., Ernst, M., Fair, D., Hampson, M., Hoptman, M.J., Hyde, J.S.,
810 Kiviniemi, V.J., Kötter, R., Li, S.J., Lin, C.P., Lowe, M.J., Mackay, C.,
811 Madden, D.J., Madsen, K.H., Margulies, D.S., Mayberg, H.S., McMahon,
812 K., Monk, C.S., Mostofsky, S.H., Nagel, B.J., Pekar, J.J., Peltier, S.J., Pe-
813 tersen, S.E., Riedl, V., Rombouts, S.a.R.B., Rypma, B., Schlaggar, B.L.,
814 Schmidt, S., Seidler, R.D., Siegle, G.J., Sorg, C., Teng, G.J., Veijola, J.,
815 Villringer, A., Walter, M., Wang, L., Weng, X.C., Whitfield-Gabrieli, S.,
816 Williamson, P., Windischberger, C., Zang, Y.F., Zhang, H.Y., Castellanos,
817 F.X., Milham, M.P., 2010. Toward discovery science of human brain func-
818 tion. Proceedings of the National Academy of Sciences of the United States
819 of America 107, 4734–9.

820 Blondel, V.D., Guillaume, J.L., Lambiotte, R., Lefebvre, E., 2008. Fast un-
821 folding of communities in large networks. Journal of Statistical Mechanics:
822 Theory and Experiment 2008, P10008.

823 Bullmore, E.T., Bassett, D.S., 2011. Brain graphs: graphical models of the
824 human brain connectome. Annual review of clinical psychology 7, 113–40.

825 Craddock, R.C., James, G., Holtzheimer, P.E., Hu, X.P., Mayberg, H.S.,
826 2012. A whole brain fMRI atlas generated via spatially constrained spectral
827 clustering. Human Brain Mapping 33, 1914–1928.

828 Dagli, M.S., Ingeholm, J.E., Haxby, J.V., 1999. Localization of cardiac-
829 induced signal change in fMRI. Neuroimage 9, 407–15.

830 Dahl, D.B., 2005. Sequentially-Allocated Merge-Split Sampler for Conjugate
831 and Nonconjugate Dirichlet Process Mixture Models. Technical Report,
832 Department of Statistics, Texas A&M University .

833 Damoiseaux, J.S., Rombouts, S., Barkhof, F., Scheltens, P., Stam, C.J.,
834 Smith, S.M., Beckmann, C.F., 2006. Consistent resting-state networks
835 across healthy subjects. Proceedings of the National Academy of Sciences
836 of the United States of America 103, 13848–13853.

837 Fortunato, S., 2010. Community detection in graphs. Physics Reports 486,
838 75–174. arXiv:arXiv:0906.0612v2.

839 Fortunato, S., Barthélemy, M., 2007. Resolution limit in community detec-
840 tion. Proceedings of the National Academy of Sciences of the United States
841 of America 104, 36–41.

842 Fox, M., Snyder, A., Vincent, J., 2005. The human brain is intrinsically
843 organized into dynamic, anticorrelated functional networks. Proceedings
844 of the National Academy of Sciences 102, 9673–9678.

Friston, K.J., Williams, S., Howard, R., Frackowiak, R.S., Turner, R., 1996. Movement-related effects in fMRI time-series. Magnetic resonance in medicine 35, 346–55.

Glover, G.H., Li, T.Q., Ress, D., 2000. Image-based method for retrospective correction of physiological motion effects in fMRI: RETROICOR. Magnetic Resonance in Medicine 44, 162–167.

Goutte, C., Toft, P., Rostrup, E., Nielsen, F., Hansen, L.K., 1999. On clustering fMRI time series. NeuroImage 9, 298–310.

Hagmann, P., Cammoun, L., Gigandet, X., Meuli, R., Honey, C.J., Wedeen, V.J., Sporns, O., 2008. Mapping the structural core of human cerebral cortex. PLoS biology 6, e159.

Herlau, T., Morup, M., Schmidt, M.N.., Hansen, L.K., 2012. Modelling dense relational data, in: Machine Learning for Signal Processing (MLSP), 2012 IEEE International Workshop on, pp. 1–6.

van den Heuvel, M.P., Sporns, O., 2011. Rich-club organization of the human connectome. The Journal of neuroscience : the official journal of the Society for Neuroscience 31, 15775–86.

van den Heuvel, M.P., Stam, C.J., Boersma, M., Hulshoff Pol, H.E., 2008. Small-world and scale-free organization of voxel-based resting-state functional connectivity in the human brain. NeuroImage 43, 528–39.

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, 158–182.

Kemp, C., Tenenbaum, J., Griffiths, T., Yamada, T., Ueda, N., 2006. Learning systems of concepts with an infinite relational model, in: Proceedings of the 21th National Conference on Artificial Intelligence (AAAI), Menlo Park, CA; Cambridge, MA; London; AAAI Press; MIT Press; 1999. pp. 381–388.

Lehmann, S., Hansen, L.K., 2007. Deterministic Modularity Optimization. The European Physical Journal B 60, 83–88.

Lund, T.E., Madsen, K.H., Sidaros, K., Luo, W.L., Nichols, T.E., 2006. Non-white noise in fMRI: does modelling have an impact? NeuroImage 29, 54–66.

McKeown, M.J., Hansen, L.K., Sejnowski, T.J., 2003. Independent component analysis of functional MRI: what is signal and what is noise? Current Opinion in Neurobiology 13, 620–629.

McKeown, M.J., Makeig, S., Brown, G.G., Jung, T.P., Kindermann, S.S., Bell, A.J., Sejnowski, T.J., 1998. Analysis of fMRI data by blind separation into independent spatial components. Human brain mapping 6, 160–188.

Meunier, D., Achard, S., Morcom, A., Bullmore, E., 2009. Age-related changes in modular organization of human brain functional networks. NeuroImage 44, 715–23.

Mørup, M., Madsen, K.H., Dogonowski, A.M., Siebner, H., Hansen, L.K., 2010. Infinite relational modeling of functional connectivity in resting state fMRI. Advances in Neural Information Processing Systems 23 , 1750–1758.

Mørup, M., Schmidt, M.N., 2012. Bayesian community detection. Neural computation 24, 2434–56.

Newman, M.E.J., 2006. Modularity and community structure in networks. Proceedings of the National Academy of Sciences of the United States of America 103, 8577–82.

Nowicki, K., Snijders, T., 2001. Estimation and prediction for stochastic blockstructures. Journal of the American Statistical Association 96, 1077–1087.

Oldfield, R.C., 1971. The assessment and analysis of handedness: The Edinburgh inventory. Neuropsychologia 9, 97–113.

Power, J.D., Cohen, A.L., Nelson, S.M., Wig, G.S., Barnes, K.A., Church, J.A., Vogel, A.C., Laumann, T.O., Miezin, F.M., Schlaggar, B.L., Petersen, S.E., 2011. Functional Network Organization of the Human Brain. Neuron 72, 665–678.

Power, J.D., Mitra, A., Laumann, T.O., Snyder, A.Z., Schlaggar, B.L., Petersen, S.E., 2014. Methods to detect, characterize, and remove motion artifact in resting state fMRI. NeuroImage 84, 320–41.

Rosvall, M., Bergstrom, C., 2008. Maps of random walks on complex networks reveal community structure. Proceedings of the National Academy of Sciences 105, 1118–1123. `arXiv:0709.4500`.

Rubinov, M., Sporns, O., 2010. Complex network measures of brain connectivity: uses and interpretations. NeuroImage 52, 1059–69.

912 Schmidt, M.N., Mørup, M., 2013. Nonparametric Bayesian Modeling of Com-
913 plex Networks. IEEE Signal Processing Magazine 30, 110–128.

914 Smith, A.M., Lewis, B.K., Ruttimann, U.E., Ye, F.Q., Sinnwell, T.M., Yang,
915 Y., Duyn, J.H., Frank, J.A., 1999. Investigation of low frequency drift in
916 fMRI signal. Neuroimage 9, 526–33.

917 Sporns, O., 2011. The human connectome: a complex network. Annals of
918 the New York Academy of Sciences 1224, 109–25.

919 Sporns, O., 2013. Network attributes for segregation and integration in the
920 human brain. Current Opinion in Neurobiology 23, 162–171.

921 Stanley, M.L., Moussa, M.N., Paolini, B.M., Lyday, R.G., Burdette, J.H.,
922 Laurienti, P.J., 2013. Defining nodes in complex brain networks. Frontiers
923 in computational neuroscience 7, 169.

924 Stevens, A.A., Tappon, S.C., Garg, A., Fair, D.A., 2012. Functional brain
925 network modularity captures inter- and intra-individual variation in work-
926 ing memory capacity. PloS one 7, e30468.

927 Strother, S.C., Anderson, J., Hansen, L.K., Kjems, U., Kustra, R., Sidtis, J.,
928 Frutiger, S., Muley, S., LaConte, S., Rottenberg, D., 2002. The quantita-
929 tive evaluation of functional neuroimaging experiments: the NPAIRS data
930 analysis framework. NeuroImage 15, 747–71.

931 Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard,
932 O., Delcroix, N., Mazoyer, B., Joliot, M., 2002. Automated anatomical
933 labeling of activations in SPM using a macroscopic anatomical parcellation
934 of the MNI MRI single-subject brain. Neuroimage 15, 273–89.

Varoquaux, G., Craddock, R.C., 2013. Learning and comparing functional connectomes across subjects. NeuroImage 80, 405–15.

Xu, Z., Tresp, V., Yu, K., Kriegel, H., 2006. Infinite hidden relational models, in: In Proceedings of the 22nd International Conference on Uncertainity in Artificial Intelligence (UAI), Citeseer.

Zalesky, A., Fornito, A., Bullmore, E., 2012. On the use of correlation as a measure of network connectivity. Neuroimage 60, 2096–2106.

Appendix A. FCON1000 data

Data included from Beijing Zang (data included in Beijing_Zang_part2.tar): sub20127, sub20246, sub20765, sub20948, sub21115, sub22201, sub22595, sub22661, sub22715, sub22890, sub26713, sub28206, sub28403, sub28698, sub28792, sub28801, sub28907, sub28965, sub29590, sub29785, sub30272, sub30310, sub30556, sub30616, sub30988, sub31058, sub31729, sub32517, sub32587, sub33747, sub33943, sub33991, sub34895, sub34943, sub35309, sub35776, sub35806, sub36580, sub36942, sub37602, sub38602, sub39725.

Data included from Leipzig: sub00321, sub01002, sub02075, sub07097, sub07374, sub07516, sub07786, sub18698, sub23427, sub25344, sub31577, sub31637, sub36858, sub37308, sub41241, sub47452, sub49383, sub52507, sub52858, sub53063, sub53394, sub59494, sub59709, sub59861, sub61373, sub63957, sub64446, sub72508, sub75022, sub75886, sub77802, sub80206, sub80552, sub85213, sub90843, sub92903, sub94784.

Appendix B. Inference

Appendix B.1. IRM

As stated in section 2.4 the generative model for the Infinite Relational Model is

Infinite Relational Model	
Cluster assignments:	$\mathbf{z} \sim \mathrm{CRP}(\alpha)$
Link probabilities:	$\rho_{k,l} \sim \mathrm{Beta}(\beta, \beta)$
Links:	$A_{i,j}^{(n)} \sim \mathrm{Bernoulli}(\rho_{z_i,z_j})$

For brevity we define the joint set of graphs as $\mathbf{A} = \{\mathbf{A}^{(1)}, ..., \mathbf{A}^{(N)}\}$. The Bernoulli likelihood can then be written as:

$$
\begin{aligned}
P(\mathbf{A}|\mathbf{z}, \boldsymbol{\rho}) &= \prod_n \prod_{j>i} \rho_{z_i,z_j}^{A_{i,j}^{(n)}} (1 - \rho_{z_i,z_j})^{\left(1 - A_{i,j}^{(n)}\right)} \\
&= \prod_{j>i} \rho_{z_i,z_j}^{\left(\sum_n A_{i,j}^{(n)}\right)} (1 - \rho_{z_i,z_j})^{\left(N - \sum_n A_{i,j}^{(n)}\right)} \\
&= \prod_{k \geq l} \rho_{k,l}^{N_{k,l}^{+}} (1 - \rho_{k,l})^{N_{k,l}^{-}},
\end{aligned}
$$

where $N_{k,l}^{+}$ and $N_{k,l}^{-}$ is the total number of links and nonlinks for all graphs between cluster k and l, respectively and N is the number of graphs (subjects). The prior for the link probabilities is a symmetric Beta distribution and can be written as

$$
P(\boldsymbol{\rho}|\beta) = \prod_{k \geq l} \frac{\Gamma(2\beta)}{\Gamma(\beta)^2} \rho_{k,l}^{\beta-1} (1 - \rho_{k,l})^{\beta-1}
$$

where $\Gamma(x) = (x-1)!$ is the gamma function. The CRP prior for the node partition can be written as

$$P(\mathbf{z}|\alpha) = \frac{\alpha^K \Gamma(\alpha) \prod_k \Gamma(n_k)}{\Gamma(J+\alpha)}, \tag{B.1}$$

where J is the number of nodes per graph, n_k is the number of nodes assigned to cluster k and K is the number of clusters. These distributions are combined to yield the joint distribution for the IRM:

$$\begin{aligned} P(\mathbf{A}, \mathbf{z}, \boldsymbol{\rho}|\alpha, \beta) &= P(\mathbf{A}|\mathbf{z}, \boldsymbol{\rho}) P(\boldsymbol{\rho}|\beta) P(\mathbf{z}|\alpha) \\ &= \Big[\prod_{k \geq l} \rho_{k,l}^{N^+_{k,l}} (1-\rho_{k,l})^{N^-_{k,l}}\Big] \Big[\prod_{k \geq l} \frac{\Gamma(2\beta)}{\Gamma(\beta)^2} \rho_{k,l}^{\beta-1} (1-\rho_{k,l})^{\beta-1}\Big] \\ &\quad \times \Big[\frac{\alpha^K \Gamma(\alpha) \prod_k \Gamma(n_k)}{\Gamma(J+\alpha)}\Big] \\ &= \Big[\prod_{k \geq l} \frac{\Gamma(2\beta)}{\Gamma(\beta)^2} \rho_{k,l}^{N^+_{k,l}+\beta-1} (1-\rho_{k,l})^{N^-_{k,l}+\beta-1}\Big] \\ &\quad \times \Big[\frac{\alpha^K \Gamma(\alpha) \prod_k \Gamma(n_k)}{\Gamma(J+\alpha)}\Big] \end{aligned}$$

Now we can marginalize $\boldsymbol{\rho}$:

$$\begin{aligned} P(\mathbf{A}, \mathbf{z}, |\alpha, \beta) &= \int P(\mathbf{A}, \mathbf{z}, \boldsymbol{\rho}|\alpha, \beta) \mathrm{d}\boldsymbol{\rho} \\ &= \Big[\prod_{k \geq l} \frac{\mathrm{B}(N^+_{k,l}+\beta, N^-_{k,l}+\beta)}{\mathrm{B}(\beta, \beta)}\Big] \Big[\frac{\alpha^K \Gamma(\alpha) \prod_k \Gamma(n_k)}{\Gamma(J+\alpha)}\Big] \end{aligned}$$

where $\mathrm{B}(x,y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}$ is the Beta function. Finally using Bayes' theorem we can find the posterior distribution of the assignment of a single node z_i

$$P(z_i = l|\mathbf{A}, \mathbf{z}_{\setminus i}, \beta, \alpha) = \frac{P(\mathbf{A}, \mathbf{z}_{\setminus i}, z_i = l|\alpha, \beta)}{\sum_{l'} P(\mathbf{A}, \mathbf{z}_{\setminus i}, z_i = l'|\alpha, \beta)}$$

where $\mathbf{z}_{\setminus i}$ is the assignments of all nodes except node i. By writing out this equation and finding parts which change when a node is assigned to a cluster

(Schmidt and Mørup, 2013) we have that:

$$P(z_i = l|\mathbf{A}, \mathbf{z}_{\setminus i}, \beta, \alpha) \propto \begin{cases} n_{l\setminus i} \prod_k \frac{\mathrm{B}(N_{k,l}^{+\setminus i}+r_{i,k}^{+}+\beta, N_{k,l}^{-\setminus i}+r_{i,k}^{-}+\beta)}{\mathrm{B}(N_{k,l}^{+\setminus i}+\beta, N_{k,l}^{-\setminus i}+\beta)} & \text{if } n_{l\setminus i} > 0 \\ \alpha \prod_k \frac{\mathrm{B}(r_{i,k}^{+}+\beta, r_{i,k}^{-}+\beta)}{\mathrm{B}(\beta,\beta)} & \text{otherwise.} \end{cases} \tag{B.2}$$

$N_{k,l}^{+\setminus i}$ and $N_{k,l}^{-\setminus i}$ is the number of links and nonlinks between clusters k and l not counting links from node i. $n_{l\setminus i}$ is the number of nodes assigned to cluster l disregarding the assignment of node i. $r_{i,k}^{+}$ and $r_{i,k}^{-}$ is the number of links and nonlinks from node i to any node in cluster k. This posterior likelihood can be evaluated efficiently since we only need to compute $\mathbf{N}^{+}$ and $\mathbf{N}^{-}$ and evaluate the Beta function for entries affected by the considered assignment change. The posterior likelihood is used in the Gibbs sampler to infer the node assignments.

Appendix B.2. IDM

The generative model for the Infinite Diagonal Model is given by:

Infinite Diagonal Model

Cluster assignments:	$\mathbf{z} \sim \mathrm{CRP}(\alpha)$
Link probabilities:	$\rho_{k,l} \sim \begin{cases} \rho_k = \mathrm{Beta}(\beta,\beta) & \text{if } k = l \\ \rho_b = \mathrm{Beta}(\beta,\beta) & \text{otherwise.} \end{cases}$
Links:	$A_{i,j}^{(n)} \sim \mathrm{Bernoulli}(\rho_{z_i,z_j})$

The Bernoulli likelihood can accordingly be written as:

$$P(\mathbf{A}|\mathbf{z}, \boldsymbol{\rho}) = \rho_b^{N_b^{+}} (1-\rho_b)^{N_b^{-}} \Big[\prod_k \rho_k^{N_k^{+}} (1-\rho_k)^{N_k^{-}} \Big],$$

where N_k^+ and N_k^- is the number of links and nonlinks within cluster k and N_b^+ and N_b^- is the total number of links and nonlinks which fall between clusters. The prior for the link probabilities can be written as

$$P(\boldsymbol{\rho}|\beta) = \frac{\Gamma(2\beta)}{\Gamma(\beta)^2}\rho_b^{\beta-1}(1-\rho_b)^{\beta-1}\Big[\prod_k \frac{\Gamma(2\beta)}{\Gamma(\beta)^2}\rho_k^{\beta-1}(1-\rho_k)^{\beta-1}\Big]$$

The prior for the node partition is the same as the IRM model (equation B.1). The joint distribution for the IDM can then be written as:

$$\begin{aligned}
P(\mathbf{A},\mathbf{z},\boldsymbol{\rho}|\alpha,\beta) &= P(\mathbf{A}|\mathbf{z},\boldsymbol{\rho})P(\boldsymbol{\rho}|\beta)P(\mathbf{z}|\alpha) \\
&= \Big[\rho_b^{N_b^+}(1-\rho_b)^{N_b^-}\Big]\Big[\prod_k \rho_k^{N_k^+}(1-\rho_k)^{N_k^-}\Big] \\
&\quad \times\Big[\frac{\Gamma(2\beta)}{\Gamma(\beta)^2}\rho_b^{\beta-1}(1-\rho_b)^{\beta-1}\Big]\Big[\prod_k \frac{\Gamma(2\beta)}{\Gamma(\beta)^2}\rho_k^{\beta-1}(1-\rho_k)^{\beta-1}\Big] \\
&\quad \times\Big[\frac{\alpha^K\Gamma(\alpha)\prod_k\Gamma(n_k)}{\Gamma(J+\alpha)}\Big] \\
&= \frac{\Gamma(2\beta)}{\Gamma(\beta)^2}\rho_b^{N_b^++\beta-1}(1-\rho_b)^{N_b^-+\beta-1} \\
&\quad \times\Big[\prod_k \frac{\Gamma(2\beta)}{\Gamma(\beta)^2}\rho_k^{N_k^++\beta-1}(1-\rho_k)^{N_k^-+\beta-1}\Big] \\
&\quad \times\Big[\frac{\alpha^K\Gamma(\alpha)\prod_k\Gamma(n_k)}{\Gamma(J+\alpha)}\Big]
\end{aligned}$$

Now marginalizing over $\boldsymbol{\rho}$:

$$\begin{aligned}
P(\mathbf{A},\mathbf{z}|\alpha,\beta) &= \int P(\mathbf{A},\mathbf{z},\boldsymbol{\rho}|\alpha,\beta)\mathrm{d}\boldsymbol{\rho} \\
&= \frac{\mathrm{B}(N_b^++\beta,N_b^-+\beta)}{\mathrm{B}(\beta,\beta)}\Big[\prod_k \frac{\mathrm{B}(N_k^++\beta,N_k^-+\beta)}{\mathrm{B}(\beta,\beta)}\Big] \\
&\quad \times\Big[\frac{\alpha^K\Gamma(\alpha)\prod_k\Gamma(n_k)}{\Gamma(J+\alpha)}\Big]
\end{aligned}$$

Finally using Bayes' theorem we can find the posterior distribution of the assignment of a single node z_i

$$P(z_i = l|\mathbf{A}, \mathbf{z}_{\setminus i}, \beta, \alpha) = \frac{P(\mathbf{A}, \mathbf{z}_{\setminus i}, z_i = l|\alpha, \beta)}{\sum_{l'} P(\mathbf{A}, \mathbf{z}_{\setminus i}, z_i = l'|\alpha, \beta)}$$

By writing out this equation and finding parts which change when a node is assigned to a cluster we find that:

$$P(z_i = l|\mathbf{A}, \mathbf{z}_{\setminus i}, \beta, \alpha) \propto \begin{cases} n_{l\setminus i} \frac{\mathrm{B}(N_l^{+\setminus i}+r_{i,l}^+ +\beta, N_l^{-\setminus i}+r_{i,l}^- +\beta)}{\mathrm{B}(N_l^{+\setminus i}+\beta, N_l^{-\setminus i}+\beta)} \frac{\mathrm{B}(N_b^{+\setminus i}+\sum_{k\neq l} r_{i,k}^+ +\beta, N_b^{-\setminus i}+\sum_{k\neq l} r_{i,k}^- +\beta)}{\mathrm{B}(N_b^{+\setminus i}+\beta, N_b^{-\setminus i}+\beta)} & \text{if } n_{l\setminus i} > 0 \\ \alpha \frac{\mathrm{B}(N_b^{+\setminus i}+\sum_{k\neq l} r_{i,k}^+ +\beta, N_b^{-\setminus i}+\sum_{k\neq l} r_{i,k}^- +\beta)}{\mathrm{B}(N_b^{+\setminus i}+\beta, N_b^{-\setminus i}+\beta)} & \text{otherwise.} \end{cases} \tag{B.3}$$

$r_{i,l}^+$ and $r_{i,l}^-$ is the number of links and nonlinks from node i to any node in cluster l.

Appendix B.3. BCD

This section will give a short description of the inference in the Bayesian Community Detection (BCD) model, for further details please refer to Mørup and Schmidt (2012). The generative model for BCD is given by:

Bayesian Community Detection

$$
\begin{aligned}
\text{Cluster assignments}: &\quad \mathbf{z} \sim \text{CRP}(\alpha) \\
\text{Cluster gap}: &\quad \gamma \sim \text{Beta}(v, v) \\
\text{Link probability}: &\quad \rho_{k,l} \sim \begin{cases} \text{Beta}(\beta, \beta) & \text{if } k = l \\ \text{BetaInc}(\beta, \beta, w_{k,l}) & \text{otherwise.} \end{cases} \\
&\quad \text{where } w_{k,l} = \min[\gamma\rho_{ll}, \gamma\rho_{kk}] \\
\text{Links}: &\quad A_{i,j}^{(n)} \sim \text{Bernoulli}(\rho_{z_i,z_j})
\end{aligned}
$$

If we let $\dot{\boldsymbol{\rho}} = \{\rho_{k,l} | k = l\}$ and $\ddot{\boldsymbol{\rho}} = \{\rho_{k,l} | k \neq l\}$ be the set of within and between link probabilities respectively. Then the joint distribution can be written as

$$
\begin{aligned}
P(\mathbf{A}, \mathbf{z}, \boldsymbol{\rho}, \gamma | \alpha, \beta) &= P(\mathbf{A}|\mathbf{z}, \boldsymbol{\rho}) P(\ddot{\boldsymbol{\rho}}|\dot{\boldsymbol{\rho}}, \gamma, \beta) P(\dot{\boldsymbol{\rho}}|\beta) P(\gamma|v) P(\mathbf{z}|\alpha) \\
&= \Big[\prod_{n=1}^{N} \prod_{j>i} \rho_{z_i,z_j}^{A_{i,j}^{(n)}} (1 - \rho_{z_i,z_j})^{1 - A_{i,j}^{(n)}} \Big] \\
&\quad \times \Big[\prod_{k>l} \frac{\rho_{k,l}^{\beta-1} (1 - \rho_{k,l})^{\beta-1}}{\text{B}_{x_{k,l}}(\beta, \beta)} \Big] \Big[\prod_{l=1}^{K} \frac{\rho_{l,l}^{\beta-1} (1 - \rho_{l,l})^{\beta-1}}{\text{B}(\beta, \beta)} \Big] \\
&\quad \times \Big[\frac{\gamma^{v-1} (1 - \gamma)^{v-1}}{\text{B}(v, v)} \Big] \Big[\frac{\alpha^K \Gamma(\alpha) \prod_k \Gamma(n_k)}{\Gamma(J + \alpha)} \Big]
\end{aligned}
$$

Integrating over $\ddot{\boldsymbol{\rho}}$:

$$\begin{aligned} P(\mathbf{A},\mathbf{z},\dot{\boldsymbol{\rho}},\gamma|\alpha,\beta) &= \int P(\mathbf{A},\mathbf{z},\boldsymbol{\rho},\gamma|\alpha,\beta)\mathrm{d}\ddot{\boldsymbol{\rho}} \\ &= \Big[\prod_{k=1}^{K}\frac{\rho_{k,k}^{N_{k,k}^{+}+\beta-1}(1-\rho_{k,k})^{N_{k,k}^{-}+\beta-1}}{\mathrm{B}(\beta,\beta)}\Big] \\ &\quad\times\Big[\prod_{k>l}\frac{\mathrm{B}_{x_{k,l}}(N_{k,l}^{+}+\beta,N_{k,l}^{-}+\beta)}{\mathrm{B}_{x_{k,l}}(\beta,\beta)}\Big] \\ &\quad\times\Big[\frac{\gamma^{v-1}(1-\gamma)^{v-1}}{\mathrm{B}(v,v)}\Big]\Big[\frac{\alpha^{K}\Gamma(\alpha)\prod_{k}\Gamma(n_k)}{\Gamma(J+\alpha)}\Big] \end{aligned}$$

Again, using Bayes theorem and eliminating terms which does not depend on $\rho_{l,l}$ the marginal posterior reduces to

$$P(\rho_{l,l}|\mathbf{A},\mathbf{z},\dot{\boldsymbol{\rho}}\backslash\rho_{l,l},\beta,\alpha,\gamma) \propto \rho_{l,l}^{N_{l,l}^{+}+\beta-1}(1-\rho_{l,l})^{N_{l,l}^{-}+\beta-1}\prod_{k\neq l}\frac{\mathrm{B}_{x_{l,k}}(N_{k,l}^{+}+\beta,N_{k,l}^{-}+\beta)}{\mathrm{B}_{x_{k,l}}(\beta,\beta)}$$

The conditional distribution for a node assignment is given as (Mørup and Schmidt, 2012):

$$\begin{aligned} P(z_i=l|\mathbf{A},\mathbf{z}_{\backslash i},\dot{\boldsymbol{\rho}},\beta,\alpha,\gamma) &\propto \rho_{l,l}^{r_{i,l}^{+}}(1-\rho_{l,l})^{r_{i,l}^{-}}\alpha^{K}n_{l\backslash i} \\ &\quad\prod_{k\neq l}\frac{\mathrm{B}_{x_{k,l}}(N_{k,l}^{+\backslash i}+r_{i,k}^{+}+\beta,N_{k,l}^{-\backslash i}+r_{i,k}^{-}+\beta)}{\mathrm{B}_{x_{k,l}}(N_{k,l}^{+\backslash i}+\beta,N_{k,l}^{-\backslash i}+\beta)} \end{aligned}$$

When terms which does not depend on γ are ignored the posterior reduces to

$$P(\gamma|\mathbf{A},\mathbf{z},\dot{\boldsymbol{\rho}},\beta,\alpha) \propto \gamma^{v-1}(1-\gamma)^{v-1}\prod_{k>l}\frac{\mathrm{B}_{x_{k,l}}(N_{k,l}^{+\backslash i}+\beta,N_{k,l}^{-\backslash i}+\beta)}{\mathrm{B}_{x_{k,l}}(\beta,\beta)}$$

Appendix C. Clusters labels

Figure C.9: Labels from the extracted clusters using IRM, BCD, and IDM. The colors correspond to the clusters from figure 3.