Nonparametric Bayesian Clustering of Structural Whole Brain Connectivity in Full Image Resolution

Karen Sandø Ambrosen*†, Kristoffer Jon Albers*, Tim B. Dyrby†, Mikkel N. Schmidt*, and Morten Mørup*
*Department of Applied Mathematics and Computer Science, Technical University of Denmark, Lyngby, Denmark
†Danish Research Centre for Magnetic Resonance, Copenhagen University Hospital Hvidovre, Hvidovre, Denmark

***Abstract*—Diffusion magnetic resonance imaging enables measuring the structural connectivity of the human brain at a high spatial resolution. Local noisy connectivity estimates can be derived using tractography approaches and statistical models are necessary to quantify the brain's salient structural organization. However, statistically modeling these massive structural connectivity datasets is a computational challenging task. We develop a high-performance inference procedure for the infinite relational model (a prominent non-parametric Bayesian model for clustering networks into structurally similar groups) that defines structural units at the resolution of statistical support. We apply the model to a network of structural brain connectivity in full image resolution with more than one hundred thousand regions (voxels in the gray-white matter boundary) and around one hundred million connections. The derived clustering identifies in the order of one thousand salient structural units and we find that the identified units provide better predictive performance than predicting using the full graph or two commonly used atlases. Extracting structural units of brain connectivity at the full image resolution can aid in understanding the underlying connectivity patterns, and the proposed method for large scale data driven generation of structural units provides a promising framework that can exploit the increasing spatial resolution of neuro-imaging technologies.**

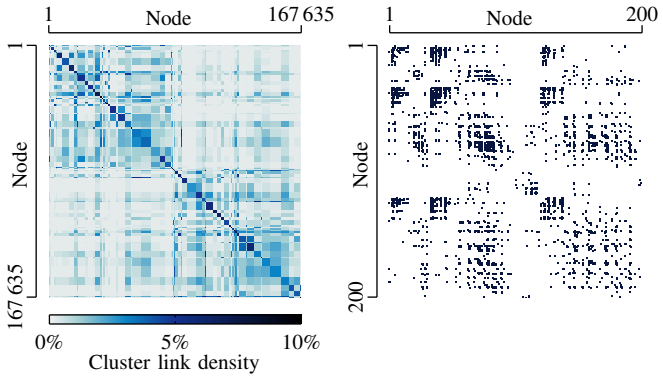

Fig. 1. Complex network of structural brain connectivity with 167 635 nodes and around one hundred million links obtained using 5000 streamlines per seed voxel. Left: Link density in each pair of the 68 regions of interest in the Desikan-Killiany atlas. Right: Links between the first 200 regions.

I. Introduction

Diffusion magnetic resonance imaging (dMRI) is an important non-invasive technique for studying the brain's structural organization. By tracking the diffusion of mainly water molecules that align with the orientation of the fibers in the brain, local estimates of fiber orientation can be obtained. These estimates are aggregated by tractography to derive maps of structural connectivity between cortical gray matter regions [5]. For the current dMRI technology these maps in full image resolution constitute complex networks of structural connectivity in the order of one hundred thousand regions and one hundred million links (see Fig. 1).

While the quantified fiber orientation within small regions of the brain as well as the subsequently derived local connectivity estimates are very noisy, these estimates can be aggregated to derive networks of whole brain connectivity within larger regions of structural units. These structural units have traditionally been based on automatic subdivision of the human brain into a fixed number of pre-specified neuroanatomical regions of interests (ROIs) [13], [7]. The Destrieux atlas [13], [8] currently has around 150 ROIs whereas the Desikan-Killiany atlas [7] has 68 ROIs. While these ROIs can be arbitrarily subdivided to provide additional regions [14] they are not explicitly based on the evidence obtained by the structural connectivity data and may therefore not optimally reflect the latent connectivity patterns of structural connectivity. Rather than fixing the structural units to a predefined atlas, we set out to learn the number of structural units and their spatial representations from the raw high resolution networks obtained using tractography. To accomplish this we develop a large scale implementation of a prominent statistical network model, the infinite relational model (IRM) [17], [24]. The IRM is able to infer structurally consistent units at a resolution which is determined based on statistical evidence. While structural connectivity graphs have previously been clustered based on IRM [3] as well as other tools such as modularity [14], this is to the best of our knowledge the first attempt at modelling structural connectivity at the full image resolution of current dMRI technology.

This paper examines the capabilities of a large scale implementation of the IRM to identify structure in high-resolution structural brain connectivity graphs. We study to what extend we can perform inference on such large scale networks and whether it is feasible to reliably detect the structural units in a data driven manner using our implementation. In particular, we investigate: i) *What is the statistically salient resolution of structural connectivity graphs*, i.e., how many clusters are used to represent high resolution structural connectivity data? ii) *How reliable can these salient structures be detected*, i.e. how consistent are the structural units with respect to initialization and convergence of the sampler as well as the number of streamlines? iii) *Are the derived structural units better at predicting connectivity than existing atlases*, i.e. how well does the connectivity patterns derived from the structural units of one graph predict the connectivity of another graph obtained from another set of whole brain diffusion weighted images from the same subject?

978-1-4799-4149-0/14/$31.00

II. Statistical model and inference

A. Infinite Relational Modelling

The Infinite Relational Model (IRM) [17], [24] is a non-parametric extension of the stochastic block model [19] in which vertices in a graph are grouped into homogenous blocks according to their structural similarity. The IRM uses the Chinese Restaurant Process (CRP) [2], [20] as prior for the partitioning of vertices to groups thereby allowing for an arbitrary number of groups. The IRM is defined by the following generative process:

$$\boldsymbol{z} \sim \mathrm{CRP}(\alpha), \qquad \text{groups}, \tag{1}$$

$$\eta_{lm} \sim \mathrm{Beta}(\beta^+, \beta^-), \qquad \text{interactions}, \tag{2}$$

$$A_{ij} \sim \mathrm{Bernoulli}(\eta_{z_i z_j}), \qquad \text{links}, \tag{3}$$

where $\boldsymbol{z}$ is the group assignment, $\boldsymbol{\eta}$ is the probability of links between each pair of groups, and $\boldsymbol{A}$ is the adjacency matrix of the graph. As the beta prior on the elements of $\boldsymbol{\eta}$ is conjugate to the Bernoulli likelihood these parameters can be analytically integrated to form the joint distribution:

$$P(\boldsymbol{A}, \boldsymbol{z}|\alpha, \beta^+, \beta^-) = \int P(\boldsymbol{A}, \boldsymbol{z}, \boldsymbol{\eta}|\alpha, \beta^+, \beta^-) d\boldsymbol{\eta} \tag{4}$$

$$= \frac{\alpha^K \Gamma(\alpha) \prod_k \Gamma(n_k)}{\Gamma(J+\alpha)} \prod_{l \leq m} \frac{\mathrm{B}(N_{lm}^+ + \beta^+, N_{lm}^- + \beta^-)}{\mathrm{B}(\beta^+, \beta^-)},$$

where K is the number of groups, J is the number of vertices, n_k is the number of vertices assigned to the k'th group, N_{lm}^+ and N_{lm}^- are the number of links and non-links between group l and m, and $B(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$ is the beta function.

B. Inference by Markov Chain Monte Carlo (MCMC)

To infer the posterior distribution, $P(\boldsymbol{z}|\boldsymbol{A}, \alpha, \beta^+, \beta^-)$, we use an MCMC procedure combining Gibbs and split-merge sampling [17]. In Gibbs sampling the posterior conditional distribution of placing one vertex at a time in any of the existing groups or in a new empty group is evaluated and the vertex is assigned according to this distribution. The probability of assigning a vertex i to group ℓ is given by:

$$P(z_i = \ell|\boldsymbol{A}, \boldsymbol{z}_{\setminus i}, h) = \frac{P(\boldsymbol{A}, \boldsymbol{z}_{\setminus i}, z_i = \ell|h)}{\sum_{\ell'=1}^{K+1} P(\boldsymbol{A}, \boldsymbol{z}_{\setminus i}, z_i = \ell'|h)}, \tag{5}$$

where $h = \{\beta^+, \beta^-, \alpha\}$ denotes the hyperparameters.

Rather than considering the assignment of a single vertex at a time split-merge sampling as presented in [15] attempts to merge or split existing clusters. Here, two vertices i and j are selected at random. If they are currently assigned to two different groups $z_i \neq z_j$, it is proposed to merge the two groups. Else it is proposed to split the single group in two. The procedure makes use of Gibbs sampling restricted to the nodes of the considered group(s) in order to define an intermediate launch state as well as to define the final split configuration and its transition probability $q(z|z^*)$. For a split configuration $q(z|z^*)$ is derived as the product of the individual transition probabilities of the vertices to move from the launch state to the final split configuration. As a merge transition is deterministic the transition from a split to a merge configuration has probability 1. Proposals are rejected or accepted according to the Metropolis-Hastings acceptance probability:

$$\alpha(z^*|z) = \min\left[1, \frac{P(\boldsymbol{A}, \boldsymbol{z}^*|\beta^+, \beta^-, \alpha) q(z|z^*)}{P(\boldsymbol{A}, \boldsymbol{z}|\beta^+, \beta^-, \alpha) q(z^*|z)}\right]. \tag{6}$$

C. Large scale computation

To get the computational performance necessary for the IRM to model structural connectivity in full image resolution we used a dedicated implementation optimized towards fully utilizing the memory structure and processor architecture of modern computers (see [1] for details). As the restricted Gibbs sweeps turns out to be the most computational demanding part of the split-merge sampling procedure, the performance of both sampling strategies benefits from most of the same optimizations. We store data in appropriate structures such that the sampling algorithms access data elements from sequential memory. In this way the access pattern takes advantage of the memory cache structure allowing for significantly faster memory accesses. To further speed up the Gibbs sampler we store and update the sufficient statistics, N^+ and N^-, instead of recalculating them in every Gibbs sweep. To ensure numeric stability within machine precision, the posterior in Eq. 5 is calculated in the log domain. The key operation then becomes calculating the logarithm of the beta function which relies on the gamma-function, $\Gamma(a)$, as:

$$\log B(a,b) = \log\Gamma(a) + \log\Gamma(b) - \log\Gamma(a+b) \tag{7}$$

As we only allow integer values for the hyperparameters, we use a lookup table of precalculated values for $\log\Gamma(a)$ which speeds up the evaluation of the posterior.

III. Data

To validate our proposed method, we used a dMRI data set previously described in [23], [22]. The data was collected at Danish Research Center for Magnetic Resonance and the study was approved by the local ethics committee. One healthy subject was scanned. The images were acquired on a Siemens VERIO 3T scanner using a 32-channel head coil. Two high resolution T1-weighted MRI images were acquired using a TR of 1,900 ms, TE of 2.32 ms, a FA of 9°, and 0.9mm^3 isotropic resolution. Two sets of whole brain diffusion weighted images (DWI) were acquired in 61 non-collinear directions with a b-value of b = 1500 s/mm^2, and ten non-diffusion weighted images (b = 0 s/mm^2). For this the twice refocused spin echo sequence with a TR of 11,440 ms and a TE of 89 ms. 61 axial slices with a resolution of 2.3 mm^3 isotropic voxels and Grappa = 2 were acquired [21]. A field map was acquired using a double gradient echo sequence with a TR of 479 ms, TE1 of 4.92 ms, TE2 of 7.38 ms, and a resolution of 3mm^3 isotropic voxels. The diffusion weighted images (DWI) were pre-processed using SPM8 (www.fil.ion.ucl.ac.uk/spm). To reduce motion artifacts and eddy current induced distortions an affine transformation between the DWIs based on normalized mutual information was applied. The voxel displacement map (VDM) was calculated based on the field map resliced to DWI resolution using the field map toolbox of SPM8 [16]. The VDM was applied to minimize geometric distortions due to susceptibility artifacts. Finally the DWIs were aligned and resliced with affine matrix to a T1 weighted MRI using 7th degree B-spline interpolation [10]. The 61 non-collinear diffusion weighting gradient directions were updated using the same rotations and transformations as the resliced

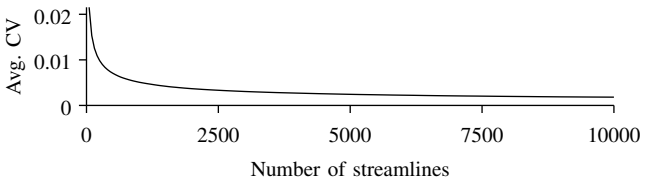

Fig. 2. Average voxel-wise coefficient of variation (CV) as function of number of streamlines in the tractography. The CV and the SNR are based on tractography results repeated five times for each number of streamlines.

images [18]. Segmentation of the white and gray matter was performed based on the high resolution structural T1w images using Freesurfer (surfer.nmr.mgh.harvard.edu) [6], [12], [11]. The Freesurfer reconstruction outputs, among others, the white matter segmentation and the gray-white matter boundary. The gray-white matter boundary for both hemispheres was converted to volumes and transformed from Freesurfer conformed space to native space. Likewise, the white matter segmentation from the Freesurfer reconstruction was transformed to native space. The diffusion parameters were estimated using FSL's BedpostX and probabilistic tractography was performed using FSL's Probtrackx2 with the omatrix3 option [4]. The transformed white matter volume was used as seed in the tractography and the transformed cortex labels as both target and stop mask in the tractography. For all other options the default settings were used. The cortex to cortex connectivity graph were output from FSL's probtrackx2 using the omatrix3 option. We obtained four 167,635×167,635 connectivity graphs (i.e., scan and rescan for 1000 and 5000 streamlines per seed voxel). Each link in the graphs took on the value of the number of streamlines connecting the two voxels in the target mask (gray/white matter boundary). The graphs were symmetrized and binarized (i.e., for each graph the graph and its transpose were added together and entries that were subsequently above zero set to one).

IV. Experiments and Results

A. Number of streamlines

To ensure that the network obtained by tractography is robust, probabilistic tractography was performed with different number of streamlines: Between 50 and 10,000 streamlines per seed voxel were used. Each number of streamlines was run five times. The voxel-wise coefficient of variation (CV) between voxels within the seed mask in the images with equal number of streamlines was calculated as $\mathrm{CV} = \frac{\sigma}{\mu}$, where σ is the standard deviation and μ is the mean. The average CV across all voxels was calculated [9] and is shown as function of number of streamlines in Fig. 2. The number of streamlines used in the subsequent experiments was selected on the basis of the average CV: As the average CV seems to have reached a stable level when using 1000 streamlines, and definitely when using 5000 streamlines (Fig. 2) we compare these two values.

B. Model parameters, inference, and convergence

For each network we performed 10 separate runs, all with the hyper parameters $\beta^+ = \beta^- = 1$ and $\alpha = \lfloor \log(J) \rfloor$, where J is the total number of nodes. For each run, we performed 100 iterations of the following sampling procedure: Each iteration began with a complete Gibbs sweep over all nodes. It was then followed by the same number of split-merge operations as the current number of clusters. In each split-merge operation we

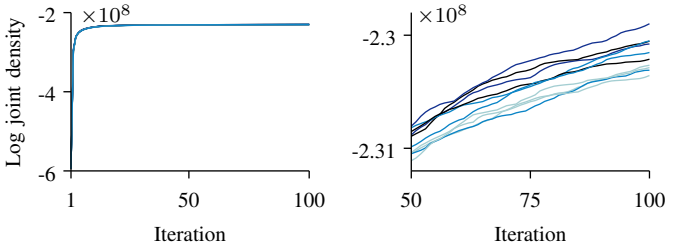

Fig. 3. Logarithm of the joint distribution for the MCMC inference procedure for the network based on 5000 streamlines. A zoom of the last 50 iterations is shown to the right.

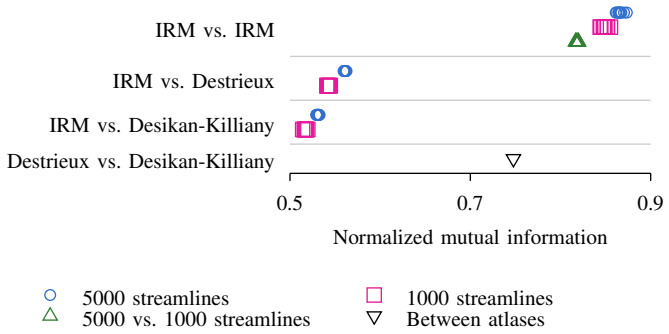

Fig. 4. Normalized Mutual Information (NMI) between 10 independent runs of the IRM and two atlases for network based on 1000 and 5000 streamlines.

performed 10 restricted Gibbs sweeps. Each of these iterations took several hours to compute. Fig. 3 shows the logarithm of the joint distribution for the different runs. It is clear that the MCMC sampler does not converge (which was also not to be expected [1]), but even when the sampler does not converge, the inferred grouping captures suboptimal but relevant structures in the network. In the following we used the inferred group structure after the last MCMC iteration.

C. Comparison and stability of estimated group structure

To compare the unsupervised groupings found by IRM with the groupings provided by the two atlases, we use the normalized mutual information (NMI) as a measure of similarity between 0 and 1. For two groupings $\boldsymbol{z}$ and $\boldsymbol{z}'$, we use: $\mathrm{NMI}(\boldsymbol{z}, \boldsymbol{z}') = \frac{2 \cdot \mathrm{I}(\boldsymbol{z}, \boldsymbol{z}')}{\mathrm{H}(\boldsymbol{z}) + \mathrm{H}(\boldsymbol{z}')}$ where $\mathrm{I}(\boldsymbol{z}, \boldsymbol{z}')$ is the mutual information between the groupings and $\mathrm{H}(\boldsymbol{z})$ is the entropy of $\boldsymbol{z}$. Fig. 4 shows NMI between all runs as well as between the runs and the two atlases. It is evident that the inferred groupings are very similar in the 10 runs as evidenced by the relatively high NMI, both within and between the networks based on 1000 and 5000 streamlines, respectively. Also, the inferred grouping is somewhat similar to the two atlases with an NMI score around 0.5-0.6.

D. Predictive performance

To assess how well the inferred structure fits the data, we use a second structural connectivity network based on a rescan of the same subject. Since any differences between the two scans are due to noise in the processes of generating the network, measuring how well we can predict the links in the second graph can be used to quantify the utility of the inferred structural units. To measure the predictive performance we use the area under the receiver operating characteristic curve (AUC) which allows us to compare predictions from the IRM

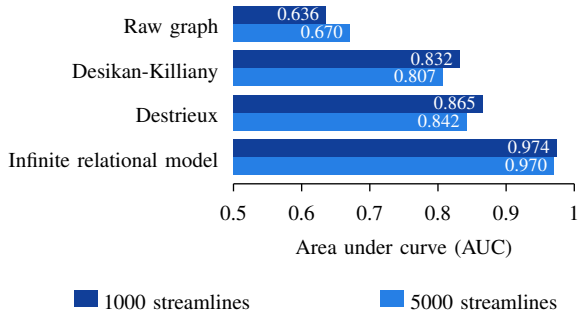

Fig. 5. Performance as measured by the area under the receiver operating characteristic curve (AUC) when predicting links in a network of structural connectivity based on data from a second scan of the same subject.

model with predictions made using other existing atlases or predicting directly from the raw network data. The results in Fig. 5 show that the IRM model outperforms predictions from the raw graph as well as both the Desikan-Killiany and Destrieux atlases both for networks based on 1000 and 5000 streamlines. However, when inspecting the extracted structural units (not shown) they were more diffuse compared to the atlases which may hamper their interpretation. This may be attributed both to the lack of convergence as well as lack of spatial constraints in the modeling.

V. Conclusion

When analyzing whole brain structural connectivity in full image resolution in the order of one thousand salient structural units were identified by our large scale implementation of the infinite relational model. The network based on 5000 streamlines had more structural units compared to the network based on 1000 streamlines. However, the estimated group structures were quite similar as quantified by NMI. Although the MCMC sampler did not reach convergence the identified groups were fairly robust to initialization while having some similarity to the Destrieux and Desikan-Killiany atlases. Notably, the extracted structural units provided significantly better predictive performances than predicting using the structural connectivity graph itself or the two considered atlases.

The present paper is to the best of our knowledge the first attempt at clustering structural connectivity in full resolution and provides a promising tool for a more detailed account of structural connectivity in general. In future work the influence of image resolution and choice of hyper-parameters should be investigated as should better sampling strategies.

Acknowledgment

This project was funded by the Lundbeck Foundation. We thank Andreas Leon Aagaard Moth for help with the large scale implementation.

References

[1] Kristoffer Jon Albers, Andreas Leon Aagard Moth, Morten Mørup, and Mikkel N. Schmidt. Large scale inference in the infinite relational model: Gibbs sampling is not enough. In *Machine Learning for Signal Processing (MLSP), 2013 IEEE International Workshop on*, pages 1–6. IEEE, 2013.

[2] D. Aldous. Exchangeability and related topics. *École d'Été de Probabilités de Saint-Flour XIII1983*, pages 1–198, 1985.

[3] Karen S Ambrosen, Tue Herlau, Tim Dyrby, Mikkel N Schmidt, and Morten Mørup. Comparing structural brain connectivity by the infinite relational model. In *Pattern Recognition in Neuroimaging (PRNI), 2013 International Workshop on*, pages 50–53. IEEE, 2013.

[4] TEJ Behrens, H Johansen Berg, Saad Jbabdi, MFS Rushworth, and MW Woolrich. Probabilistic diffusion tractography with multiple fibre orientations: What can we gain? *Neuroimage*, 34(1):144–155, 2007.

[5] TEJ Behrens, MW Woolrich, M Jenkinson, H Johansen-Berg, RG Nunes, S Clare, PM Matthews, JM Brady, and SM Smith. Characterization and propagation of uncertainty in diffusion-weighted mr imaging. *Magnetic resonance in medicine*, 50(5):1077–1088, 2003.

[6] Anders M Dale, Bruce Fischl, and Martin I Sereno. Cortical surface-based analysis: I. segmentation and surface reconstruction. *Neuroimage*, 9(2):179–194, 1999.

[7] Rahul S Desikan, Florent Ségonne, Bruce Fischl, Brian T Quinn, Bradford C Dickerson, Deborah Blacker, Randy L Buckner, Anders M Dale, R Paul Maguire, Bradley T Hyman, et al. An automated labeling system for subdividing the human cerebral cortex on mri scans into gyral based regions of interest. *Neuroimage*, 31(3):968–980, 2006.

[8] Christophe Destrieux, Bruce Fischl, Anders Dale, and Eric Halgren. Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. *Neuroimage*, 53(1):1–15, 2010.

[9] Annette J Dobson. *An introduction to generalized linear models*. CRC press, 2001.

[10] TB Dyrby, HM Lundell, MG Liptrot, MW Burke, M Ptito, and HR Siebner. Interpolation of dwi prior to dti reconstruction, and its validation. In *Proc. Intl. Soc. Mag. Reson. Med*, volume 19, 2011.

[11] Bruce Fischl, David H Salat, Evelina Busa, Marilyn Albert, Megan Dieterich, Christian Haselgrove, Andre van der Kouwe, Ron Killiany, David Kennedy, Shuna Klaveness, et al. Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neuron*, 33(3):341–355, 2002.

[12] Bruce Fischl, Martin I Sereno, and Anders M Dale. Cortical surface-based analysis: Ii: Inflation, flattening, and a surface-based coordinate system. *Neuroimage*, 9(2):195–207, 1999.

[13] Bruce Fischl, André van der Kouwe, Christophe Destrieux, Eric Halgren, Florent Ségonne, David H Salat, Evelina Busa, Larry J Seidman, Jill Goldstein, David Kennedy, et al. Automatically parcellating the human cerebral cortex. *Cerebral cortex*, 14(1):11–22, 2004.

[14] Patric Hagmann, Leila Cammoun, Xavier Gigandet, Reto Meuli, Christopher J Honey, Van J Wedeen, and Olaf Sporns. Mapping the structural core of human cerebral cortex. *PLoS biology*, 6(7):e159, 2008.

[15] S. Jain and Radford M Neal. A split-merge markov chain monte carlo procedure for the dirichlet process mixture model. *Journal of Computational and Graphical Statistics*, 13(1), 2004.

[16] Peter Jezzard and Robert S Balaban. Correction for geometric distortion in echo planar images from b0 field variations. *Magnetic resonance in medicine*, 34(1):65–73, 1995.

[17] C. Kemp, J. B. Tenenbaum, T. L. Griffiths, T. Yamada, and N. Ueda. Learning systems of concepts with an infinite relational model. In *AAAI*, volume 3, page 5, 2006.

[18] Alexander Leemans and Derek K Jones. The b-matrix must be rotated when correcting for subject motion in dti data. *Magnetic Resonance in Medicine*, 61(6):1336–1349, 2009.

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

[20] J. Pitman et al. Combinatorial stochastic processes. Technical report, Springer, 2002.

[21] TG Reese, O Heid, RM Weisskoff, and VJ Wedeen. Reduction of eddy-current-induced distortion in diffusion mri using a twice-refocused spin echo. *Magnetic Resonance in Medicine*, 49(1):177–182, 2003.

[22] Nina Linde Reislev, Ron Kupers, Hartwig R. Siebner, Maurice Ptito, and Tim B. Dyrby. Blindness differentially affects the integrity of the dorsal and ventral visual streams, 2013.

[23] Nina Linde Reislev, Maurice Ptito, Ron Kupers, Hartwig R. Siebner, and Tim B Dyrby. Alterations of the inferior longitudinal fasciculus in congenital and late blindness, 2012.

[24] Z. Xu, V. Tresp, K. Yu, and H.P. Kriegel. Learning infinite hidden relational models. *Uncertainty in Artificial Intelligence (UAI2006)*, 2006.