Bayesian Connectomics

A probabilistic perspective on brain networks

ISBN

978-94-6295-650-6

COVER DESIGN

Arief Hühn and Max Hinne

Bayesian Connectomics: a probabilistic perspective on brain networks

Bayesian Connectomics

A probabilistic perspective on brain networks

Proefschrift

ter verkrijging van de graad van doctor
aan de Radboud Universiteit Nijmegen
op gezag van de rector magnificus prof. dr. J.H.J.M. van Krieken,
volgens besluit van het college van decanen
in het openbaar te verdedigen op vrijdag 9 juni 2017
om 14.30 uur precies

door

Max Hinne

geboren op 16 oktober 1983
te Nijmegen

PROMOTOREN:

Prof. dr. Tom Heskes
Prof. dr. Peter W. M. Desain

COPROMOTOR:

Dr. Marcel A. J. van Gerven

MANUSCRIPTCOMMISSIE:

Prof. dr. Paul H. E. Tiesinga
Prof. dr. Mark W. Woolrich (University of Oxford, Verenigd Koninkrijk)
Dr. Lourens J. W. Waldorp (Universiteit van Amsterdam)

[There is] no duality between the existence of a thing and its properties: properties are all there is. Indeed: there are no things.

— Piet Hut, paraphrasing David Mermin

CONTENTS

1

INTRODUCTION

The human brain is arguably one of the most interesting systems around. Its sheer complexity is awe-inspiring, as it consists of roughly 86 billion neurons [Herculano-Houzel, 2009] and 164 trillion synapses [Tang et al., 2001], giving rise to a number of unique configurations that baffles imagination[1] [Churchland, 2000]. Impressive numbers aside, the brain houses the intricate computations that give rise to our cognition and behavior [Churchland and Sejnowski, 1994]. Small wonder that researchers have tried to make sense of this organ since ancient times [Finger, 2001; Glickstein, 2014]. This endeavor has gained a lot of traction in the latter half of the previous century, as the development of neuroimaging techniques such as functional magnetic resonance imaging (fMRI) greatly improved the ability to study the brain of subjects *in vivo* [Ogawa et al., 1990].

However, correctly interpreting neuroimaging data is far from trivial. Each neuroimaging technique has its own assumptions and weaknesses [Jones et al., 2013; Haller and Bartsch, 2009] and advanced statistical techniques are needed to extract the signal of interest (representing for example the neuronal activity in a particular area) from the noise [Bowman et al., 2007; Bowman, 2014]. This thesis aims to contribute to this task, in particular to the subdiscipline of neuroscience that is concerned with networks of connected brain regions. I argue that brain connectivity studies frequently place too much confidence in neuroimaging techniques and fail to acknowledge the uncertainty associated with their estimated networks and analyses thereof. I investigate whether a Bayesian treatment of brain network estimation offers a solution to this issue.

In the remainder of this introduction, I first describe the necessary background regarding brain networks. I proceed by describing Bayesian generative modeling, which forms the central methodology used throughout this dissertation. In addition, I explain how Bayesian modeling may be applied to explicitly characterize the uncertainty in brain network estimates. Lastly, I provide the outline of this thesis.

[1] It was estimated at $10^{10^{14}}$ [Churchland, 2000]. For comparison, the number of elementary particles in the universe is roughly 10^{86}. Note however that this comparison is not entirely fair, as it is between a number of apples and a number of unique states of combinations of oranges. Still, the reader will have understood that the number of unique configurations of connected neurons is immense.

1.1 BRAIN NETWORKS AND CONNECTOMICS

A crucial insight in modern neuroscience is that the *interaction* between brain structures is as important as the activity in the individual regions themselves [Catani et al., 2012b]. In itself, this realization is far from novel, as these ideas have been around since the work by René Descartes in the 17th century. He realized that white matter did not merely function as scaffolding for the ventricles, but contained many fibres connecting different areas of the brain [Descartes, 1662; Catani et al., 2013]. However, only recently technological advancements in imaging techniques and computing have made it possible to obtain and process huge collections of high-quality data that allow us to study the connections of the brain *in vivo* on a large scale [Lichtman et al., 2014]. Consequently, a new avenue of research opened up, and a substantial part of the neuroscience community has shifted its attention away from studying individual areas and towards the study of *networks* of such regions. This shift has spawned a novel (sub-)discipline in neuroscience, which is devoted to estimating and analyzing connections in the brain, and bears the name *connectomics*[2] [Sporns et al., 2005; Hagmann, 2005].

Connectivity studies come in three flavors. The first is *structural connectivity* (also known as anatomical connectivity), which is concerned with the mapping of white-matter fiber bundles that connect spatially segregated brain regions [Gong et al., 2009; Hagmann et al., 2007]. The second type is *functional connectivity*, which tries to identify the statistical dependencies between the activation patterns of different regions [Friston, 2011; van den Heuvel and Hulshoff Pol, 2010]. Finally, *effective connectivity* goes one step further and not only finds statistical dependence, but also aims to find the causal relationships between brain regions. That is, it describes how activity in one region causes another region to behave in a certain way [Friston, 2011; McIntosh, 2010; Stephan and Roebroeck, 2012]. Structural, functional and effective connectivity are of course related, as causation implies a statistical dependency [Robinson et al., 2014; Iyer et al., 2013], and regions can only be statistically dependent if there is an anatomical pathway through which this dependency can be mediated [Messé et al., 2015c; Honey et al., 2009].

Structural, functional and effective connectivity represent connections in the brain using networks, consisting of nodes and edges. Depending on the scale of the study, the nodes of the network may reflect anything from up to large regions of interest down to individual neurons. The edges of the network indicate the connections between the nodes and may simply be on or off, have weights indicating connection strength and, in the case of effective connectivity, may contain directionality. As networks estimated from neuroimaging data are often large (ranging from roughly one hundred to many thousands of nodes), a visual representation usually shows little more than a tangled mess. Tools from network analysis and graph theory may be used as summary statistics so that properties of the network as a whole may be represented by a few numbers [Bullmore and Sporns, 2009; Rubinov and Sporns, 2010; Sporns, 2014].

A prominent example of such a statistic that is frequently attributed to brain networks is *small-worldness* [Watts and Strogatz, 1998], which I will explain here as an illustration of possible analyses: For a network to be a small-world network, it has to fulfill two criteria simul-

[2] After the interest in genomics, it has become fashionable to have your own -omics field.

taneously. First, the network has to show a high degree of clustering. This can intuitively be understood using the phrase 'I am friends with friends of my friends', which in network terminology means that many connected pairs of nodes in the network form a triangle with a third node. The second criterion is that from any node in the network, it should take only a few 'hops' of intermediate nodes to reach any other node. This too has an intuitive interpretation known as the 'six degrees of separation', referring to the conjecture that anyone on the world is only six handshakes (i.e. edges) away from any other person [Milgram, 1967]. When a network sufficiently meets both criteria, it is considered a small-world network. Small-worldness allows one to quantify for example the efficiency of information processing of a network. Because of the short path length and clustering, information from any node in the network can easily reach any other node, which is not the case for many other network architectures. Many more network properties exist next to small-worldness, but I will not go into detail regarding them here. The interested reader is referred to [Newman, 2010] for more on this topic.

Network analysis is often used as a tool to distinguish between the network(s) found within healthy subjects and those of patients [Fornito and Bullmore, 2012, 2014]. For example, network analysis of the connectome has proven instrumental in study of the brain in patients with Alzheimer's disease. It has been reported that networks of patients with Alzheimer's disease show a decrease in small-worldness, due to an increasing path length (due to damage in fibers longer routes are needed for two regions to communicate) [He et al., 2008; Yao et al., 2010; Lo et al., 2010]. For many other pathologies, such as multiple sclerosis, schizophrenia and epilepsy, as well as for cognitive ability, similar studies have been conducted that show correlations between behavioral changes and properties of brain networks [He et al., 2009; Bai et al., 2012; Griffa et al., 2013; Baronchelli et al., 2013; Catani, 2007; Craddock et al., 2013; Fornito and Bullmore, 2012; Smith et al., 2015]. An important consequence of these findings is that brain network analysis may serve to create biomarkers that can be used for early diagnosis, before behavioral deficits become apparent [Castellanos et al., 2013], as well as for patient stratification [Deco and Kringelbach, 2014] and prognosis [Fornito et al., 2015].

1.2 Uncertainty and Bayesian modeling

The previous paragraph underscores the importance of connectomics as a research discipline. However, one should realize that there is a large gap between acquired data and the actual wiring of the brain. Several elaborate preprocessing steps are needed to transform the raw data into something that can represent connectivity, but despite our best efforts, technological limitations and indirect measurement will at best result in a noisy representation of the underlying connectivity [Lichtman et al., 2014; Akil et al., 2011; Schultz et al., 2014; Jbabdi et al., 2015; Craddock et al., 2013; Tournier et al., 2011]. As a consequence, the networks that we analyze are likely to contain some degree of error. Just as a blurred picture leaves the viewer with multiple possibilities of what is being portrayed, neuroimaging techniques leave us at best with a distorted image of the true connectivity. This implies that we could put forward different networks as the true network based on the same data. In short, there is *uncertainty* about what the actual connectivity is, given the noisy data at hand.

The goal of this dissertation is to make the uncertainty associated with connectivity estimates explicit. Instead of postulating that a certain network reflects the true connectivity, I propose a methodology to describe all networks supported by the combination of data and our prior assumptions. In addition, this methodology provides for each network the *probability* that indicates how likely a network is, so that the confidence in an estimate may be quantified. To accomplish this goal, I propose to use Bayesian statistics [Laplace, 1814; Jeffreys, 1961; Gelman et al., 2013], as this framework allows us to specifically reason with probabilities and uncertainty. In a Bayesian model, the interactions between stochastic variables of interest (i.e. the parameters of a network and the data you observe) are described. It consists particularly of the following two ingredients:

1. *The likelihood function*. The likelihood function describes how parameters generate the data we observe[3]. For instance, if a particular fiber exists, then this will generate the white matter we observe in MRI. Here, the (presence of the) fiber is the parameter of interest, while the measurement of white matter acts as data.

2. *The prior*. The prior distribution captures the beliefs we had about the parameters before observing any data. For instance, before any data is observed we might expect to find a certain number of connections. The prior serves as an *explicit* representation of the assumptions of the modeler.

In Bayesian modeling, both the likelihood function and the prior are described in terms of *probability distributions*. The former may be written as $P(x \mid \phi)$, in which x represents the observed variables (i.e. the data, such as fMRI BOLD response) and ϕ the parameter of interest (e.g. the covariance structure between brain activity in different regions). The latter, $P(\phi)$, is the distribution of the parameters in absence of any observed data. The functional form of either of these probability density functions depends on what is being modeled and the decisions made by the modeler. For example, the likelihood function could be a Gaussian distribution with mean ϕ, and the prior could be a uniform distribution in the unit interval.

By using Bayes' theorem (named after the reverend Thomas Bayes, 1701–1761), we can use these two components to compute the *posterior* distribution[4] that combines both our a priori knowledge as well as the observations:

$$P(\phi \mid x) \propto P(x \mid \phi)P(\phi) \quad . \tag{1.1}$$

This fundamental equation describes our beliefs about ϕ, once data x is observed.

It deserves re-iterating that this equation describes the distribution of all possible parameter values (here: all possible values for ϕ), rather than the probability of a single instance. This explains the computational difficulty of this seemingly trivial equation. Even for relatively small

[3] Note that the term 'generate' should in this context be interpreted as 'generates observations with certain statistical regularities', rather than a description of e.g. fiber growth – this is a statistical description, not necessarily a mechanistic one.

[4] Actually, we usually compute the posterior up to the constant term $P(x)$, known as the marginal likelihood or model evidence. When we are interested in the distribution over the parameters, this term is not required as it is constant for all ϕ. This is convenient, as the model evidence is typically difficult to compute for any but the most trivial models.

networks, the state space of ϕ (i.e. all the possible networks for a certain set of regions) is exceedingly large or may even be infinite if ϕ is continuous. Because of this, exact computation of $P(\phi \mid x)$ is intractable in almost all nontrivial models. Instead, the posterior distribution may be approximated using a technique known as Markov chain Monte Carlo (MCMC) [Andrieu et al., 2003; Gilks et al., 1996]. As the mathematical details of MCMC are outside of the scope of this introduction (they will be discussed in the remaining chapters where appropriate), its core ideas may instead be explained using the following analogy: Imagine a probability distribution as mountainous terrain, with mountains and valleys corresponding to areas of high and low probability, respectively. We are tasked with drawing an approximation of the elevation map of this terrain. However, there is one problem: a dense fog surrounds us so we cannot see anything except our current position. Furthermore, we can sense the difference in height between a proposed, adjacent position and our current one. By cleverly deciding when to step in a randomly proposed direction and when to stay at our current position, we are ensured to have stayed at all of the terrain for a duration proportional to the probability of that area. We will have visited high mountains frequently and only occasionally we went through low terrain, perhaps to visit another mountain on the other side.

For well-behaved probability distributions, the MCMC strategy sketched here (known as Metropolis-Hastings MCMC) suffices. But when models become more complicated, for example when the terrain spans multiple planets, or contains multiple spike-like mountains with almost infinitely deep ravines separating them, more advanced methods are required. A number of these, such as split-merge sampling [Richardson and Green, 1997] (which allows us to make large jumps between hills far apart) or reversible-jump MCMC [Green, 1995], (used when each position in the landscape itself contains a probability density) are discussed and applied throughout this thesis.

Bayesian connectomics

In the current context, equation (1.1) describes the distribution over connectivity given data. The parameter ϕ represents a brain network, such as the functional connectivity between regions in a predefined atlas. The observations x that inform us about ϕ could, for example, be the fMRI BOLD signal correlation between pairs of regions. The (approximated) posterior distribution of ϕ given x describes how likely each of the possible networks is after having seen the data and taking into account our prior beliefs. Whereas alternative threshold-based approaches would consider a pair of regions connected if their pairwise correlation exceeds a certain limit, a Bayesian approach yields the posterior probability of this connection being present as well as, in the case of weighted networks, the credible interval of the connection weight. I refer to this probabilistic take on brain networks as *Bayesian connectomics*, or BaCon for short. A schematic toy example is shown in Figure 1.1.

Note that any subsequent network analysis that could be performed on a connectivity point estimate, may also be applied to (each element of) the probability distribution. This then provides for instance the credible interval of the small-world index, rather than a single value. Most

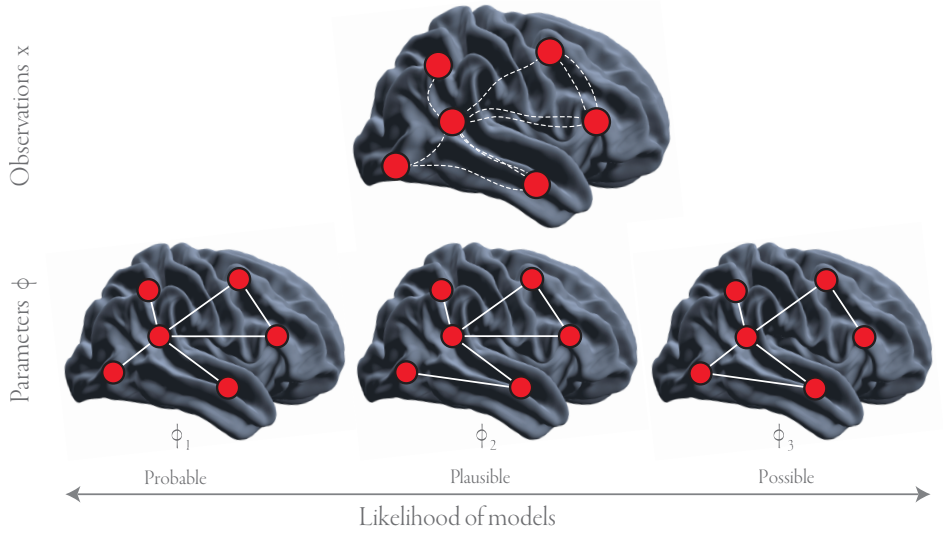


FIGURE 1.1: Schematic toy example of Bayesian connectomics. The top row shows a network of six nodes, representing six regions of interest in a brain. Measurements of connectivity are taken and indicated as dashed lines. This provides an indication, but no absolute certainty, as to which regions are connected. The bottom row shows a potential posterior distribution of connectivity given these samples, i.e. $P(\phi \mid x)$. Three different models are supported by these data. The first, ϕ_1, is the most likely, but ϕ_2 and ϕ_3 could also be the true network, albeit with a lower probability.

importantly, if such a credible interval spans a large range of values, our estimates are uncertain and should be handled with care.

1.3 THESIS OUTLINE

The thesis is structured as follows. In Chapter 2, a generative model is postulated that describes how structural connectivity may be derived from probabilistic measurements of possible fiber trajectories. This is combined with either a prior on the density of networks, or a hierarchical prior through which different subjects become conditionally dependent. In order to validate the model, the resulting estimates of structural connectivity are compared with the conditional (in)dependence structure found in resting-state functional connectivity, which has been collected for the same subjects.

Chapter 3 focuses on a different prior distribution for structural connectivity, known as the 'Chinese restaurant process', in which I explicitly capture the belief that structural connectivity should show clustering. By approximating the joint distribution of connectivity and clustering, both are inferred at once. The approach allows for probabilistic estimates of clustered connectivity, which may be used to define a parcellation.

Subsequently, the focus is shifted towards functional connectivity. Functional connectivity may be quantified using partial correlations, but finding the matrix of partial correlations

is difficult in practice as the results may be unstable if there is not enough collected data. In Chapter 4 I propose to use the G-Wishart distribution [Atay-Kayis and Massam, 2005; Dobra and Lenkoski, 2011] as a Bayesian alternative to regularization approaches that are typically employed when estimating partial correlation. The G-Wishart distribution restricts the search space by only allowing a functional connection between nodes that are connected in a provided network G. Here, G is constructed using structural connectivity, so that functional connectivity becomes constrained by an anatomical prior.

Chapter 5 continues where the previous chapter left off. Again, the G-Wishart distribution is used to express that functional connectivity is constrained by conditional (in)dependencies. This time however, I approximate the *joint* posterior of functional connectivity and conditional (in)dependence, i.e. $P(G, \mathbf{R} \mid \mathbf{X})$ instead of $P(\mathbf{R} \mid \mathbf{X}, G)$. As the G-Wishart distribution is doubly intractable, I propose an improved reversible-jump algorithm [Green, 1995] to approximate it. Although the algorithm is applied in order to estimate functional connectivity, it should be noted that the proposed methodology is generic and suitable for all applications of Gaussian graphical models.

Lastly, Chapter 6 provides an extensive comparison of our unconstrained functional connectivity model with the graphical lasso [Friedman et al., 2008]. The basis of this comparison is formed by a series of simulations of fMRI BOLD response signals [Smith et al., 2011]. In addition, the chapter introduces a Bayesian approach to data fusion, combining the likelihood functions for structural and functional connectivity into a single model that allows estimation of connectivity informed by two modalities.

The thesis concludes in Chapter 7 with a summary of the presented work and its main contributions. Furthermore, the limitations of the proposed approaches are discussed and a number of suggestions for further research directions are provided.

2

BAYESIAN INFERENCE OF STRUCTURAL CONNECTIVITY

Human behavior ultimately arises through the interactions between multiple brain regions that together form networks that can be characterized in terms of structural, functional and effective connectivity [Penny et al., 2006]. Structural connectivity presupposes the existence of white-matter tracts that connect spatially segregated brain regions which constrain the functional and effective connectivity between these regions. Hence, structural connectivity provides the scaffolding that is required to shape neuronal dynamics. Changes in structural brain networks have been related to various neurological disorders. For this reason, optimal inference of structural brain networks is of major importance in clinical neuroscience [Catani, 2007]. Inference of these networks entails two steps. First is the estimation of the white matter tracts. The second step consists of obtaining the network that captures which regions are connected, based on the earlier identified fibre tracts. In this chapter, we focus on the latter step.

For the first step, we use diffusion weighted imaging (DWI), which is a prominent way to estimate structural connectivity of whole-brain networks in vivo. It is a variant of magnetic resonance imaging (MRI) which measures the restricted diffusion of water molecules, thereby providing an indirect measure of the presence and orientation of white-matter tracts. By following the principal diffusion direction in individual voxels, streamlines can be drawn that represent the structure of fibre bundles, connecting separate regions of grey matter. This process is known as deterministic tractography [Conturo et al., 1999; Chung et al., 2010; Shu et al., 2011]. Alternatively, fibres may be estimated using probabilistic tractography [Behrens et al., 2003a, 2007; Friman et al., 2006; Jbabdi et al., 2007]. This comprises a model for the principal diffusion direction that is then used to sample distributions of streamlines. Ultimately, the procedure results in a measure of uncertainty about where a hypothesized connection will terminate. A benefit of the probabilistic approach is that it explicitly takes uncertainty in the streamlining process into account.

This chapter is based on: M. Hinne, T. Heskes, C.F. Beckmann and M.A.J van Gerven, 2013. "Bayesian inference of structural brain networks." *NeuroImage* 66, pp. 543–552.

Apart from studies focusing on particular tracts, much research has been devoted to the derivation of macroscopic connectivity properties, that is, whole-brain structural connectivity. Several approaches have been suggested to extract whole-brain networks from probabilistic tractography results [Robinson et al., 2008; Hagmann et al., 2007; Gong et al., 2009]. Unfortunately, inference of whole-brain networks from probabilistic tractography estimates remains somewhat ad hoc. Typically the underlying brain network is derived by thresholding the streamline distribution such that counts above or below threshold are taken to reflect the presence or absence of tracts, respectively. This approach is easy to implement but it has a number of issues. First, the threshold is arbitrarily chosen to have a particular value. In a substantial part of the literature, the threshold that is used to transform the streamline distribution into a network is actually set to zero [Hagmann et al., 2007, 2008; Zalesky et al., 2010; Vaessen et al., 2010; Chung et al., 2011]. However, probabilistic streamlining depends on the arbitrary number of samples that are drawn per voxel. This implies that, as more samples are drawn, more brain regions are likely to eventually become connected given a threshold at zero. Alternatively, the number of streamlines can be interpreted as connection weight [Bassett et al., 2011; Zalesky et al., 2010; Robinson et al., 2010], or a relative threshold can be applied [Kaden et al., 2007]. This way, the relative differences between connections remain respected. Unfortunately, the connection weights do not have a straightforward (probabilistic) interpretation. Simply normalizing these weights does not yield a true notion of connection probability. At most, it can be regarded as the conditional probability that a streamline ends in a particular voxel given the starting point of the streamline. In the case of a streamline distribution with, say, half of the streamlines starting at node A ending in node B, and the other half ending in node C, normalized streamline counts cannot distinguish between one edge with an uncertain end point, or two edges with definite end points. Finally, several graph-theoretical measures such as characteristic path length and clustering coefficient are ill-defined for non-binary networks.

In general, it is problematic to use thresholding since it ignores the relative differences between streamline counts. Intuitively, one would expect that if, say, ninety percent of the streamlines connect from voxel A to voxel B, and ten percent connect voxel A to voxel C, then at the least the former has a higher probability of having a corresponding edge in the network than the latter, but both edges are possible as well. This is related to the *burstiness* phenomenon of words in document retrieval, where the occurrence of a rare word in a document makes its repeated occurrence more likely [Xu and Akella, 2010]. Summarizing, the issue with thresholding approaches is that they consider each tract in isolation. This ignores the information that can be gained from the possible symmetry in streamline counts, as well as from the relative differences within a streamline distribution.

Another important observation is that the mentioned approaches do not easily support the integration of probabilistic streamlining data with other sources of information. Data is often not collected in isolation but rather acquired for multiple subjects, potentially using a multitude of imaging techniques. Multi-modal data fusion is needed in order to provide a coherent picture of brain function [Horwitz and Poeppel, 2002; Groves et al., 2011]. The integration of multi-subject data is required for group-level inference, where the interest is in estimating a network

that characterizes a particular population, for example, when comparing patients with controls in a clinical setting [Simpson et al., 2011].

In the following, we provide a Bayesian framework for the inference of whole-brain networks from streamline distributions. In our approach, we consider the distribution of (binary) networks that are supported by our data, instead of generating a single network based on an arbitrary threshold. Our approach relies on defining a generative model for whole-brain networks which extends recent work in systems biology [Mukherjee and Speed, 2008] and consists of two ingredients. First, the classical Erdős-Rényi model [Erdős and Rényi, 1960] is used to define a network prior. This prior is later extended to handle multi-subject data, capturing the notion that different subjects' brains tend to be similar. Second, we propose a forward model based on a Dirichlet compound multinomial distribution which views the streamline distributions produced by probabilistic tractography as noisy data, thus completing the generative model.

In order to validate our Bayesian framework we make use of the often reported observation that resting-state functional connectivity reflects structural connectivity [Koch et al., 2002; Greicius et al., 2009; Honey et al., 2009; Lv et al., 2010; Skudlarski et al., 2008; Park et al., 2008; Damoiseaux and Greicius, 2009]. We show that structural networks that derive from our generative model informed by the connectivity for other subjects provide a better fit to the (in)dependencies in resting-state functional MRI (rs-fMRI) data than the standard thresholding approach.

2.1 MATERIAL AND METHODS

Data acquisition

Twenty healthy volunteers were scanned after giving informed written consent in accordance with the guidelines of the local ethics committee. A T1 structural scan as well as resting-state functional data and diffusion-weighted images were obtained using a Siemens Magnetom Trio 3T system at the Donders Centre for Cognitive Neuroimaging, Radboud University Nijmegen, The Netherlands. The rs-fMRI data were acquired at 3 Tesla using a multi echo-echo planar imaging (ME-EPI) sequence (voxel size 3.5 mm isotropic, matrix size 64×64, TR = 2000 ms, TEs = 6.9, 16.2, 25, 35 and 45 ms, 39 slices, GRAPPA factor 3, 6/8 partial Fourier). A total of 1030 volumes were obtained. An optimized acquisition order described by Cook et al. [2006] was used in the DWI protocol (voxel size 2.0 mm isotropic, matrix size 110×110, TR = 13000 ms, TE = 101 ms, 70 slices, 256 directions at b = 1500 s/mm^2 and 24 directions at b=0).

Preprocessing of resting-state data

The multi-echo images obtained using the rs-fMRI acquisition protocol were combined using a custom Matlab script (MATLAB 7.7, The MathWorks Inc., Natick, MA, USA) which implements the procedure described by Poser et al. (2006) and also incorporates motion correction using functions from the SPM5 software package (Wellcome Department of Imaging Neuro-

science, University College London, UK). Of the 1030 combined volumes, the first six were discarded to allow the system to reach a steady state. Tools from the Oxford FMRIB Software Library (FSL, FMRIB, Oxford, UK) were used for further processing. Brain extraction was performed using FSL BET [Smith, 2002]. For each subject, probabilistic brain tissue maps were obtained using FSL FAST [Zhang et al., 2001]. A zero-lag 6th order Butterworth bandpass filter was applied to the blood-oxygen-level dependent (BOLD) response timeseries to retain only frequencies between 0.01 and 0.08 Hz. After preprocessing, the fMRI data were parcellated according to the automated anatomical labeling (AAL) atlas [Tzourio-Mazoyer et al., 2002]. Regions without voxels with gray-matter probability $\geqslant 0.5$ were discarded. This resulted in an average region count of 115.7 ± 0.1. For these regions the functional data was summed and then standardized to have zero mean and unit standard deviation. The resulting data were used to compute the empirical covariance matrix $\hat{\mathbf{C}}$.

Preprocessing of diffusion imaging data

The preprocessing steps for the diffusion data were conducted using FSL FDT tools [Behrens et al., 2003a] and consisted of correction for eddy currents and estimation of the diffusion parameters. To obtain a measure of white-matter connectivity, FDT Probtrackx 2.0 [Behrens et al., 2003a, 2007] was used. As seed voxels for tractography we used those voxels that live on the boundary between white matter and gray matter. For each of these voxels 5000 streamlines were drawn, with a maximum length of 2000 steps. The streamlines were restricted by the voxel fractional anisotropy to prevent them from wandering around in gray matter. Streamlines in which a sharp angle (>80 degrees) occurred or that had a length less than 2 mm were discarded. The output thus obtained is a matrix $\mathbf{S}$ with s_{ij} the number of streamlines drawn from voxel i to voxel j. To transform this into the parcellated scheme as dictated by the AAL atlas, the streamlines were summed over all voxels per region, resulting in an aggregated connectivity matrix which ranges over regions instead of voxels. Regions that had been removed after preprocessing the fMRI data were removed from the aggregated connectivity matrix as well.

Framework for structural connectivity estimation

In this section we derive our Bayesian approach to the inference of whole-brain structural networks. The quantity of interest in our framework is the posterior over structural networks represented by the adjacency matrix $\mathbf{G}$ given observed probabilistic streamlining data $\mathbf{S}$ and hyperparameters ζ. An element $g_{ij} \in \{0, 1\}$ represents the absence or presence of an edge between brain region i and j. $\mathbf{G}$ is taken to be symmetric, undirected, and without self-loops, such that $g_{ij} = g_{ji}$ and $g_{ii} = 0$. A brain region can either be interpreted as a voxel or as an aggregation of voxels as defined by a gray matter parcellation. The posterior expresses our knowledge on structural connectivity given the data and background knowledge and is given by:

$$P(\mathbf{G} \mid \mathbf{S}, \zeta) \propto P(\mathbf{S} \mid \mathbf{G}, \delta_1, \delta_0) P(\mathbf{G} \mid \theta) \ , \tag{2.1}$$

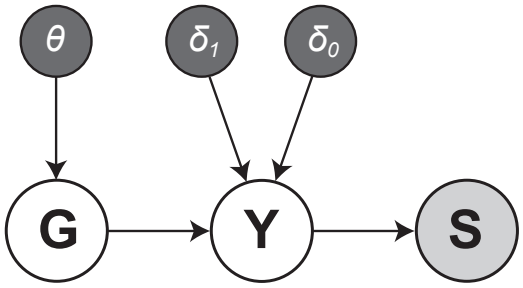


FIGURE 2.1: The generative model that describes how the observed streamline distribution $\mathbf{S}$ depends on the (hidden) connectivity probabilities $\mathbf{Y}$. These in turn depend on the hyperparameters δ_1 and δ_0 as well as the connectivity $\mathbf{G}$, which is determined by hyperparameter θ of the prior.

with hyperparameters $\zeta = (\delta_1, \delta_0, \theta)$, for which an interpretation will be given later on. In the following, for convenience, we will sometimes suppress the dependence on the hyperparameters.

To infer the posterior distribution, we must specify a prior $P(\mathbf{G} \mid \theta)$ and a forward model $P(\mathbf{S} \mid \mathbf{G})$ which together define a generative model of probabilistic streamlining data. Given these components, the posterior can be approximated using a Markov chain Monte Carlo algorithm, as described in detail in Section 2.1. We now proceed to formally define the components of the generative model as shown in Figure 2.1.

Forward model

We begin with a specification of the forward model $P(\mathbf{S} \mid \mathbf{G})$. Here, we describe how the observed streamline distributions $\mathbf{S}$ depend on the underlying network $\mathbf{G}$ through latent streamline probabilities $\mathbf{Y}$.

Assume there are p brain regions for which we want to estimate the structural connectivity. We start by considering one region i and the possible targets in which a postulated tract may terminate. Let s_{ij} denote the number of streamlines which start in region i and terminate in region j. We assume that $s_{ii} = 0$. Probabilistic tractography produces a distribution over target vertices $\mathbf{s}_i = (s_{i1}, \ldots, s_{ip})^T$ by drawing L streamlines, $S_i = \sum_{j=1}^{p} n_{ij} \leqslant L$ of them ending up in a target region.[1] A particular distribution $\mathbf{s}_i$ depends on the streamline probabilities. That is, we expect many streamlines between two regions when there is a high streamline probability and vice versa. This is captured by expressing the probability of a distribution $\mathbf{s}_i$ in terms of a multinomial distribution

$$\mathbf{s}_i \mid \mathbf{y}_i \quad \sim \quad \text{Multinomial}(\mathbf{y}_i) \ , \tag{2.2}$$

in which $\mathbf{y}_i = (y_{i1}, \ldots, y_{ip})$ is a probability vector with $\sum_j y_{ij} = 1$. Each y_{ij} represents the probability of drawing a streamline from region i to region j. This streamlining probability

[1] It is possible that streamlines end up in voxels outside any region of the parcellation, hence the inequality.

itself depends on whether or not there actually exists a physical tract between region i and region j.

Let $\mathbf{g}_i$ denote the i-th row of $\mathbf{G}$ indicating the connectivity between region i and all other regions. Intuitively, we expect a high streamline probability when there is an edge in the network. Conversely, we expect a low probability when two regions are disconnected. Thus, the streamline probabilities depend on the actual white-matter connectivity as modeled by $\mathbf{G}$. This is captured by modeling the distribution of streamline probabilities using a Dirichlet distribution

$$\mathbf{y}_i \,|\, \mathbf{G}, \delta_1, \delta_0 \quad \sim \quad \text{Dirichlet}(\mathbf{b}_i) \tag{2.3}$$

where shorthand notation $\mathbf{b}_i = (b_{i1}, \ldots, b_{ip})$ with $b_{ij} \equiv g_{ij}\delta_1 + (1 - g_{ij})\delta_0$ is used. The term b_{ij} can be interpreted as the parameters that determine the probability of streamlining from region i to region j when an edge g_{ij} is either present (δ_1) or absent (δ_0).

To obtain a single expression for the likelihood of an adjacency matrix, let $\mathbf{S} = (\mathbf{s}_1; \ldots; \mathbf{s}_p)$ represent the combined probabilistic tractography data, i.e. for each of the p nodes a distribution of streamlines to all other nodes. Similarly, let $\mathbf{Y} = (\mathbf{y}_1; \ldots; \mathbf{y}_p)$ denote the combined hidden connection probabilities and lastly $\mathbf{G} = (\mathbf{g}_1; \ldots; \mathbf{g}_p)$ the adjacency matrix for all brain regions. The likelihood of the network $\mathbf{G}$ is expressed as

$$P(\mathbf{S} \,|\, \mathbf{G}, \delta_1, \delta_0) = \int P(\mathbf{S} \,|\, \mathbf{Y})\, P(\mathbf{Y} \,|\, \mathbf{G}, \delta_1, \delta_0)\, d\mathbf{Y}. \tag{2.4}$$

By recognizing that the Dirichlet distribution is the conjugate prior for the multinomial distribution, it follows that (2.4) is a product of Dirichlet compound multinomial distributions [Madsen et al., 2005; Xu and Akella, 2010; Minka, 2000]. The DCM distribution assumes that, given a network, a probability vector can be drawn with large values where the network has edges and small values where the network is disconnected. This probability vector, in turn, can be used to sample from a multinomial distribution that reflects the probabilistic tractography outcome.

For sufficiently small choices of the hyperparameters of the DCM, sampling from this multinomial reflects the burstiness behavior we observe in the streamline distributions, where some pairs of nodes are connected by many streamlines, while most pairs have few or even zero streamlines.

Network prior

In order to define a prior on adjacency matrices, we adopt the Erdős-Rényi model which states that the probability of an edge between region i and j is given by parameter θ [Erdős and Rényi, 1960] and that the individual edges are *a priori* independent. This allows the prior to be expressed in terms of a product of Bernoulli distributions:

$$g_{ij} \,|\, \theta \quad \sim \quad \text{Bernoulli}(\theta) \ , \tag{2.5}$$

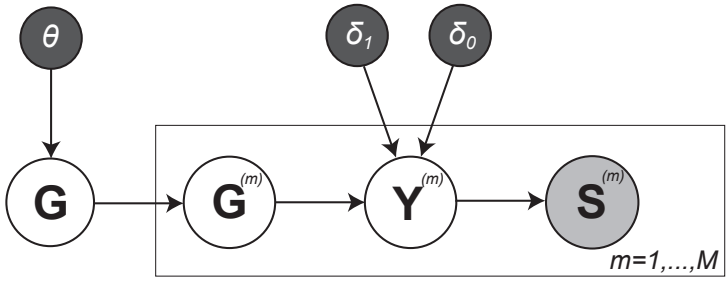


FIGURE 2.2: The hierarchical model describes how the connectivity for a subject depends on its streamline distribution but also on the connectivity in other subjects as mediated through parent network $\bar{\mathbf{G}}$.

where $P(\mathbf{G} \mid \theta) = \prod_{i<j} P(g_{ij} \mid \theta)$. Recall that $g_{ji} \equiv g_{ij}$ by definition. Here we specify $\theta = 0.5$, which represents that we have no preference for either the presence or absence of a connection.

Hierarchical model

So far, we assumed that data for each subject is analyzed independently. However, in practice, data for multiple subjects may be available and data for one subject might inform the inference for another subject. The intuition is that brain connectivity will, to a certain extent, be similar across subjects. Therefore, borrowing statistical strength from other subjects should lower the susceptibility to noise and artifacts in a single subject. This can be achieved by formulating a hierarchical model, where subject-dependent parameters at the first level are tied by subject-independent parameters at the second level. Figure 2.2 depicts this hierarchical model.

Suppose streamline data $\mathcal{S} = (\mathbf{S}^{(1)}, \ldots, \mathbf{S}^{(M)})$ is acquired for M subjects. Furthermore, let $\mathcal{G} = (\mathbf{G}^{(1)}, \ldots, \mathbf{G}^{(M)})$ denote a vector whose elements $\mathbf{G}^{(m)}$ refers to the connectivity matrix for subject m. In the hierarchical model, we assume that the different subjects are related through parent connectivity $\bar{\mathbf{G}}$. The different $\mathbf{G}^{(m)}$ are conditionally independent given $\bar{\mathbf{G}}$. For a new subject $m+1$, the quantity of interest is the posterior marginal

$$\begin{aligned} &P\left(\mathbf{G}^{(m+1)} \mid \mathcal{S}, \mathbf{S}^{(m+1)}, \zeta\right) \\ &\quad \propto P\left(\mathbf{S}^{(m+1)} \mid \mathbf{G}^{(m+1)}, \delta_1, \delta_0\right) P\left(\mathbf{G}^{(m+1)} \mid \mathcal{S}, \zeta\right). \end{aligned}$$

We could approximate this marginal by sampling from the hierarchical model. However, this is a computationally demanding task as it requires the simultaneous estimation of all of the adjacency matrices belonging to each of the subjects, as well as the parent network $\bar{\mathbf{G}}$. Instead, we specify a prior based on the connectivity obtained for other subjects. This improvement over the Erdős-Rényi model defines a separate connection probability for each individual edge instead

of using a single parameter θ to specify the connection probability for complete networks. This multi-subject prior is derived from the hierarchical model in Appendix A.1 and is equal to:

$$P\left(\mathbf{G}^{(m+1)} \mid \mathcal{S}, \zeta\right) = \prod_{i<j} p_{ij}^{g_{ij}^{(M+1)}} (1-p_{ij})^{(1-g_{ij}^{(M+1)})} \quad , \tag{2.6}$$

where $p_{ij} \equiv (\sum_{r=1}^{M} \hat{g}_{ij}^{(r)} + 1)/(M+2)$ with $\hat{g}_{ij}^{(r)}$ the maximum likelihood (ML) estimate for subject r. Hence, we derive a prior for subject $M+1$ from the ML estimates for subjects $1, \ldots, M$. These estimates can be obtained by running the single-subject models together with a flat prior. The multi-subject prior can then be plugged into (2.1), replacing (2.5), to produce the posterior for subject $M+1$.

Approximate inference

Since the posterior (2.1) cannot be calculated analytically, we resort to an MCMC scheme to sample from this distribution [Mukherjee and Speed, 2008]. We always start the sampling chain with a random symmetric adjacency matrix without self-loops. A new sample is proposed based on a previous network $\mathbf{G}$ by flipping an edge, resulting in a network $\mathbf{G}'$ (which, because of the symmetry of $\mathbf{G}$, implies $g'_{ij} = 1 - g_{ij}$ and $g'_{ji} = 1 - g_{ji}$). The acceptance of the proposed sample is determined by the ratio

$$\gamma = \frac{P\left(\mathbf{G}' \mid \mathbf{S}, \zeta\right)}{P\left(\mathbf{G} \mid \mathbf{S}, \zeta\right)} \quad .$$

A proposed network is accepted with probability $\min(1, \gamma)$ with $\log \gamma = \Delta L_{kl} + \Delta P_{kl}$. Here, ΔL_{kl} and ΔP_{kl} define the change in log-likelihood and log-prior respectively, after flipping edge g_{kl}. A complete derivation of these terms is given in Appendix A.2.

The sample distributions were obtained for each subject by drawing ten parallel chains of 300 000 samples (discarding the first 60 000 samples as burn-in phase and keeping only each 600th sample to assure independence). The collection of T accepted samples $\{\mathbf{G}^{(1)}, \ldots, \mathbf{G}^{(T)}\}$ forms an approximation of the posterior $P\left(\mathbf{G} \mid \mathbf{S}, \zeta\right)$. The samples can be used to estimate posterior probabilities of network features, such as the probability of a specific connection. Assuming the Markov chain has converged, the posterior probability of a single connection is given by $\mathbb{E}\left[g_{ij} \middle| \mathbf{S}\right] = \frac{1}{T} \sum_{t=1}^{T} g_{ij}^{(t)}$. Other summary statistics for the distribution may be estimated in a similar manner.

Validation of structural connectivity estimates

Structural connectivity constrains functional connectivity [Honey et al., 2010; Cabral et al., 2012]. In other words, when there is functional connectivity, there is often structural connectivity, although structural connectivity is not a necessary requirement for functional connectiv-

ity [Honey et al., 2009]. We exploit this relationship in the validation of structural connectivity estimates. This is achieved by constraining the conditional independence structure of functional activity by structural connectivity [Marrelec et al., 2006; Smith et al., 2011; Varoquaux et al., 2010; Deligianni et al., 2011]. Assume that a $p \times 1$ vector of BOLD responses $\mathbf{x}$ can be modeled by a zero-mean Gaussian density with inverse covariance matrix $\mathbf{K}$, i.e.

$$\mathbf{x} \,|\, \mathbf{K} \sim \mathcal{N}(0, \mathbf{K}^{-1}) \ . \tag{2.7}$$

Then, given acquired resting-state data $\mathbf{X} = (\mathbf{x}_1; \ldots; \mathbf{x}_n)$ for n time points, model estimation may be performed by finding the maximum likelihood solution

$$\hat{\mathbf{K}} = \underset{\mathbf{K} \in \mathbb{P}_p}{\operatorname{argmax}} \, L(\mathbf{X} \,|\, \mathbf{K}) \ ,$$

in which $\mathbb{P}_p$ is the space of positive definite matrices of size $p \times p$ and $L(\mathbf{x} \,|\, \mathbf{K}) = n/2 \log |\mathbf{K}| - 1/2 \left\langle \mathbf{K}\hat{\mathbf{C}} \right\rangle$ is, up to a constant, the log-likelihood function of $\mathbf{K}$, with $\hat{\mathbf{C}} = \frac{1}{n}\mathbf{X}^T\mathbf{X}$ the empirical covariance matrix.

However, in general for fMRI data, $p > n$, which implies that the covariance matrix is not full rank. Hence, finding its inverse requires suboptimal solutions such as the generalized inverse or pseudo-inverse [Ryali et al., 2012]. As a solution to this problem, regularized sparse approximations of the inverse covariance matrix have been proposed [Friedman et al., 2008; Huang et al., 2010]. In our setup, the sparsity structure is readily available in the form of structural connectivity $\mathbf{G}$. In order to use $\mathbf{G}$ as a constraint when estimating $\hat{\mathbf{K}}$, we can make use of the fact that variables i and j are conditionally independent if and only if $k_{ij} = 0$ [Dempster, 1972]. That is, we can interpret (2.7) as a Gaussian Markov random field with respect to network $\mathbf{G}$ such that $g_{ij} = 0 \Leftrightarrow k_{ij} = 0$ for all $i \neq j$ [Whittaker, 2009]. Let the notation $\mathbf{K} \in \mathbb{P}_\mathbf{G}$ denote that the independence structure of the positive definite matrix $\mathbf{K}$ is dictated by $\mathbf{G}$. As shown by Dahl et al. [2008], the ML estimate can be formulated as the following convex optimization problem:

$$\hat{\mathbf{K}} = \underset{\mathbf{K} \in \mathbb{P}_\mathbf{G}}{\operatorname{argmax}} \, L(\mathbf{X} \,|\, \mathbf{K}) \quad \text{subject to} \quad k_{ij} = 0 \Leftrightarrow g_{ij} = 0, i \neq j \ .$$

A standard convex solver [Schmidt et al., 2007] was used to find this constrained maximum likelihood estimate and use it to define the score for a particular network $\mathbf{G}$: $S(\mathbf{G}) \equiv L(\hat{\mathbf{K}})$. By comparing scores for different structural connectivity estimates, we are able to quantify the performance of a structural network in terms of how well it fits the functional data.

Since we compare different models, we have to take model complexity into account. We could opt for the use of a penalty term such as the Bayesian information criterion. Here, however, we use a more stringent approach, where we enforce constant model complexity. This is implemented by constraining the number of edges for all networks from one subject to be equal to that of the maximum likelihood (ML) solution $\mathbf{G}_{ML} = \operatorname{argmax}_\mathbf{G} P(\mathbf{S} \,|\, \mathbf{G}, \delta_1, \delta_0)$ of that particular subject. Recall that this maximum likelihood solution is equivalent to the so-

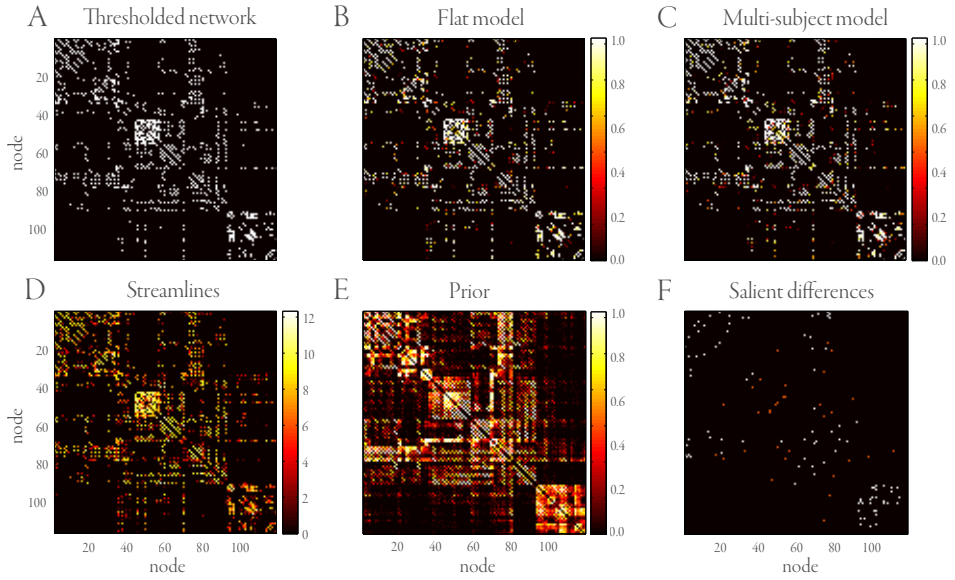


FIGURE 2.3: A.–E. Connectivity results for that subject for which sampling in conjunction with the multi-subject prior showed the largest improvement. Shown are A the network that is obtained through the thresholding approach, B the posterior connection probabilities according to the flat model, C the posterior connection probabilities according to the multi-subject model, D the streamline distribution on a log scale and E the multi-subject prior based on the other subjects as used in the multi-subject model. Panel F shows the most salient differences in connectivity between the maximum a posteriori networks and the thresholding approach, across all subjects. The edges are color-coded. White edges indicate those connections that were present in at least six subjects whereas these edges would not be part of the thresholded network. Orange edges show converse findings. All matrices are ordered according to the order in the AAL template.

lution obtained when using a flat prior in our generative model. For the multi-subject prior, the constraint on edge count is achieved by starting out with the converged ML solution and, subsequently, drawing new samples by simultaneously adding and removing an edge. For the thresholded networks, we choose a threshold such that the resulting number of edges is the same as that of the ML solution. Note that this approach is only a way to obtain a fair comparison between different structural networks and not a requirement of the model itself. The threshold was applied to the asymmetric streamline data, normalized according to the number of streamlines emanating from each node. Note that all added edges were symmetric.

2.2 RESULTS

In order to validate our framework, we made use of resting-state functional data which was acquired in conjunction with the diffusion imaging data. Specifically, we compared the fit to the functional data for structural networks either obtained by the standard thresholded approach or obtained using the developed generative model. The fit to the functional data is quantified

in terms of the score $S(\mathbf{G})$. We performed a comparison using either a flat prior (by choosing $\theta = 0.5$) or the multi-subject prior. For simplicity, the hyperparameters δ_1 and δ_0 were manually set to 1 and 0.1, respectively, as small values for the hyperparameters capture the burstiness phenomenon described in the introduction to this chapter. For the thresholded approaches, we have one structural network estimate, denoted by $\mathbf{G}_T$. In contrast, for our generative model, we have a posterior over structural networks, which gives rise to a distribution of scores $S(\mathbf{G}^{(t)})$ where t denotes sample index.

Comparing ML estimates with thresholded networks

The sparsity of the maximum likelihood estimates $\mathbf{G}_{ML}$, as obtained with the flat prior, was fairly constant (1019.2 ± 39.4 out of 6670 possible edges). As an example, Figure 3.2 shows connectivity results for one subject.

Although thresholding of streamline distributions is common practice, how exactly the threshold is applied varies between studies. To have a fair comparison, we investigated the impact of different thresholding approaches. We considered applying the threshold to the maximum, the mean and the minimum of s_{ij} and s_{ji}, respectively. To compare our model with these approaches, we computed for each subject the fraction of samples of the posterior network distributions that scored higher than thresholding. Let $f_{F\text{-}T}$ be the fraction of samples where the generative model with a flat prior scored higher than the thresholded network. The results for the distribution of $f_{F\text{-}T}$ over 20 subjects, given the different threshold methods, are shown in Table 2.1. When the threshold is applied to the maximum of s_{ij} and s_{ji}, the generative model outperforms thresholding. However, when either the mean or the minimum of s_{ij} and s_{ji} is used, samples obtained from the posterior with a flat prior score the same as thresholded networks, on average. To explain this behavior, it is instructive to consider (A.4) in Appendix A.2. Given hyperparameters δ_1 and δ_0 very small compared to elements of $\mathbf{S}$, the change in log-likelihood after flipping edge g_{ij} from absent to present boils down to

$$\Delta L_{ij} \approx (\delta_1 - \delta_0) \left[\log \left(\frac{s_{ij}}{\sum_k s_{ik}} \right) + \log \left(\frac{s_{ji}}{\sum_k s_{jk}} \right) \right] \quad .$$

This expression nicely summarizes the ramifications of our model. When sampling over networks, the generative model takes symmetry between streamlines into account (which follows from the sum) and it considers the relative distribution of streamlines (which follows from the fractions). Note that the latter is equivalent to normalizing the streamlines; a required step for thresholding. Thresholding approaches can imitate the behavior of the Dirichlet compound multinomial distributions by thresholding on either the mean or the minimum of s_{ij} and s_{ji} and by normalizing the streamline distribution by the number of outgoing streamlines.

TABLE 2.1: The fraction of samples that have a higher score than thresholded networks. The fraction of samples from the distribution with a flat and an multi-subject prior are represented by $f_{F\text{-}T}$ and $f_{M\text{-}T}$, respectively. The different threshold approaches are max, mean and min. The p-values were obtained using a one-sample t-test with $\mu_0 = 0.5$.

Threshold	$f_{F\text{-}T}$	p	$f_{M\text{-}T}$	p
max	0.60 ± 0.06	0.07	0.76 ± 0.07	<0.001
mean	0.50 ± 0.06	0.47	0.67 ± 0.07	0.02
min	0.49 ± 0.06	0.45	0.66 ± 0.08	0.03

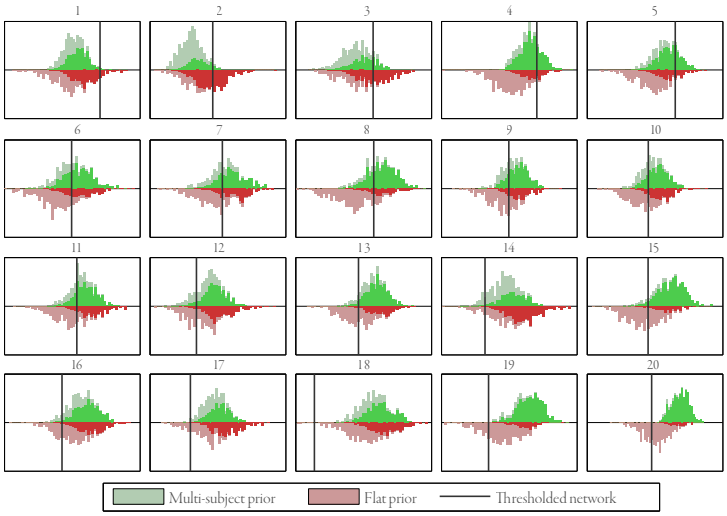
Multi-subject prior

With optimal threshold settings, it is possible to have thresholded networks that perform similar to the networks we infer through the posterior distribution with a flat prior. However, our model is capable of incorporating additional constraints, such as the multi-subject prior. Let $f_{M\text{-}T}$ be the fraction of samples where the DCM with the multi-subject prior scored higher than the thresholded network. The results for the distribution of $f_{M\text{-}T}$ over 20 subjects, given the different threshold methods, are shown in Table 2.1. In addition, we compared the fraction of samples obtained with the multi-subject prior that scored higher than samples with the flat prior, $f_{M\text{-}F}$. We found that this distribution had a mean of 0.64 ± 0.04 ($p < 10^{-3}$).

The likelihood scores estimated for the distributions over samples, obtained using our approach in the presence of either the flat prior or the multi-subject prior, are shown in Figure 2.4. In addition, the figure shows the score for the thresholded network, with a threshold applied to the minimum of s_{ij} and s_{ji}. The distributions obtained using the multi-subject prior are narrower and therefore more consistent than those obtained with the flat prior. Moreover, likelihood scores obtained using the multi-subject prior tend to be of higher magnitude than those obtained using the flat prior. From these results we can conclude that our model is up to par with the most optimal threshold approaches, but that it is capable of surpassing thresholded networks by using informative priors.

2.3 DISCUSSION

Standard thresholding approaches for the inference of whole-brain structural networks suffer from the fact that they rely on arbitrary thresholds while assuming independence between tracts and ignoring prior knowledge. In order to overcome these problems, we have put forward a Bayesian framework for inference of structural brain networks from diffusion-weighted imaging. Our approach makes use of a Dirichlet compound multinomial distribution to model the streamline distribution obtained by probabilistic tractography. In addition, we defined a simple prior on node degrees as well as a multi-subject prior that uses connectivity estimates from other subjects as an additional source of information.



#	$f_{F\text{-}T}$	$f_{M\text{-}T}$	$f_{M\text{-}F}$
1	0.01	0.01	0.52
2	0.35	0.03	0.26
3	0.35	0.12	0.39
4	0.09	0.23	0.78
5	0.10	0.24	0.72
6	0.31	0.56	0.69
7	0.40	0.62	0.63
8	0.24	0.67	0.80
9	0.42	0.69	0.68
10	0.47	0.70	0.65
11	0.43	0.71	0.65
12	0.68	0.84	0.57
13	0.56	0.91	0.71
14	0.89	0.91	0.40
15	0.46	0.93	0.85
16	0.75	0.93	0.61
17	0.88	0.99	0.52
18	0.99	1.00	0.59
19	0.78	1.00	0.86
20	0.64	1.00	0.92

FIGURE 2.4: The scores $S(\mathbf{G})$ (horizontal axis) for the thresholded network $\mathbf{G}_T$, samples from the generative model given a flat prior (lower histogram, red) and given the multi-subject prior (upper histogram, green). The fraction of each bin that has a bright color corresponds with the fraction of the other distribution that is outperformed by this bin. These fractions are also shown in the table to the right; $f_{F\text{-}T}$ is the fraction of samples with a flat prior that outperform thresholding, $f_{M\text{-}T}$ is the fraction of samples with the multi-subject prior that outperform thresholding and $f_{M\text{-}F}$ is the fraction of samples with the multi-subject prior that outperform samples with the flat prior. The subjects are ordered according to the performance of the multi-subject prior approach relative to the thresholded network.

The proposed methodology was validated using simultaneously acquired resting-state functional MRI data. The outcome of our experiments revealed that the generative model combined with a flat prior performs equally well as the most optimal thresholded network. The use of an informative multi-subject prior instead created networks that significantly outperformed the thresholding approach. A comparison between the networks obtained with the multi-subject or flat prior showed that the former improved on the latter, thereby motivating the use of the multi-subject prior.

In our setup, the hyperparameters δ_1 and δ_0 were set by hand and the edge probability θ was chosen to result in a flat prior. Instead these parameters could have been estimated from the streamline data using empirical Bayes, they could have been integrated out entirely in a full Bayesian sampling approach, or they could be optimized according to the resting-state functional data. Note further that a fair comparison between networks required model complexity to be controlled. This was achieved via the constraint that networks obtained with either the multi-subject prior or with the thresholding approach had the same number of edges as the most probable network with a flat prior. While this is to the advantage of the thresholding approach, since no arbitrary threshold needs to be chosen, it can only impede networks obtained using the multi-subject prior since that might support a different number of tracts.

Even given optimal settings for the thresholding approach, our approach shows clear benefits. Foremost, the DCM model intuitively assigns probabilities to the existence of edges in the inferred networks, providing a mechanism to cope with the uncertainty in the data. Moreover, the proposed generative model allows for intuitive and well founded priors, such as the described multi-subject prior. The hierarchical model in Figure 2.2 also allows for group-level inference [Robinson et al., 2010]. This means that, given streamline data for multiple subjects, the generative model can be used to infer individual subject connectivity G as well as the group-level parent network $\bar{\mathbf{G}}$. This allows one to get a handle on group differences, for instance, in the context of clinical neuroscience. The current work focused mainly on the empirical validation of our theoretical framework using functional data. In future work, we will focus more on interpretation of the obtained structural connectivity estimates.

In this chapter we used resting-state fMRI data as a means to validate whole-brain structural networks derived from diffusion-weighted imaging. A logical extension of our work is to derive connectivity based on the integration of these two imaging modalities. This example of Bayesian data fusion requires that we extend the generative model to take functional data into account as well [Rykhlevskaia et al., 2008; Sui et al., 2011]. We can then use structural networks as an informed prior for inference of functional connectivity or infer structural connectivity using both modalities simultaneously.

An additional benefit of our framework is that the network sparsity follows directly from optimizing (2.1). In the thresholding approach, the network sparsity is a consequence of the specific threshold setting. As a byproduct of our study, we have observed that thresholding of streamlines benefits from considering the mean or minimum of the number of streamlines connecting A to B and vice versa. This in itself may lead to improvements in the analysis of structural connectivity.

Summarizing, we proposed a Bayesian framework which lays the foundations for a theoretically sound approach to the inference of whole-brain structural networks. This framework does not suffer from the issues which plague current thresholding approaches to structural connectivity estimation and has been shown to give rise to substantially improved structural connectivity estimates. The proposed generative model is easily modified to incorporate other sources of information, thereby further facilitating the estimation of whole-brain structural networks in vivo.

3

A CLUSTERING PRIOR FOR STRUCTURAL CONNECTIVITY

The brain can be described as a vast network of interconnected neurons. The acquisition and subsequent analysis of this network has resulted in a discipline known as *connectomics* [Sporns et al., 2005; Hagmann, 2005]. Part of connectomics focusses on *structural connectivity*, which concerns the layout of physical tracts consisting of axonal fibers. At a macroscopic scale, structural connectivity is defined in terms of brain regions that consist of neuronal populations which are inter-connected via white matter fiber bundles. As neuronal activity is constrained by neuroanatomy, correctly identifying these structural connections aids in understanding how spatially remote regions of the brain cooperate [Greicius et al., 2009; Honey et al., 2009]. Both functional and structural connectivity have been shown to be relevant in clinical applications, for instance by characterizing the connectomes that correspond to neurological and/or psychological disorders [Catani, 2007; Craddock et al., 2013; Fornito and Bullmore, 2012]. Similarly, connectivity studies have become a useful aid in understanding cognition by elucidating which networks are related to particular functions [Bressler and Menon, 2010; Baronchelli et al., 2013; Sporns, 2014; Ekman et al., 2012].

An important application of structural connectivity is the delineation of functionally specialized clusters of brain regions based on their structural connectivity patterns. Importantly, connectivity-based parcellation can be obtained *in vivo*, contrasting it with other methods such as histological analysis of cytoarchitecture [Schleicher et al., 1999; Caspers et al., 2006]. Other non-invasive methods that delineate structural cortical boundaries exist, based on identification of major anatomical landmarks, but these approaches have been shown to be susceptible to large inter-subject variability [Klein et al., 2007]. Intuitively, a cluster consists of a set of regions that are more similar to other regions in the cluster than to regions outside of it, but this idea can be operationalized in different ways.

This chapter is based on: M. Hinne, M. Ekman, R.J. Janssen, T. Heskes and M.A.J. van Gerven, 2015. "Probabilistic clustering of the human connectome identifies communities and hubs". *PLoS ONE* 10(1), pp. e0117179.

Under the first interpretation, which we refer to as *profile-based clustering*, regions may be clustered based on similarity of their connectivity profiles (see e.g. [Knösche and Tittgemeyer, 2011; Johansen-Berg et al., 2004, 2005; Anwander et al., 2006; Mars et al., 2011], or see [Cloutman and Lambon Ralph, 2012] for a literature review). In other words, it assumes that regions in the same cluster connect with the same areas. Several studies have used this approach to parcellate regions of interest such as the frontal pole [Liu et al., 2013], posteromedial cortex [Zhang et al., 2014], occipital lobes [Thiebaut De Schotten et al., 2014], cingulate cortex [Beckmann et al., 2009] and thalamus [Serra et al., 2013; Behrens et al., 2003b; O'Muircheartaigh et al., 2011]. Note that in this interpretation of connectivity-based clusters, regions within a cluster are not necessarily mutually connected. Implicitly, this approach aims to find regions that similarly integrate information from other parts of the brain.

Under the second interpretation, which we refer to as *community-based clustering*, parcellations are taken to consist of densely connected clusters that are only sparsely connected to regions outside the cluster. This approach is typically used for whole-brain parcellation [Hilgetag and Kaiser, 2004; van den Heuvel et al., 2008; Meunier et al., 2009; Craddock et al., 2012; Power et al., 2011]. Here, one implicitly assumes that clusters are structurally (and thus, indirectly, functionally) specialized and mostly interact with the other regions in their respective cluster.

Either operationalization of connectivity-based clustering is applicable to the human brain. As a consequence, choosing either perspective may hinder finding meaningful clusters that only adhere to the other definition of a cluster. Furthermore, clustering behavior may not be uniform across the brain, and instead may be a mixture of profile-based and community-based clusters. In this study we introduce a probabilistic model that is able to parcellate structural connectivity and is sufficiently flexible to incorporate the different cluster interpretations. The model reveals community-based clusters in the human connectome, but also a set of nodes that are assigned to small profile-based clusters that function as hubs. These small clusters all contain regions that have been labeled the 'rich club' [van den Heuvel and Sporns, 2011]. In addition, we find that the whole-brain parcellations which the model provides, are up to par with algorithms that have previously been used to parcellate connectivity. Moreover, in contrast to parcellations obtained with the other approaches, parcellations estimated by the probabilistic model can be obtained without the need to choose the number of clusters in advance. Finally, because the model is probabilistic, it is able to explicitly represent the uncertainty in the obtained parcellations. These visualizations show that the cluster assignment for particular regions is uncertain, indicating that care should be taken when interpreting connectivity-based parcellations obtained using alternative deterministic procedures.

3.1 PROBABILISTIC MODEL FOR CONNECTIVITY-BASED CLUSTERING

First, we describe a probabilistic model for clustering based on structural connectivity. Next, we extend this model to allow for direct estimation of cluster structure from probabilistic tractography data. The complete model is shown in Figure 3.1A.

Stochastic block model

Let us assume that each cluster has a distinct connectivity profile. In other words, each cluster has a set of parameters that govern its connectivity behavior. This is known as a stochastic block model (SBM) [Nowicki and Snijders, 2001]. Within a cluster, all nodes (i.e. brain regions) use the same connectivity parameters. As a consequence, regions within a cluster are stochastically exchangeable. Formally, each node $1, \ldots, p$ is assigned a cluster label, using the unobserved (latent) cluster assignment variable $z_{ij} = 1$ if node i is in cluster j, and 0 otherwise. Each node is assigned to exactly one cluster. Structural connectivity is described by a symmetric and binary adjacency matrix $\mathbf{G}$. Because of the stochastic exchangeability assumption, the probability of a connection g_{ij} depends solely on the clusters to which nodes i and j are assigned. The cluster connection probabilities between clusters a and b are given by ρ_{ab}. The set of cluster connection probabilities is collected in the connection probability matrix $\boldsymbol{\rho}$. In practice, $\boldsymbol{\rho}$ is unknown so we use a prior on $\boldsymbol{\rho}$ to reflect our assumptions about it. Here, this is a Beta prior that depends on two hyperparameters α and β, which model the probability of a connection or non-connection between different clusters, respectively. We assume an uninformative prior with $\alpha = \beta = 1$. After observing the data, the posterior expectation for $\boldsymbol{\rho}$ will reflect cluster connection probabilities and therefore informs about the contributions of profile-based and community-based clusters [Andersen et al., 2014]. As an example of how block models may represent different connectivity patterns, Figure 3.1B shows the posterior expectations for the connection probability matrix using a toy network of twelve nodes distributed evenly over three clusters. Here, the network on the left shows how profile-based clusters may be captured by the off-diagonal weights of $\boldsymbol{\rho}$, while the network on the right shows how strong weighs on the diagonal of $\boldsymbol{\rho}$ generate a traditional community-based network.

SBM have seen widespread application in literature, ranging from discovery of roles in social networks [Nowicki and Snijders, 2001] to identification of protein-protein interactions [Airoldi et al., 2009; Guimerà and Sales-Pardo, 2009]. In addition, several model variants have been introduced, such as approaches that deal with overlapping clusters [Wolfe and Jensen, 2004] or SBM tailored to weighted networks [Aicher et al., 2015]. However, these approaches assume that the number of clusters K is known a priori, which is frequently not the case with empirical data. A nonparametric extension was introduced to learn the number of clusters from data as well [Kemp et al., 2006]. This is achieved by placing a prior distribution on the cluster assignment matrix $\mathbf{Z}$. Doing so allows the model to accommodate a (potentially) infinitely large number of clusters, rather than needing to specify the number of clusters K beforehand. Specifically, we draw $\mathbf{Z}$ from a Chinese restaurant process (CRP) [Aldous, 1985]. This distribution over partitions can be used to generate samples from in the following way. Consider nodes that are assigned to clusters one by one, as customers entering a restaurant and choosing a table to sit at. Each customer is assigned to a non-empty table k with probability $\frac{m_k}{p-1+\xi}$, with m_k the number of customers already assigned to table k and with probability $\frac{\xi}{p-1+\xi}$ to an empty (new) table. Its concentration parameter ξ determines how likely it is for a customer to sit at an

empty table, which affects the total number of tables with customers, i.e. the number of clusters. Using the CRP, the generative model is then given by

$$\begin{aligned} \mathbf{Z} \mid \xi &\sim \mathrm{CRP}(\xi) \\ \rho_{ab} \mid \alpha, \beta &\sim \mathrm{Beta}(\alpha, \beta) \\ g_{ij} \mid \rho_{ab}, \mathbf{Z} &\sim \mathrm{Bernoulli}(z_i \rho z_j^{\mathsf{T}}) \ , \end{aligned} \tag{3.1}$$

where we introduce the notation m_i to indicate the ith row of a matrix $\mathbf{M}$. The model is known as the infinite relational model (IRM) [Kemp et al., 2006; Mørup et al., 2010; Xu et al., 2007; Andersen et al., 2014]. The infinite relational model is easily generalized to encompass multiple (conditionally independent) subjects that share a parcellation [Mørup et al., 2010] by changing the model into

$$\begin{aligned} \rho_{ab}^{(m)} \mid \alpha, \beta &\sim \mathrm{Beta}(\alpha, \beta) \\ g_{ij}^{(m)} \mid \rho_{ab}^{(m)}, \mathbf{Z} &\sim \mathrm{Bernoulli}(z_i \rho^{(m)} z_j^{\mathsf{T}}) \end{aligned} \tag{3.2}$$

where the superscript m indicates the subject index. Note that although the clustering $\mathbf{Z}$ is shared across subjects, the cluster-to-cluster connection probabilities are subject-specific. When the IRM is applied to binary connectivity data, we will refer to the model as the bIRM.

Forward model for streamlines

To infer cluster assignments, structural connectivity data must be provided in the form of a binary adjacency matrix $\mathbf{G}$. This matrix can be obtained using probabilistic tractography [Behrens et al., 2003a]. Probabilistic tractography proceeds by drawing streamlines between brain regions based on local estimates of anisotropic diffusion. These streamlines are collected in a streamline count matrix $\mathbf{S}$, reflecting the number of streamlines between pairs of brain regions. Thresholding of the streamline count matrix produces a binary matrix that reflects structural connectivity. However, as a threshold results in a point estimate, we make use of a probabilistic model that describes how a structural network generates the distributions of probabilistic streamlines that are obtained through tractography (see Chapter 2). Using this forward model, a streamline threshold is no longer required.

Ideally, probabilistic streamlines show a distribution that perfectly reflects the underlying structural connectivity. However, tractography is prone to noise and errors, in particular in the presence of kissing, splaying and crossing fibres [Bassett et al., 2011]. Hence, we distinguish between true connections along which we expect to observe streamlines, and false connections that may occasionally display streamlines, but do not correspond to actual anatomical pathways. The probability of a streamline between a pair of regions is represented by the matrix $\mathbf{X}$. Formally, the streamline probability vector y_i for a particular region i is determined by a Dirichlet distribution with parameter δ_1 for true connections and δ_0 for false connections. Ac-

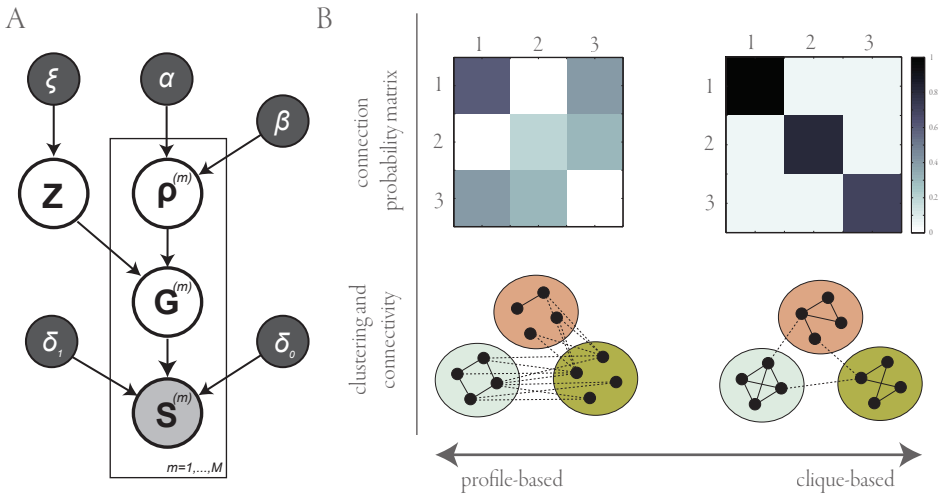


FIGURE 3.1: Probabilistic model for connectivity-based parcellation. A. The streamline infinite relational model combines a forward model for streamline data $\mathbf{S}$ with an infinite relational model that allows estimation of the cluster assignment matrix $\mathbf{Z}$ as well as the connection probability matrix $\boldsymbol{\rho}$. Hyperparameters $\{\xi, \alpha, \beta, \delta_1, \delta_0\}$ complete the model. B. The probabilistic model supports both profile-based clustering as well as community-based clustering. The top row shows simulated connection probability matrices that correspond to profile-based clustering (left) and community-based clustering (right). Example networks that correspond to these probabilities are shown at the bottom.

cording to these probabilities, streamlines are distributed amongst the target regions using a multinomial distribution. By integrating out the streamline probability vectors, we obtain the following forward model:

$$\mathbf{s}_i \,|\, \mathbf{g}_i, \delta_1, \delta_0 \quad \sim \quad \text{DirMul}(\delta_1 \mathbf{g}_i + \delta_0(\mathbf{1}_n - \mathbf{g}_i)) \tag{3.3}$$

where $\text{DirMul}(\boldsymbol{\alpha})$ stands for the Dirichlet compound multinomial distribution with $\boldsymbol{\alpha}$ its vector of parameters. Note that this formulation assumes undirected structural connectivity, but allows streamline counts s_{ij} and s_{ji} to be different.

To estimate connectivity from streamline data, (3.3) is supplied with a prior on $\mathbf{G}$. In the most straightforward case, this prior is uniform, i.e. $P(\mathbf{G}) \propto 1$ (see Chapter 2). Alternatively, this uninformative prior may be replaced by the IRM (see Figure 3.1A), expressing our assumption of clustering in the network. The interpretation of the integrated model is that most observed streamlines indicate a structural connection (although there is some noise in the tractography process), and regions in the same cluster share their connectivity preferences (although some exceptions are allowed). We refer to the combined model that operates on streamline data as sIRM. Details of the approximate inference algorithm used to compute posterior estimates of interest are provided in Appendix B.1.

Visualizing clustering uncertainty

In addition to the display of individual parcellations, such as the maximum a posteriori (MAP) estimate, the proposed method may be used to visualize uncertainty about a parcellation. However, as the number of clusters may vary within a clustering distribution, and cluster labels are arbitrary for each sample, such a visualization is not trivial. A representation that allows different samples to be compared is the cluster co-assignment matrix $\mathbf{M} = \mathbf{Z}^T\mathbf{Z}$. The expectation of $\mathbf{M}$, i.e. $\mathbb{E}[\mathbf{M}] = \frac{1}{T}\sum_{s=1}^{S}\mathbf{M}_s$, with T the number of obtained samples (see Appendix B.1), describes the posterior probability that two regions are assigned to the same cluster. The probability of co-assignment is used to color a region i with a weighted color coding given by $c_i = \sum_j m_{ij}\hat{c}_j / \sum_j m_{ij}$, with $\hat{c}_j$ the color representation in the MAP estimate for region j. For example, suppose that the MAP estimate consists of two clusters, one colored red and one colored yellow. If a region i is co-assigned to a region in the red cluster in half the samples, and to a region in the yellow cluster in the other half of the samples, it will be colored orange.

Evaluating parcellation quality

Since a clustering ground truth is unavailable, we used the reproducibility of the parcellations as a indicator of parcellation quality [Klein et al., 2007; Blumensath et al., 2013; Thirion et al., 2014]. We quantify the reproducibility as the similarity between the parcellations of different participants (or groups of participants), expressed using the adjusted mutual information (AMI) [Vinh et al., 2010]. The AMI measure differs from the more traditional normalized mutual information measure in that it compensates for possible bias as a result of different numbers of clusters per parcellation. The measure is defined as

$$S(\mathbf{Z}_1, \mathbf{Z}_2) = \frac{\mathrm{MI}(\mathbf{Z}_1, \mathbf{Z}_2) - \mathbb{E}\left[\mathrm{MI}(\mathbf{Z}_1, \mathbf{Z}_2)\right]}{\max\left(\mathrm{H}(\mathbf{Z}_1), \mathrm{H}(\mathbf{Z}_2)\right) - \mathbb{E}\left[\mathrm{MI}(\mathbf{Z}_1, \mathbf{Z}_2)\right]} \quad , \tag{3.4}$$

with $\mathrm{MI}(\mathbf{Z}_1, \mathbf{Z}_2)$ the mutual information between two clusterings, $\mathrm{H}(\mathbf{Z})$ the entropy of a clustering and $\mathbb{E}\left[\mathrm{MI}(\mathbf{Z}_1, \mathbf{Z}_2)\right]$ the expected mutual information between two clusterings.

For comparison, we also obtain parcellations using K-means, the canonical algorithm for profile-based clustering, and using Infomap [Rosvall and Bergstrom, 2008], an algorithm that relies on community-based clustering (see Appendix B.3). Both methods have been used before in the context of brain parcellation [Anwander et al., 2006; Power et al., 2011]. Each of the comparison algorithms was applied to the empirical streamline distributions. For K-means and Infomap the streamline matrices were made symmetric, i.e. $\mathbf{S}' = \mathbf{S} + \mathbf{S}^T$. The number of clusters for K-means and Infomap is fixed to be the same as for the MAP estimate of the sIRM approach. To apply the bIRM, we first obtained the MAP estimate of connectivity with a flat prior, i.e. $P(\mathbf{G}) \propto 1$, which was subsequently clustered using the IRM approach. For the group-level analysis, a group-level streamline matrix was created by summing the streamline counts of all ten participants per split. This matrix was provided as input for K-means and

Infomap. The bIRM method uses the ten individual MAP estimates of the forward model as input, while the sIRM method uses the ten streamline matrices.

3.2 RESULTS

The sIRM was used to obtain posterior distributions as well as MAP estimates for twenty participants, using the parameter settings described in Appendix B.2. On average, the MAP clusterings of these participants consisted of 12.10 (SD 1.29) clusters. The posterior distributions of the numbers of clusters are centered tightly around the MAP estimates, as evidenced by a mean range of the 95% credible intervals of only 0.85 (SD 0.81) (i.e. most of the samples of the approximated distribution have the exact same number of clusters). To analyze the behavior of group-level parcellations, we created 20 random splits of the set of subjects and obtained the MAP parcellation for all of the 40 halves. These 40 parcellations consisted on average of 15.03 (SD 0.83) clusters. At both the individual and the group-level we find that all identified clusters are spatially contiguous, with the exception that the superior frontal gyrus and the precuneus are sometimes assigned to the same cluster as their functional homologue in the contralateral hemisphere.

As an example, the MAP estimate for one participant is shown in Figure 3.2A–D. The other MAP estimates are shown in Fig. S2 and Fig. S3 for the individual participants and the 40 halves, respectively. Figure 3.2A shows the connectivity matrix $\mathbf{G}$. Within-cluster connections are colored with the color of their respective clusters and between-cluster connections are colored black. Figure 3.2B shows the probability of a connection between pairs of clusters, as represented by the connection probability matrix $\boldsymbol{\rho}$. The number of connections between pairs of clusters, $\mathbf{Z}\mathbf{G}\mathbf{Z}^{\mathsf{T}}$, is shown in Figure 3.2C. Note that the number of possible connections grows with m_k^2, so while the amount of connection increases, the cluster connection probability may decrease. A visualization of the layout of the connectivity $\mathbf{G}$ and clustering $\mathbf{Z}$ of this MAP estimate is shown in Figure 3.2D.

Figure 3.2E shows the interpolated clustering colors based on the approximated posterior distribution of $\mathbf{Z}$ for this participant. The figure reveals that although the expectation is not very different from the MAP estimate, there remains some room for uncertainty. For instance, the right parietotemporal cortex and the right superior temporal cortex are seen as separate clusters in the MAP estimate, but appear to be merged into a single cluster in a substantial part of the distribution. Further uncertainty is shown in the assignment of left thalamus, right precuneus and left inferior parietal cortex. Zooming in on these regions results in what we will refer to as a 'cluster probability map'; a map that, for any given region of interest i, shows the probability that it is assigned to the same cluster as another region. This corresponds to a row $\mathbf{m}_i$ of $\mathbf{M}$. Figure 3.3 visualizes the cluster probability maps for the right inferior frontal gyrus and the left postcentral gyrus, for the same participant as shown in Figure 3.2. Maps like these serve as further illustration that there may be substantial uncertainty in cluster assignments, and that point estimates should be used with care.

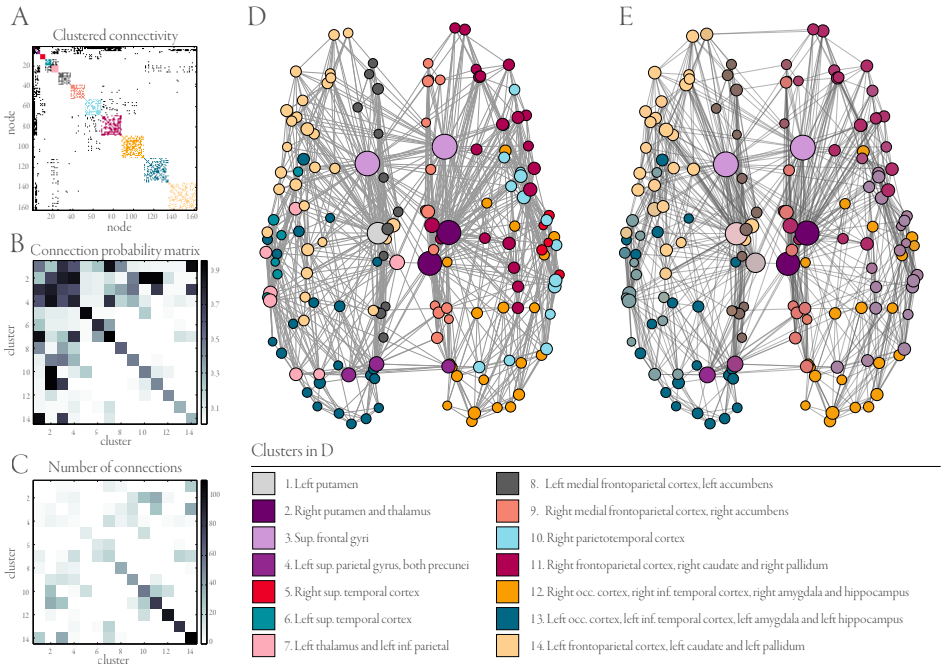


FIGURE 3.2: Maximum a posteriori parcellations for one participant. A. The adjacency matrix $\mathbf{G}$. B. The connection probability matrix $\boldsymbol{\rho}$. C. The number of connections between clusters $\mathbf{Z}\mathbf{G}\mathbf{Z}^{\mathsf{T}}$. D. Visualization of the maximum a posteriori network structure and parcellation. The clusters are color coded to be able to compare the network with the adjacency matrix in A. Node sizes are scaled by their degree. E. Visualization of the expectation of network structure and parcellations. Colors are interpolated with the MAP estimate as point of reference (see text). To keep the visualization uncluttered, only the m most probable edges are shown, where m is the number of edges in the MAP estimate.

Comparison with other methods

The parcellations for K-means, Infomap and bIRM are shown in Figures S4–S6, respectively, together with the connections that correspond to the top 5% streamline counts. For the bIRM approach, this resulted in 8.25 (SD 0.55) clusters. At the group level, 10.18 (SD 0.45) clusters were found for the bIRM.

The results of the reproducibility comparisons are shown in Figure 3.4. At the level of individual parcellations, only K-means performs notably less consistent than the other methods. The sIRM results in sparse connectivity (the MAP estimates of the twenty subjects have a density of $\pm 7\%$, on average) and its parcellations are up to par with the bIRM and Infomap. At the group level, Infomap appears to be the most consistent method and K-means is again the method that is least consistent. The group-level parcellations obtained by the sIRM are similar in consistency to those obtained by the bIRM.

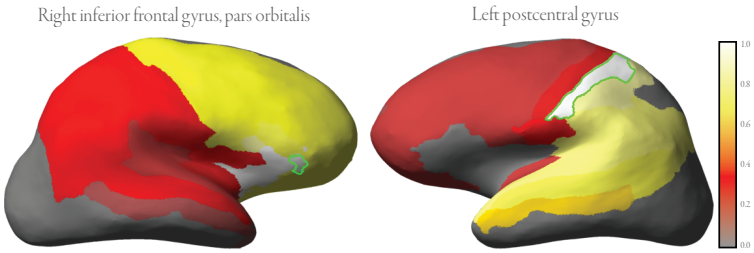


FIGURE 3.3: Cluster probability map to visualize the uncertainty of the resulting cluster assignments. (left) Cluster probability for the inferior frontal gyrus. The map shows that the highlighted region is likely to be assigned to regions in the frontal cortex, and to a much lesser degree to regions in the parietal cortex. The opposite pattern is shown for the postcentral gyrus (right).

Community-based versus profile-based clusters

The MAP estimates of ρ as well as the number of connections between clusters in Figure 3.2 show that clustering behavior depends strongly on cluster type and that some clusters show community-like tendencies while others do not. To quantify the extent to which a cluster forms a community, and to be able to compare this with the other methods, we computed for each cluster c the ratio r_c of within-cluster streamlines versus the total number of streamlines connected to this cluster. When this ratio approaches 1, the corresponding cluster is a community. If instead the ratio approaches 0, it can be regarded a profile-based cluster. As the number of clusters can be different between the sIRM and the bIRM methods, we then assigned to each node the score r_c its corresponding cluster had.

Figure 3.5 shows for both the single-participant as well as for the group-level parcellations the ratio r_c for all nodes, as averaged over all participants and all group-level parcellations, respectively. The results for the sIRM method clearly reveal that parcellation can be divided into two regimes. The first consists of clusters with relatively high r_c values (up to 0.76). These are large clusters that correspond predominantly to major cortical areas or lobes that are highly intra-connected. They are connected via the second regime with low r_c values (as low as 0.02), that contains small clusters (containing less than five nodes) and have few within-cluster streamlines. In at least 15 out of 20 participants, the small clusters contain the bilateral putamina, bilateral superior frontal gyri and right thalamus. For the group-level parcellations we observe similar behavior. Most clusters tend to be communities, connected via small clusters consisting of one or two regions. Small-cluster regions that occur in at least 26 out of 40 group-level parcellations are the bilateral precunei, superior frontal gyri, thalami, putamina and left superior parietal gyrus.

The other approaches show different patterns. For the single-participant parcellations, bIRM, K-means or Infomap show no regions that consistently appear in clusters with low r_c. For group-level parcellations, bIRM finds bilateral superior frontal gyri and right putamen (a subset of the regions found by the sIRM), K-means finds right occipital gyrus, right orbital sulcus and right medial olfactory sulcus and Infomap finds bilateral pallida and left putamen. Remarkably, the regions that stand apart according to the sIRM approach (for both the single-subject and

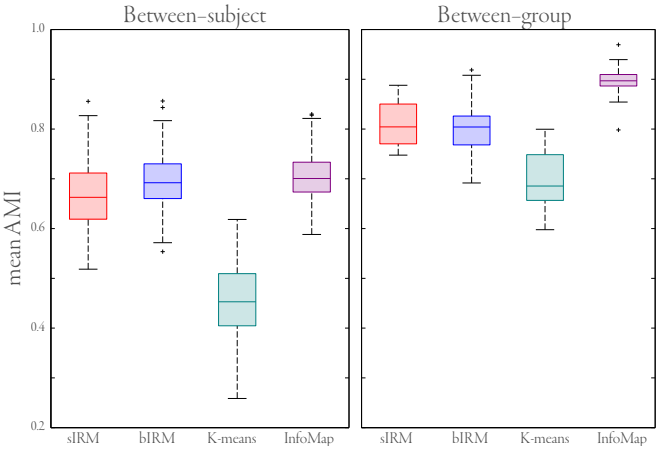


FIGURE 3.4: The reproducibility of parcellations. A. Mean adjusted mutual information (AMI, see main text) for pairwise comparisons between parcellations of all participants. B. Mean AMI for each of the pairs of parcellations that were created by randomly splitting the participant group into halves and obtaining a parcellation for each half.

the group-level parcellations) are all known to be part of the 'rich club' regions that integrate more remote cortical regions [van den Heuvel and Sporns, 2011]. Note that these regions do not mutually form a single large cluster, as each of them has substantially different connections to the rest of the brain.

3.3 DISCUSSION

Human brain connectivity is shown to exhibit clustering according to two different principles. Some regions show community-based clustering where clusters are both spatially contiguous and densely intraconnected. These clusters are tied together by brain regions that reveal connectivity-based clustering. The latter contain only one or two nodes and do not 'fit in' with other clusters due to substantially different connection profiles. These clusters consist predominantly of the superior frontal gyri, the superior parietal gyri, the precunei, the thalami and the putamina. All these regions have previously been identified as members of the 'rich club' [van den Heuvel and Sporns, 2011]. In addition, all regions except for the thalamus have been pointed out as being the most vulnerable and central regions in structural brain networks [Iturria-Medina et al., 2008]. Since the thalamus can be considered to act as a relay station, it stands to reason that it connects different cortical clusters. The rich-club regions stand apart from the rest of the connectome, as they are not part of the community-like clusters. Rather, they are each assigned their own cluster or group together with a similar region. Intuitively, this is an appealing observation. The community-like clusters contain regions dedicated to specialized processing. Here, an abundance of local connectivity, required for extensive within-cluster communication, results in dense intra-cluster connectivity. Simultaneously, the

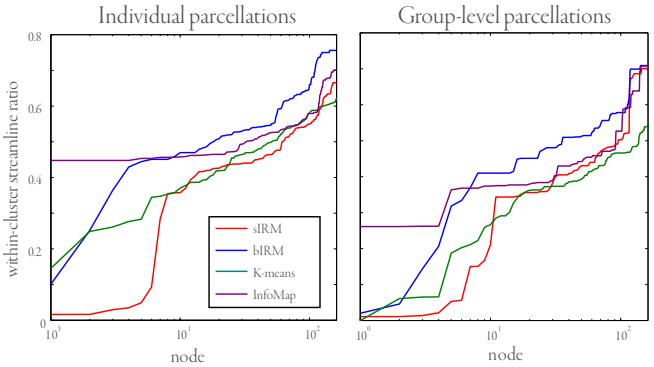


FIGURE 3.5: The ratio describing how community-like the clusters are that each region is assigned to. The plots show, for the cluster each region is assigned to, the ratio of within-cluster streamlines versus the total number of streamlines connected to that cluster, for parcellations of individual participants (left) and group-parcellations (right). This reveals that in particular for the sIRM method, this ratio is small for a number of regions, indicating that these regions are not part of community-based clusters. Regions that consistently have a low ratio, are described in the text. Note that the nodes corresponding to each line are ordered differently.

signal from these clusters must be integrated and disseminated, which is presumably achieved via the rich club regions [van den Heuvel and Sporns, 2013]. Note that although small clusters consisting of one or a few regions will not be community-based clusters by definition, because there are simply not enough possible internal edges, it is the finding that these regions are not assigned to bigger clusters containing other nodes that is interesting. This means that a substantial fraction of the connections of these regions is to several different clusters, which emphasizes their integrative role.

Validating the results from connectivity-based parcellation remains a difficult task [Cloutman and Lambon Ralph, 2012]. Since ground truth is not available, reproducibility across subjects is often used as a proxy for parcellation quality [Klein et al., 2007; Blumensath et al., 2013; Thirion et al., 2014]. Based on this criterion, we have shown that our approach performs well. Yet, by visual inspection of the parcellations (see Fig. S2,S4–S6) some pairs of regions appear grouped together that are not immediately obvious. For instance, in many parcellations we find that thalamus and putamen are assigned to the same cluster, as well as amygdala and hippocampus and bilateral superior frontal gyri. One may argue that in particular for the subcortical areas, these regions should appear in separate clusters as they have specialized functionality. However, since these parcellations are based on connectivity and these regions project similarly to the cortex, they are put together in a cluster. This is inherent to connectivity-based clustering and occurs for K-means and Infomap as well. In fact, of the algorithms we considered, sIRM appears to be the only method that consistently assigns many of the rich club regions to small clusters instead of agglomerating them into large clusters. Finally, we note that anatomical constraints such as enforcing that subcortical areas constitute singleton clusters may easily be added into the prior distribution.

Previously used methods are limited to one particular clustering behavior. A notable exception is a recent study that reveals that the connectome of the C. elegans roundworm also consists of a number of densely intraconnected clusters that are integrated via a core cluster that strongly connects to each community [Pavlovic et al., 2014]. As we described above, the human connectome reveals both densely connected clusters as well as disconnected clusters with very similar connectivity. Consequently, picking one perspective inevitably neglects part of the available structure, which is exemplified by the parcellations found using K-means and Infomap that assign the rich club regions to larger clusters. Although, in the case of Infomap, this results in more consistent parcellations, the more detailed picture of communities and integrating regions is lost.

Estimating parcellation with the sIRM approach provides a number of additional interesting quantities apart from the parcellation itself. The cluster connection probabilities $\boldsymbol{\rho}$ act as a cluster-level estimation of connectivity, expressed in terms of the rich-club regions and major cortices. Region-to-region connectivity $\mathbf{G}$ is estimated alongside clustering, and is potentially less prone to noise due to the prior that encourages regions in the same cluster to connect similarly, unless the data provides strong evidence to the contrary. Furthermore, other connectivity-based parcellation strategies, such as K-means, which groups regions with similar connectivity, and Infomap, which optimizes for densely connected components, provide point estimates of a parcellation. Although these approaches have provided valuable insight in the organization of both structural and functional connectivity [Power et al., 2011; Skudlarski et al., 2010; Nanetti et al., 2009], they do not quantify the uncertainty in their results (this issue is also discussed in [Thirion et al., 2014]). Instead, our proposed method characterizes the full posterior distribution of all variables involved and thus provides a richer representation of parcellation. We find that a number of regions show substantial uncertainty in their assignments to a cluster. This illustrates that point estimates should be used with care, as a number of regions could easily be assigned erroneously.

Integrating the forward model for structural connectivity with the IRM as prior leads to qualitatively different parcellations than when the bIRM is applied post-hoc to the connectivity estimates from the forward model. This is visible from the different number of clusters that both pipelines provide (the combined approach results in roughly 50% more clusters). In addition, the clear distinction between the two kinds of clusters is only marginally visible in the bIRM parcellations. Note that regardless of these differences, the reproducibility of the parcellations is similar for both methods.

The IRM is a nonparametric method, which is helpful since the number of clusters is not known a priori. Still, parameters remain that affect the resulting parcellation. In particular the parameters δ_1 and δ_0 that govern the estimated connectivity and concentration parameter ξ, that affects the number of identified clusters (although we observe that its influence is drowned out by the contribution of the likelihood) may be estimated from the data using empirical Bayes [Robbins, 1964]. However, this will incur substantial additional computation costs that must be overcome to arrive at an efficient model.

There are a number of directions one could take the proposed approach. To start with, alternative generative models may be used instead. For instance, the connectivity forward model

that consists of a Dirichlet compound multinomial distribution, modeling the probabilities of streamlines, may be replaced by a Poisson model [Schmidt and Mørup, 2013], modeling streamline counts instead. Additional analyses will be needed to identify which model best captures the underlying connectivity. Furthermore, alternative stochastic block models may be used, such as described in a recent study [Andersen et al., 2014] that applies the IRM and two variants to resting-state functional MRI.

Lastly, an interesting avenue to pursue is identifying what causes the uncertainty in the posterior distribution of parcellation. Presumably, this can be attributed to a large extent to noise in data acquisition and tractography. But aside from methodological reasons, there may be other causes that could provide insight in the functional organization of the brain. For instance, uncertainty may in some cases be a result of overlapping clusters [Rubinov and Sporns, 2011; Yeo et al., 2013]. It is likely that regions that are part of multiple, overlapping clusters show higher uncertainty in their cluster assignments. This may be clarified by embedding the infinite relational model within a larger framework to infer clusters at different levels of a nested hierarchy [Meunier et al., 2009; Blundell and Teh, 2013]. One would expect that those clusters that show overlap and uncertainty become merged at a higher level of the hierarchy.

3.4 CONCLUSION

In this chapter, we have described an approach for connectivity-based parcellation that encompasses both community-like clusters as well as clusters that have similar connections, without being densely intraconnected. We find that both kinds of clusters are represented in the human connectome, and that the division into these two types corresponds to previous knowledge of structural connectivity. The model is able to quantify which regions are difficult to assign to a cluster, and it learns the number of clusters from the data. Finally, it does not depend on thresholded connectivity, but derives connectivity simultaneously with the parcellation. We hope that connectivity-based parcellation based on probabilistic models such as the one presented here will help to better understand the structural organization principles of human brain networks.

4

CONSTRAINING FUNCTIONAL CONNECTIVITY WITH STRUCTURE

One of the oldest and most influential ideas in cognitive neuroscience is that the brain, and in particular the cortex, can be divided into specialized functional regions [Friston, 2011]. In recent times, the neuroscience community has become increasingly interested in determining how these regions are organized as large functional networks and how their modulation reflects ongoing cognitive processing [Bullmore and Sporns, 2009]. The organization of these functional networks can be described using the umbrella term 'functional connectivity', defined as the deviations from statistical independence between distributed and often spatially remote neuronal units [Friston, 1994; Craddock et al., 2013]. Despite the indirect nature of the blood oxygenation level dependent (BOLD) signal, functional magnetic resonance imaging (fMRI) has proven to be able to extract patterns of co-activation between clusters of voxels [Lowe et al., 2000].

The easiest way to operationalize the notion of functional connectivity is to calculate a covariance matrix which, in case of standardized variables, is equivalent to the correlation structure between brain regions. However, this approach is not able to identify direct (monosynaptic) functional connections as it is also sensitive to indirect (polysynaptic) functional interactions. For example, if regions A and B as well as regions B and C display correlated activity, then A and C will also show correlated activity even if they are not directly connected [Smith, 2012; Varoquaux and Craddock, 2013].

In contrast, the precision matrix, defined as the inverse of the covariance matrix, captures conditional independence between brain regions [Lauritzen, 1996; Whittaker, 2009]. That is, elements of the precision matrix are related to partial rather than full correlations and zero elements of the precision matrix imply an absence of direct functional connectivity. Therefore, sparse precision matrices provide us with valuable information about how different regions

This chapter is based on: M. Hinne, L. Ambrogioni, R.J. Janssen, T. Heskes and M.A.J. van Gerven, 2014. "Structurally-informed Bayesian functional connectivity analysis." *NeuroImage* 68, pp. 294–305.

interact, though the estimates need to be interpreted with care [Marrelec and Benali, 2009; Friston, 2011; Woolrich and Stephan, 2013; Hutchison et al., 2013].

A common approach to obtain a point estimate for a sparse precision matrix is by means of the graphical lasso [Friedman et al., 2008; Varoquaux et al., 2010; Smith et al., 2011], which achieves sparseness through ℓ_1 regularization. Although the graphical lasso provides a reasonable point estimate, it is biased due to the induced shrinkage of the partial correlations. Furthermore, it does not directly provide a measure of uncertainty regarding the partial correlation estimates. This could lead to possibly erroneous conclusions about functional connectivity.

From a Bayesian perspective we are interested in the posterior density of the precision matrix given observed data. Ultimately, this should lead to more reliable inferences about a subject's cognitive state. In order to facilitate the estimation problem, we will not resort to shrinkage, as in the graphical lasso. Rather, we assume that the conditional independence structure between brain regions is given by an independent estimate of structural connectivity.

Structural connectivity refers to the presence of white matter tracts between spatially segregated brain regions [Hagmann et al., 2008]. In humans, these tracts can be estimated in vivo by diffusion weighted imaging (DWI) which measures the anisotropy in the diffusion of water molecules [Le Bihan et al., 2001]. The final result is usually a binary undirected graph which reports whether or not two areas are structurally connected. Clearly two brain regions can be directly functionally coupled only if they are physically connected, therefore the concepts of functional and structural connectivity are intimately related [Damoiseaux and Greicius, 2009]. The idea is to infer structural connectivity from DWI data and use it as an additional constraint in our Bayesian model. The validity of this approach is supported by several recent experimental studies which found a substantial overlap between structural and functional networks both inside specific cortical areas [Koch et al., 2002] and on a whole brain scale [Hagmann et al., 2008; Honey et al., 2009, 2007; Damoiseaux and Greicius, 2009; Greicius et al., 2009; Cabral et al., 2012]. Related approaches have been used before in the context of functional and effective connectivity analysis [Stephan et al., 2009; Deligianni et al., 2011; Ng et al., 2012].

In the following we present a new Bayesian framework for estimating functional connectivity. The framework, which we refer to as Bayesian functional connectivity (BFC) analysis, makes use of a G-Wishart prior [Roverato, 2002]. This prior allows the sparseness structure of estimated precision matrices to be determined by a graph G, corresponding to structural connectivity. BFC analysis then amounts to computing a posterior density over sparse precision matrices. This posterior may then be used to compute marginal densities for partial correlations of interest. Our approach is compared with existing approaches using both simulated data and empirical data. We show that our approach provides robust partial correlation estimates while at the same time quantifying the uncertainty about functional connectivity.

4.1 MATERIALS AND METHODS

Conventional functional connectivity estimation

Traditionally, functional connectivity estimation has relied on estimating covariance structure between pbrain regions from timeseries data $\mathbf{X} \in \mathbb{R}^{n \times p}$. Here, each vector $\mathbf{x}_i = (x_{i1}, \ldots, x_{ip})$ reflects neuronal activity (e.g. BOLD responses) for p brain regions. Without loss of generality, we assume that data is standardized to have zero mean and unit standard deviation such that covariance coincides with correlation. It is assumed that the data are generated according to a zero-mean multivariate Gaussian density

$$P(\mathbf{X} \mid \mathbf{K}) = \prod_{i=1}^{n} \mathcal{N}(\mathbf{x}_i \mid 0, \mathbf{K}) \propto |\mathbf{K}|^{n/2} \exp\left[-\frac{1}{2}\langle \hat{\mathbf{\Sigma}}\mathbf{K}\rangle\right] \tag{4.1}$$

with precision (inverse covariance) matrix $\mathbf{K} = \mathbf{C}^{-1}$, empirical scatter matrix $\hat{\mathbf{\Sigma}} = \mathbf{X}^\top\mathbf{X}$ and trace operator $\langle\cdot\rangle$. The choice of this distribution is justified by the fact that it is the maximum entropy distribution among all distributions with a specified mean and covariance [Cover and Thomas, 2006]. Alternatively, the likelihood may be characterized in terms of the scatter matrix $\hat{\mathbf{\Sigma}}$ which follows a Wishart distribution $\mathcal{W}_p(\mathbf{C}, n)$ if its density is

$$P(\hat{\mathbf{\Sigma}} \mid \mathbf{C}, n) = \frac{|\hat{\mathbf{\Sigma}}|^{n/2}}{Z(n, \hat{\mathbf{\Sigma}})} \exp\left[-\frac{1}{2}\langle \hat{\mathbf{\Sigma}}\mathbf{C}^{-1}\rangle\right] \quad , \tag{4.2}$$

with $Z(n, \hat{\mathbf{\Sigma}})$ the partition function that normalizes the probability distribution. This perspective can be applied more easily for distributions with a mean different from zero [Anderson, 1984].

We focus on estimating the precision matrix $\mathbf{K} = \mathbf{C}^{-1}$ rather than the covariance matrix. As mentioned before, zero elements in $\mathbf{K}$ reflect the absence of direct interactions. More formally, the sparseness structure of $\mathbf{K}$, represented in terms of an undirected graph G where $V(G)$ is a set of nodes and $E(G)$ is a set of undirected edges between nodes, is equivalent to the conditional independence structure of a Gaussian Markov random field [Lauritzen, 1996; Whittaker, 2009]. In other words, in the context of connectivity analysis, $k_{ij} = 0$ corresponds to the absence of structural connectivity between brain regions i and j.

In order to estimate the precision matrix $\mathbf{K}$ of a zero-mean multivariate Gaussian density from data $\mathbf{X}$ one may maximize the log likelihood

$$\log P(\mathbf{X} \mid \mathbf{K}) = \frac{1}{2}\left[n \log |\mathbf{K}| - \langle \hat{\mathbf{\Sigma}}\mathbf{K}\rangle\right]$$

which gives the maximum likelihood estimate (MLE):

$$\hat{\mathbf{K}} = \arg\max_{\mathbf{K} \in \mathbb{P}_p} \left[n \log |\mathbf{K}| - \langle \hat{\mathbf{\Sigma}}\mathbf{K}\rangle\right] = n\hat{\mathbf{\Sigma}}^{-1} \tag{4.3}$$

where the maximization is constrained to precision matrices in the family of $p \times p$ positive definite matrices $\mathbb{P}_p$.

In practice, however, this empirical estimate does not contain zero elements. Furthermore, in case $n < p$, the maximum likelihood solution does not exist since $\hat{\mathbf{\Sigma}}/n$ becomes singular. Even in case $n > p$, the MLE is often poorly behaved, and regularization is called for [Pourahmadi, 2011]. The graphical lasso [Friedman et al., 2008] regularizes the preceding MLE through sparsification by solving

$$\hat{\mathbf{K}} = \arg\max_{\mathbf{K} \in \mathbb{P}_p} \left[\log|\mathbf{K}| - \frac{1}{n}\langle \hat{\mathbf{\Sigma}}\mathbf{K} \rangle - \lambda \|\mathbf{K}\|_1 \right] . \tag{4.4}$$

The employed ℓ_1 regularizer encourages sparse precision matrices as determined by the regularization parameter λ. This maximization problem can be solved using established coordinate descent methods [Friedman et al., 2008]. The graphical lasso has been proposed as the method of choice for functional connectivity estimation [Varoquaux et al., 2010; Smith et al., 2011; Varoquaux and Craddock, 2013].

Even though the graphical lasso is commonly used to estimate sparse precision matrices, it suffers from two issues. First, since the graphical lasso employs shrinkage, pushing precision values towards zero, the resulting functional connectivity estimate is biased. Second, the graphical lasso produces a point estimate which does not directly allow inferences to be drawn about the uncertainty in our estimates arising from sampling noise and finite sample size.

Bayesian functional connectivity estimation

To tackle the aforementioned issues, we developed a Bayesian framework for inferring functional connectivity which does not rely on shrinkage but rather assumes that the sparseness structure G of $\mathbf{K}$ is given [Dempster, 1972]. Specifically, we assume that the graph G is given by the structural connectivity as estimated from DWI data.

We start by assuming a G-Wishart distribution as the conjugate prior on precision matrices $\mathbf{K}$. The G-Wishart is defined for the cone $\mathbb{P}_G$ of positive-definite symmetric matrices with off-diagonal elements $k_{ij} = 0$ whenever $(i, j) \notin E(G)$. A zero-constrained random matrix $\mathbf{K}$ has the G-Wishart distribution $\mathcal{W}_G(\delta_0, \mathbf{D})$ if its density is [Wang and Li, 2012]:

$$P(\mathbf{K} \mid G) = \frac{|\mathbf{K}|^{(\delta_0 - 2)/2}}{Z_G(\delta_0, \mathbf{D})} \exp\left(-\frac{1}{2}\langle \mathbf{D}\mathbf{K} \rangle \right) 1_{\{\mathbf{K} \in \mathbb{P}_G\}}$$

where δ are the prior degrees of freedom, $\mathbf{D}$ a symmetric positive definite prior scatter matrix, and $Z_G(\delta, \mathbf{D})$ the normalizing constant. The indicator function 1_x evaluates to 1 if its argument x is true and to 0 if its argument is false. In our experiments, we set $\delta_0 = 3$ and choose $\mathbf{D} = \mathbf{I}_{p \times p}$ [Moghaddam et al., 2009]. This amounts to a vague prior for the precision matrix

in (4.1), except that its support is restricted by G. We may now use Bayes' rule to obtain the posterior density for $\mathbf{K}$ according to

$$\begin{aligned} P(\mathbf{K} \mid \mathbf{X}, G) &\propto P(\mathbf{X} \mid \mathbf{K}) P(\mathbf{K} \mid G) \\ &= \frac{|\mathbf{K}|^{(\delta_n - 2)/2}}{Z_G(\delta_n, \mathbf{B})} \exp\left(-\frac{1}{2}\langle \mathbf{B}\mathbf{K} \rangle\right) 1_{\{\mathbf{K} \in \mathbb{P}_G\}} \ . \end{aligned} \tag{4.5}$$

Because the G-Wishart prior is conjugate to the likelihood, the resulting posterior once again follows a G-Wishart distribution with $\delta_n = \delta_0 + n$ the posterior degrees of freedom and $\mathbf{B} = \mathbf{D} + \hat{\mathbf{\Sigma}}$ the posterior scatter matrix.

In order to approximate this posterior density we employed an efficient edgewise block Gibbs sampler described in detail in Wang and Li [2012]. The algorithm is similar to a Bayesian iterative proportional scaling algorithm, but instead of updating $\mathbf{K}$ in large blocks (e.g. maximum cliques in G, as is used in [Lenkoski and Dobra, 2011]), it is updated per edge. This can be done efficiently, as the authors show that this only requires 2×2 matrix inversions. Source code implementing the Gibbs sampler can be freely obtained from the author's website.[1]

In practice, we find it more convenient to express functional connectivity in terms of partial correlation rather than precision. This can be easily achieved via the following transformation

$$\rho_{ij \mid W} = \begin{cases} -\frac{k_{ij}}{\sqrt{k_{ii} k_{jj}}} & \text{for } i \neq j \\ 1 & \text{for } i = j \end{cases} \tag{4.6}$$

reflecting the correlation between brain regions i and j when we condition on all other brain regions $W = \{1, \ldots, n\} \setminus \{i, j\}$. Observe that, given this transformation, the interpretation in terms of conditional independence still holds. Hence, in practice, we use the posterior density $P(\mathbf{R} \mid \mathbf{X}, G)$ for the partial correlation matrix $\mathbf{R}$ as our estimate of Bayesian functional connectivity. This density is computed by applying the transformation (4.6) to each of the samples of the precision matrix as generated by the Gibbs sampler.

Experimental validation

We compared conventional functional connectivity analysis with Bayesian functional connectivity analysis using both simulated data and empirical data.

For the conventional analysis, we examined the maximum likelihood estimate (4.3) and the graphical lasso (4.4). In order to obtain an optimal estimate for the graphical lasso, we used a five-fold cross-validation procedure where the log likelihood computed on hold-out data was used to fine-tune the regularization parameter, as proposed in [Friedman et al., 2008].

For the BFC analysis, we computed the posterior density using Gibbs sampling, where we discarded the first 2 000 burn-in samples and stored the subsequent 5 000 samples while using the transformation (4.6). The mode of this posterior (maximum a posteriori (MAP) estimate) was

[1] http://www.stat.sc.edu/~wang345

TABLE 4.1: Left-hemisphere AAL regions used in functional connectivity analysis. Right-hemisphere AAL regions 46–90 follow the same labeling.

#	*region*	#	*region*	#	*region*
1	precentral	16	cingulum_ant	31	parietal_inf
2	frontal_sup	17	cingulum_mid	32	supramarginal
3	frontal_sup_orb	18	cingulum_post	33	angular
4	frontal_mid	19	hippocampus	34	precuneus
5	frontal_mid_orb	20	parahippocampal	35	paracentral_lobule
6	frontal_inf_oper	21	amygdala	36	caudate
7	frontal_inf_tri	22	calcarine	37	putamen
8	frontal_inf_orb	23	cuneus	38	pallidum
9	rolandic_oper	24	lingual	39	thalamus
10	supp_motor	25	occipital_sup	40	heschl
11	olfactory	26	occipital_mid	41	temporal_sup
12	frontal_sup_medial	27	occipital_inf	42	temporal_pole_sup
13	frontal_mid_orb	28	fusiform	43	temporal_mid
14	rectus	29	postcentral	44	temporal_pole_mid
15	insula	30	parietal_sup	45	temporal_inf

used to compare with point estimates obtained using conventional analyses. Both this MAP estimate as well as the graphical lasso estimates were computed using fast optimization procedures due to Schmidt [2010].[2]

In the remainder of this section, we describe the empirical data which were used to validate our approach.

Data acquisition

Six healthy volunteers were scanned after giving informed written consent in accordance with the guidelines of the local ethics committee. These subjects represent a subset of the data previously used in Chapter 2. T1 anatomical scans, resting-state functional images and diffusion-weighted images were obtained using a Siemens Magnetom Trio 3T system at the Donders Centre for Cognitive Neuroimaging, Radboud University Nijmegen, The Netherlands. Resting-state fMRI data were acquired at 3 Tesla by using a multi-echo echo-planar imaging (ME-EPI) sequence (voxel size 3.5 mm isotropic, matrix size 64×64, TR=2 000 ms, TEs=6.9, 16.2, 25, 35 and 45 ms, 39 slices, GRAPPA factor 3, 6/8 partial Fourier). A total of 1030 volumes were obtained. An optimized acquisition order described by Cook et al. [2007] was used in the DWI

[2] http://www.di.ens.fr/~mschmidt/Software

protocol (voxel size 2.0 mm isotropic, matrix size 110×110, TR=13 000 ms, TE=101 ms, 70 slices, 256 directions at b=1500 s/mm^2 and 8 images at b=0 s/mm^2).

Tools from the Oxford FMRIB Software Library (FSL, FMRIB, Oxford, UK) were used to preprocess the data. For each subject brain extraction from the anatomical scan was performed using FSL BET [Smith, 2002], tissue segmentation was performed using FSL FAST [Zhang et al., 2001] and subcortical structures were segmented using FSL FIRST [Patenaude et al., 2011].

Preprocessing of diffusion imaging data

DWI data were preprocessed using FSL FDT [Behrens et al., 2003a] and consisted of correction for eddy currents and estimation of the diffusion parameters. A measure of white matter connectivity was obtained by FDT Probtrackx 2.0 [Behrens et al., 2003a, 2007]. The voxels that live on the boundary between white matter and gray matter were used as seeds for tractography, gray matter was used as a target mask. For each seed voxel 5 000 streamlines were drawn with a maximum length of 2 000 steps. The streamlines were restricted by the fractional anisotropy to prevent them from wandering around in gray matter. Streamlines in which a sharp angle ($> 80°$) occurred were discarded.

This procedure resulted in a matrix providing voxel-to-voxel connectivity for each subject s. Each matrix was reduced to a 90×90 streamline count matrix $\mathbf{S}^{(s)}$ by summing the streamline counts over voxels belonging to 90 distinct brain regions, as defined by the automatic anatomical labelling (AAL) atlas [Tzourio-Mazoyer et al., 2002] while excluding cerebellar structures. The regions defined by the AAL template are displayed in Table 4.1 for ease of reference. Finally, edges of the structural connectivity graph G were determined by the streamline count matrices as follows:

$$(i,j) \in E(G) \leftrightarrow \bigcap_{s\in\{1,\ldots,6\}} (s_{ij}^{(s)} > 0 \wedge s_{ji}^{(s)} > 0)\,. \tag{4.7}$$

This particular definition corresponds to the following intuition: at the coarse scale that is given by the AAL template, we require that each subject shows support for a particular connection. Per subject we are lenient and threshold at zero. Alternatively, using the union of the thresholded networks for each subject would result in extremely dense estimates. The structural graph G was used as a constraint for Bayesian functional connectivity analysis.

Preprocessing of functional data

Multi-echo images obtained using the rs-fMRI acquisition protocol were combined using a custom Matlab script (MATLAB 7.7, The MathWorks Inc., Natick, MA, USA) which implements the procedure described in [Poser et al., 2006] and also incorporates motion correction using functions from the SPM5 software package (Wellcome Department of Imaging Neuroscience, University College London, UK). Of the 1030 combined volumes, the first six were discarded to allow the system to reach a steady state. The data was then factorized into 77 independent

components using FSL MELODIC [Beckmann and Smith, 2004], of which 37 ± 1 were retained by manually removing artifact components. After preprocessing, the fMRI data were parcellated according to the AAL atlas. For these regions, the functional data was summed and standardized to have zero mean and unit standard deviation. The resulting BOLD timeseries for all six subjects were used in subsequent functional connectivity analyses.

4.2 RESULTS

Simulation study

We performed a simulation study in order to compare the performance of the G-Wishart approach with the maximum likelihood estimate and the graphical lasso. We assumed the structural estimate defined in (4.7) as the ground truth. Using this structure, 100 random precision matrices $\mathbf{K}_s$ were drawn from the posterior $p(\mathbf{K}|\mathbf{X}, G)$ with $\delta_n = 3 + n$ and $\mathbf{B} = \mathbf{I}_{p\times p} + \hat{\mathbf{\Sigma}}$, with $\hat{\mathbf{\Sigma}}$ the mean covariance matrix of the six subjects. The consequence of this approach was that the used precision matrices were ensured to be positive definite, follow the structure of G, and have comparable conditioning as the MAP estimates in our experiments using real data. From the sampled precision matrices $\mathbf{K}_s$, an $n \times p$ data matrix $\mathbf{X}_s \sim \mathcal{N}(0, \mathbf{K}_s^{-1})$ was generated, with $n = \{128, 256, \ldots, 4\,096\}$. For each ground truth precision matrix $\mathbf{K}_s$ and the different sample sizes, we created reconstructions $\mathbf{K}_r$ of $\mathbf{K}_s$ using four strategies: the graphical lasso estimate, the maximum likelihood estimate (MLE), the maximum a posteriori (MAP) estimate using the G-Wishart prior and G_s as structural estimate and the MAP estimate using the Wishart prior, which corresponds to a G-Wishart prior with a full graph. An example precision matrix, as well as the different reconstructions, is shown in Figure 4.1A.

We quantified the quality of the reconstruction of the ground truth precision using the Kullback-Leibler (KL) divergence:

$$D_{KL}(\mathbf{K}_s \| \mathbf{K}_r) = \frac{1}{2\log 2}\left[\log \frac{|\mathbf{K}_s|}{|\mathbf{K}_r|} + \langle \mathbf{K}_r \mathbf{K}_s^{-1} \rangle - p\right].$$

The results of the simulation are provided in Figure 4.1B. The figure shows that compared to the maximum likelihood estimator and the MAP estimate using the Wishart distribution, the G-Wishart approach does not require many samples in order to reconstruct $\mathbf{K}_s$. Also, it outperforms the graphical lasso estimates. However, this critically depends on the quality of G. To see how well the G-Wishart MAP estimate approximates the ground truth, we ran additional simulation runs in which the structural graph G was increasingly perturbed. We considered three cases. In the first, the graph is rewired by removing edges at random and simultaneously connecting randomly selected disconnected nodes, thus keeping the density of G constant. The results of this rewiring are shown in Figure 4.1C. In the second and third case we either removed edges, representing false negatives, or we added edges, representing false positives. The results of these perturbations are shown in Figure 4.1D. The results show that the performance of the G-Wishart approach deteriorates linearly with increasing rewiring. In the low-powered setting

of $n = 128$, we observe that the approach outperforms the alternatives for up to 20% of the edges replaced, but for the high-powered setting of $n = 4\,096$ this percentage drops to 5%. For false negative errors in the structural estimate, the number of samples has a clear effect on the performance. For few samples, the effect of mere sparsity keeps the Kullback-Leibler divergence lower than the MLE or the Wishart MAP, although the graphical lasso obviously performs better than the (near) empty structural graph. Once more samples are acquired, false negatives have an even stronger negative influence on the performance. Less troublesome are false positives, for which we observe the converse behavior. For few samples, false positives provide difficulty for the G-Wishart approach, but once a large number of samples is obtained, an increase in density of the structural estimate of up to 90% still shows the G-Wishart approach as the most successful.

Finally, we considered how the posterior density of partial correlation changes depending on the number of samples and the quality of the provided structural estimate. We considered three cases: the estimation of partial correlation for a ground truth edge using G, the same estimation using a perturbed graph with 20% of the edges replaced, but with the ground truth edge still intact and finally the estimation of partial correlation for an edge that was absent in the ground truth, yet is present in the perturbed graph. We used the same structural estimate for G as before, and used one random sampled precision matrix as ground truth, from which data was generated for $n = \{128, 1\,024, 4\,096\}$. As an example, Figure 4.2A and B show the estimated partial correlation between the supplementary motor areas for the true graph and the perturbed graph that contains this connection, respectively. Notably, the distributions become more narrow when more samples are acquired (as expected). In Figure 4.2C the partial correlation between the left cuneus and the right temporal pole is shown. Note that this connection is not in our structural estimate. The distributions show that for erroneous edges (i.e. false positives), the G-Wishart approach correctly estimates a partial correlation of zero.

Empirical validation

We now turn to the empirical validation where we used a structural connectivity graph G as estimated from DWI data to constrain the correlation structure which explains resting-state fMRI data.

Estimated structural connectivity graph

The structural connectivity as estimated via the procedure described in (4.7) is shown in Figure 4.3, separated into intra-hemispheric connectivity in left and right hemispheres as well as interhemispheric connectivity between left and right hemispheres. The region numbers correspond to those displayed in Table 4.1. A total of 774 structural links were identified of which 329 were left intrahemispheric connections, 328 were right intrahemispheric connections and 117 were interhemispheric connections. This gives a network density of 0.19. Note that his is denser than the density of 0.11 reported by Gong et al. [2009]. However, for our purposes, the

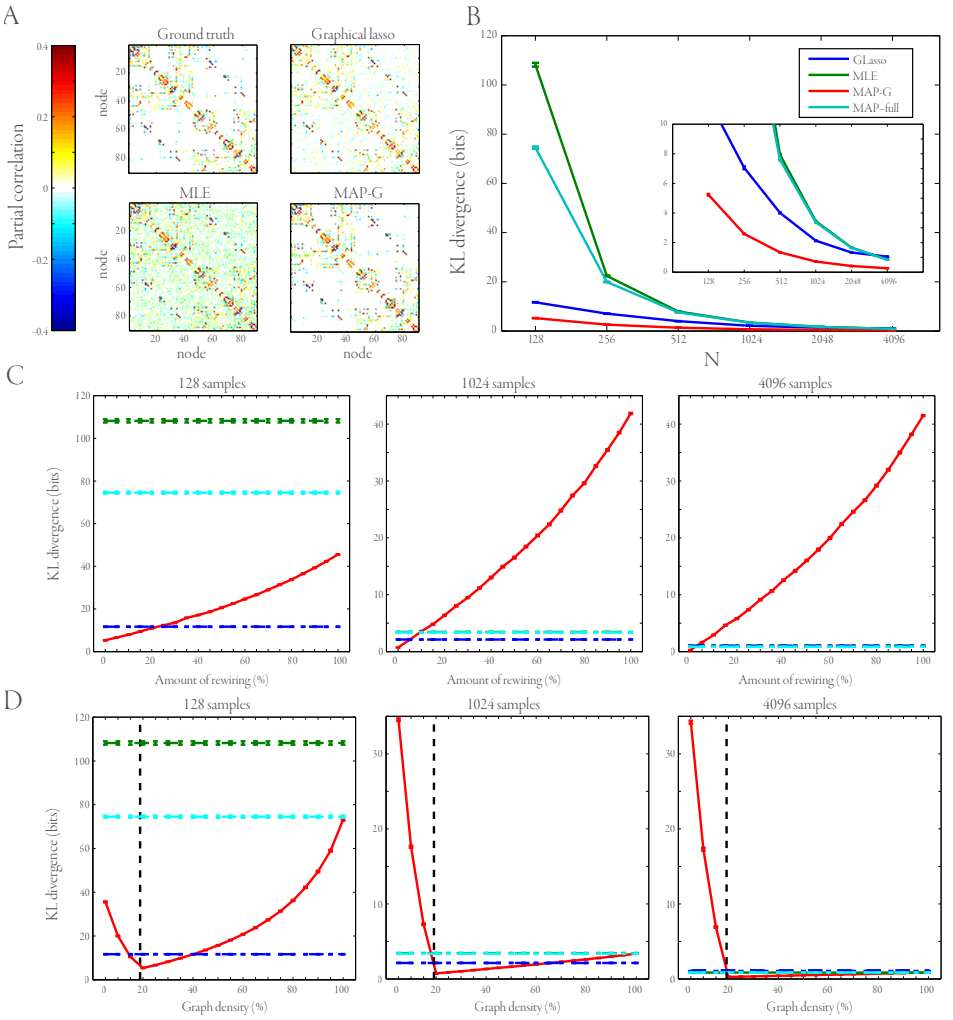


FIGURE 4.1: A. The simulated precision matrix $\mathbf{K}_S$, and reconstructions $\mathbf{K}_T$ using the graphical lasso, MLE and MAP-G solution, for $n = 1024$. Note that the diagonal elements have been set to zero to increase the visibility of the differences between the off-diagonal elements. The reconstruction for the MAP estimate using a full graph has been omitted, as it was indistinguishable from the MLE solution. B. The Kullback-Leibler divergence for the four different methods as a function of the number of samples. The inset shows a detailed view. C. The Kullback-Leibler divergence for the four different methods as a function of the amount of rewiring applied to the ground truth graph. Note that the MLE, the graphical lasso and the MAP estimate using the full graph are unaffected by this rewiring; the dotted lines are provided as a visual aid. D. The Kullback-Leibler divergence for the four different methods as a function of false positives and false negatives. The vertical dashed line indicates the actual density. The effect of false negatives is shown on the left side of this line (until the graph is near empty) whereas the right side shows the effect of false positives (until the graph is the complete graph). Similar to C, the other methods are unaffected by the changes in G and are only shown for convenience.

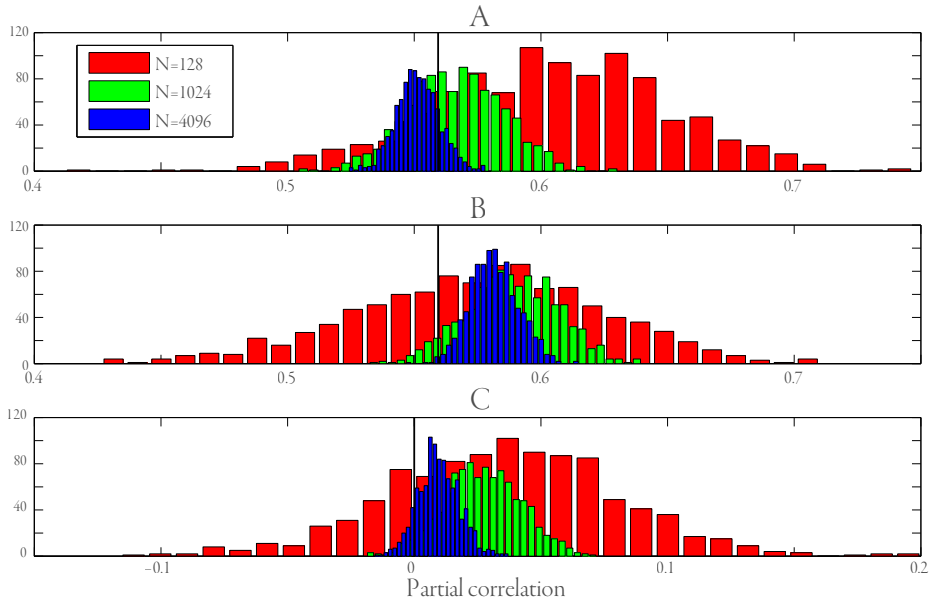


FIGURE 4.2: A. Partial correlation for an example true edge. B. Partial correlation using a perturbed graph that still contained this edge. C. Partial correlation for an example edge that was absent in the ground truth. The true partial correlation is indicated with a vertical black line.

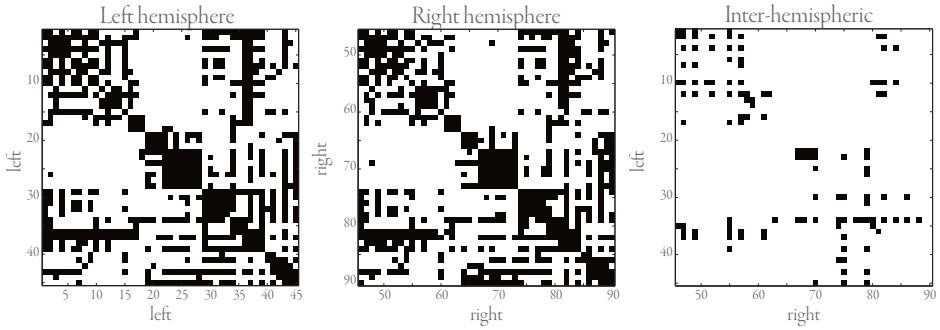


FIGURE 4.3: Connectivity of the structural connectivity matrix used as a constraint in BFC estimation. Submatrices representing intrahemispheric and interhemispheric connectivity are shown separately here for ease of reference. Black (white) squares indicate the presence (absence) of a structural connection.

BFC analysis can still provide evidence against edges which were erroneously included in the structural graph as we saw in the simulation.

TABLE 4.2: Five interhemispheric and intrahemispheric connections with highest mean partial correlations. Hemispheres H are indicated by 'L' (left hemisphere) and 'R' (right hemisphere). Mean and standard deviation of the partial correlation, averaged over subjects, are denoted by μ and σ, respectively. Logarithm of the total number of streamlines between regions, averaged over subjects, is denoted by n.

#	*region 1*	H	*region 2*	H	μ	σ	$\log n$
1	supp_motor_area	R	supp_motor_area	L	0.60	0.018	9
2	paracentral_lobule	R	paracentral_lobule	L	0.59	0.018	7
3	cingulum_ant	R	cingulum_ant	L	0.57	0.018	7
4	precuneus	R	precuneus	L	0.57	0.016	7
5	caudate	R	caudate	L	0.56	0.016	4
1	frontal_mid	L	frontal_sup	L	0.65	0.019	10
2	pallidum	R	putamen	R	0.60	0.017	10
3	pallidum	L	putamen	L	0.59	0.016	11
4	frontal_inf_tri	L	frontal_inf_oper	L	0.56	0.019	9
5	frontal_inf_tri	R	frontal_inf_oper	R	0.56	0.019	9

Bayesian functional connectivity analysis

BFC analysis was performed on the resting-state data for each of the six subjects and amounted to computing posterior densities $p(\mathbf{R} \mid \mathbf{X}, G)$. Figure 4.5 shows the mean partial correlations between all brain regions for each of the six subjects. Mean partial correlations between regions ranged from -0.43 to 0.79. Furthermore, partial correlation estimates had standard deviations ranging up to 0.04, illustrating the fact that absolute certainty cannot be achieved using a finite amount of data. It can be observed that functionally homologous regions in the left and right hemispheres are strongly partially correlated.

In order to gain more insight into the partial correlations between particular regions, we focus on connections with either high or low mean partial correlations. Since qualitatively different behavior can be observed for inter- and intra-hemispheric connections, we handle them separately.

Table 4.2 depicts five interhemispheric connections and five intrahemispheric connections which showed the strongest partial correlations. Evidence for non-zero partial correlations between the regions identified in Table 4.2 is supported by literature. Specifically, partial correlations between supplementary motor areas [Salvador et al., 2005], cingulate cortices [Salvador et al., 2005], as well as prefrontal areas [Damoiseaux et al., 2006; Beckmann et al., 2005], have been identified before. The strong partial correlation between putamen and pallidum can be explained by monosynaptic feedforward and feedback pathways between the striatum and the globus pallidus [Bolam et al., 2000; Smith and Bolam, 1990]. Structural connectivity between left and right precuneus is also supported by anatomical studies [Cavanna and Trimble, 2006].

Table 4.3: Five interhemispheric and intra-hemispheric connections with lowest mean partial correlations. Hemispheres H are indicated by 'L' (left hemisphere) and 'R' (right hemisphere). Mean and standard deviation of the partial correlation, averaged over subjects, are denoted by μ and σ, respectively. Logarithm of the total number of streamlines between regions, averaged over subjects, is denoted by n.

#	*region 1*	H	*region 2*	H	μ	σ	$\log n$
1	frontal_sup_medial	R	putamen	L	0.02	0.016	5
2	putamen	R	frontal_mid	L	0.02	0.014	4
3	parietal_sup	R	putamen	L	0.02	0.014	2
4	putamen	R	parietal_sup	L	0.02	0.014	2
5	putamen	R	precuneus	L	0.02	0.015	4
1	putamen	L	calcarine	L	0.01	0.009	5
2	caudate	L	precuneus	L	0.02	0.013	4
3	putamen	R	lingual	R	0.02	0.017	5
4	putamen	L	cingulum_ant	L	0.02	0.019	5
5	lingual	R	frontal_sup_orb	R	0.02	0.007	4

Likewise, evidence exists for direct connections between left and right caudate [Mensah and Deadwyler, 1974; Medina and Pazo, 1981].

Table 4.3 shows five interhemispheric connections and five intrahemispheric connections which showed the weakest partial correlations. A comparison between Tables 4.2 and 4.3 suggests that connections with strong partial correlations are supported by many more streamlines compared to the connections that showed weak partial correlations. This is confirmed by highly significant positive correlations between streamline count and absolute partial correlation values ($\rho > 0.40, p < 10^{-31}$ in all subjects). This relationship has been observed before [Hermundstad et al., 2013]. This indicates that there is a clear correspondence between the information conveyed by the structural connectivity as estimated using DWI and the Bayesian functional connectivity analysis applied to resting-state fMRI data.

Comparison with the graphical lasso

While the Bayesian approach to functional connectivity analysis has been shown to lead to interpretable estimates of partial correlation while at the same time quantifying the uncertainty in these estimates, it is important to compare its behavior with conventional approaches to functional connectivity analysis. In the following, we compare BFC estimates with estimates produced by the graphical lasso. Note that the cross-validation procedure used to select the value of the regularization parameter for the graphical lasso led to quite dense partial correlation matrices. On average, only 53 region pairs were estimated to have zero partial correlation.

Figure 4.4A shows the whole-brain point estimates produced by the graphical lasso. A comparison of these estimates with those obtained by BFC analysis, as depicted in Figure 4.5, in-

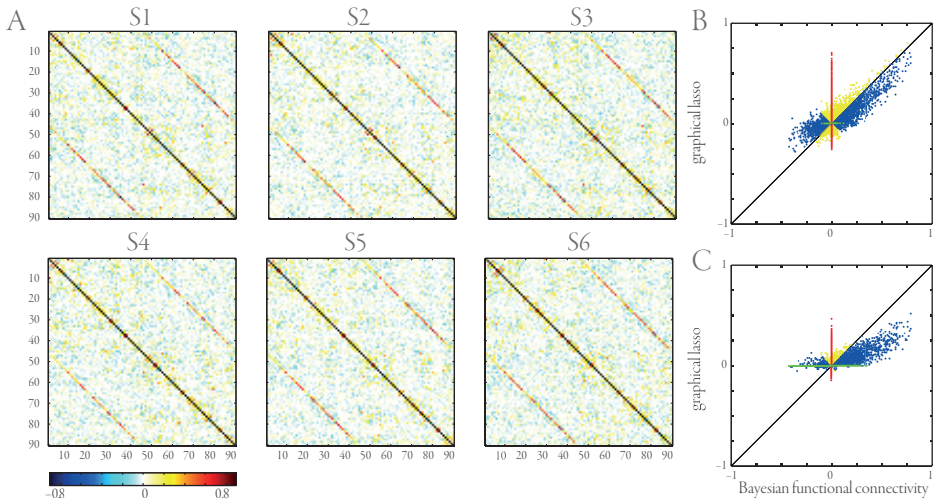


FIGURE 4.4: Graphical lasso results. (A) Graphical lasso functional connectivity estimates converted to partial correlations for six subjects. (B) Scatter plot comparing partial correlations for all six subjects as estimated using BFC analysis (mean partial correlation) or the graphical lasso. For the graphical lasso, a point estimate was obtained using nested cross-validation. Blue datapoints denote non-zero partial correlations that are larger in magnitude for BFC than for the graphical lasso. Yellow datapoints denote non-zero partial correlations that are lower in magnitude for BFC than for the graphical lasso. Red datapoints denote partial correlations which were zero for BFC analysis and non-zero for the graphical lasso. Green datapoints denote partial correlations which were zero for the graphical lasso and non-zero for BFC analysis. (C) Same as (B) but now with graphical lasso estimates obtained with λ fixed to 0.15.

dicate that both approaches show some correspondence in terms of block diagonal structure and strong partial correlations between interhemispheric functionally homologous areas. At the same time, results show that the estimated partial correlations tend to be weaker for the graphical lasso.

Figure 4.4B shows a scatter plot comparing partial correlation estimates obtained using both approaches. The clouds of blue and yellow datapoints indicate that the graphical lasso leads to smaller partial correlation estimates as expected by the shrinkage property. This is confirmed by the fact that 72% of the connections according to G have lower absolute partial correlations for the graphical lasso compared to the Bayesian approach. Green datapoints show non-zero partial correlations for BFC analysis which were forced to zero by the graphical lasso. Red datapoints signify non-zero partial correlations for the graphical lasso which were forced to be zero according to the structural graph G. The two outliers with non-zero partial correlations above 0.7 according to the graphical lasso reflect connectivity between left and right posterior cingulate cortex in two subjects, which has been reported previously [Salvador et al., 2005].

An alternative to using cross-validation for selecting the regularization parameter λ is to tune it such that a predetermined network density is achieved. Figure 4.4C compares partial correlations between BFC analysis and graphical lasso estimates obtained with λ set to 0.15. This gives much sparser point estimates that more closely resemble the structural graph G. Two main

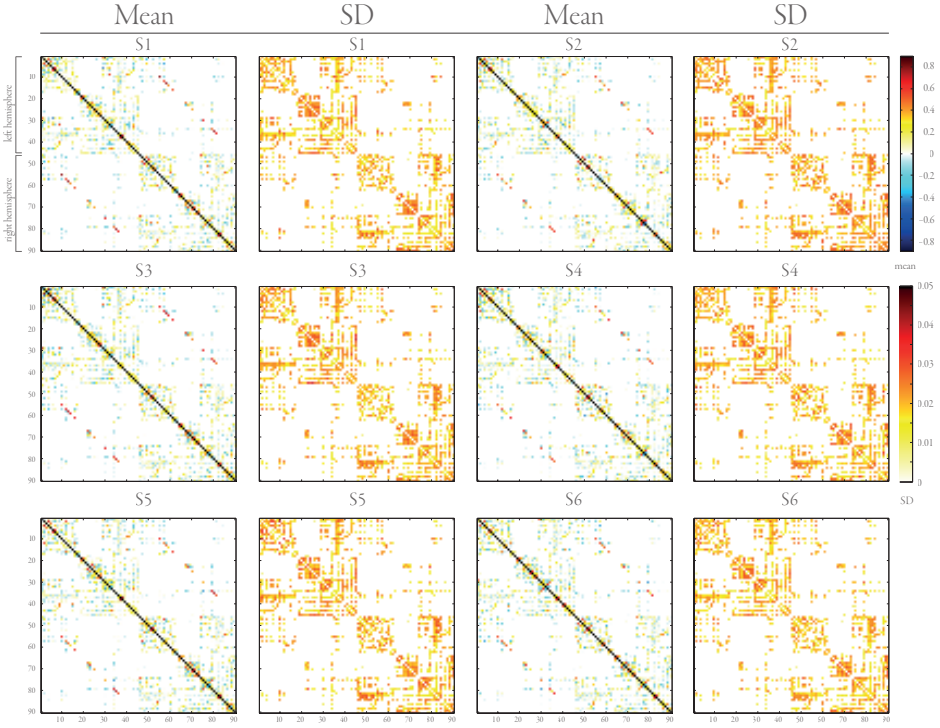


FIGURE 4.5: Whole-brain BFC estimates. Mean and standard deviation of the partial correlation matrices are shown for each of the six subjects.

effects can be observed. First, with increased λ partial correlations are much more affected by shrinkage. That is, 94% of the connections given by the structural graph G have lower absolute partial correlations for the graphical lasso compared to the Bayesian approach. Second, there is a substantial increase in the number of connections implied by the structural graph that are set to zero.

In order to quantify the fit of the different models to the data we used a hold-out scheme where the first 512 samples were used as training data, and the remaining samples as test data. If we compare the G-Wishart approach, the cross-validated graphical lasso and the graphical lasso with sparsity matched with G, we find a log-likelihood on the test data of $-4.81 \cdot 10^4 \pm 0.48 \cdot 10^4$, $-5.90 \cdot 10^4 \pm 0.11 \cdot 10^4$ and $-4.36 \cdot 10^4 \pm 0.18 \cdot 10^4$, respectively. From these results we conclude that the sparsity of G is valuable for correct estimation of partial correlations, but that the structural estimate we used was sub-optimal.

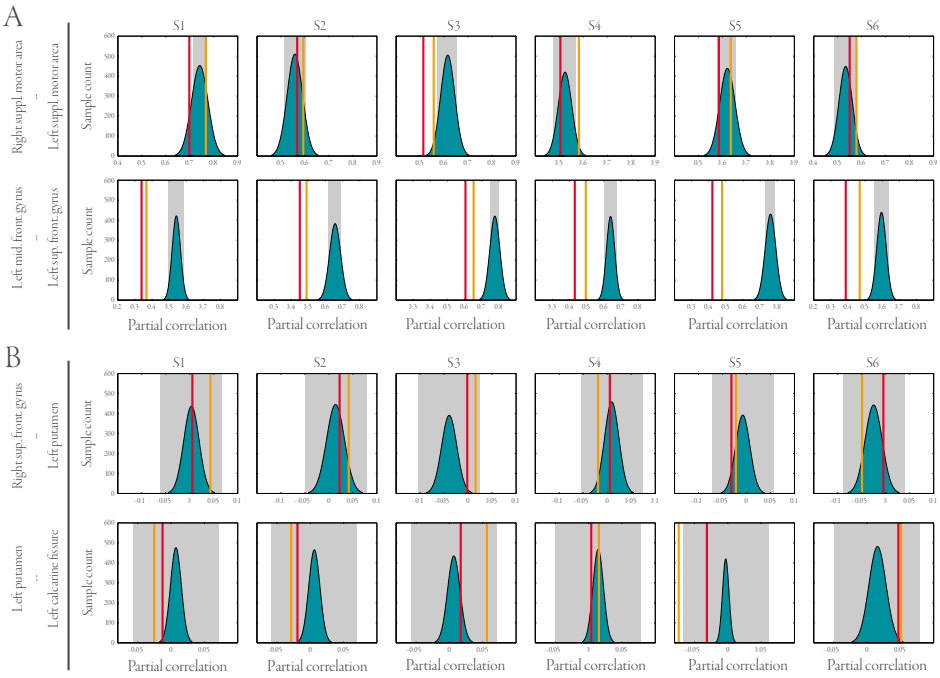


FIGURE 4.6: Posterior densities for individual partial correlations as estimated by BFC analysis (blue), point estimates produced by the graphical lasso (red) and the MLE (yellow). The shaded areas show the 95% confidence interval around the MAP estimate. (A) Estimates obtained for the inter- and intra- hemispheric connections with the strongest partial correlation. (B) Estimates obtained for the inter- and intrahemispheric connections with the weakest partial correlation.

Posterior densities

In order to gain additional insight into the posterior densities estimated by BFC analysis and to allow further comparison with the point estimates produced by the graphical lasso, Figure 4.6 shows the estimates in each subject for the inter- and intrahemispheric connections with strongest and weakest partial correlation according to the BFC analysis. For the strong partial correlations, shown in Figure 4.6A, the graphical lasso estimates tend to be much smaller than the estimates obtained using the BFC approach. In part, this is likely due to the shrinkage induced by the graphical lasso. MLE estimates are also shown for comparison. The weak partial correlations shown in Figure 4.6B are especially interesting. While the graphical lasso makes binary statements about the presence or absence of a non-zero partial correlation, Bayesian functional connectivity analysis quantifies the uncertainty in our estimates, thereby providing a more nuanced view.

For comparison, Figure 4.6 shows the 95% confidence intervals around the MAP estimates of the distribution, which were obtained by applying the Fisher transform to the estimates [Fisher, 1915] with $n - 3 - p = 1\,024 - 3 - 90 = 931$ degrees of freedom. The distributions and

intervals show that for high partial correlations, the 95% confidence interval is enclosed in the MAP distribution, while for low partial correlation the confidence interval is wider than the (bulk of the) distribution.

4.3 DISCUSSION

We proposed Bayesian functional connectivity analysis as a new approach for analyzing the interactions between BOLD timeseries in multiple brain regions. The approach produces a posterior density for the partial correlation matrix and relies on the use of a G-Wishart distribution. This distribution has been used extensively for analyzing covariance structure in high-dimensional problems in biology [Jones et al., 2005], economics [Carvalho et al., 2007] and epidemiology [Dobra and Lenkoski, 2011]. In a neuroscientific context, we use the G-Wishart distribution as a prior which constrains the partial correlation matrix via a structural graph G, as derived from diffusion imaging data.

As demonstrated by both simulated and empirical data, Bayesian functional connectivity analysis based on the G-Wishart is a promising approach for functional connectivity analysis in cognitive neuroscience. The main advantages of our approach compared to existing approaches are as follows. First, we constrain the functional connectivity estimates using a structural graph G as estimated from diffusion imaging data, thereby effectively achieving multi-modal data fusion [Biessmann et al., 2011; Groves et al., 2011]. The constraints imposed by G alleviate the need to impose shrinkage, thereby reducing bias in the partial correlation estimates. Second, we can quantify the uncertainty in our estimates, which allows sound inferences to be drawn about the presence of non-zero partial correlations between BOLD timecourses for multiple regions of interest.

Results based on simulated data show that the G-Wishart approach outperforms the maximum likelihood estimate, the normal Wishart MAP estimate and the graphical lasso for limited data (cf. Figure 4.1B). As more data becomes available the different approaches converge to the same estimate. Note however that these results are obtained when the G-Wishart approach has access to the ground truth graph G. Figure 4.1C and Figure 4.1D show that as the structural graph is perturbed more and more, eventually, the G-Wishart is outperformed by the MLE and the graphical lasso.

Empirical results show that the connections with high partial correlations correspond to anatomical tracts that are known from literature. In our empirical comparison with the graphical lasso, we observed salient differences between respective functional connectivity estimates. This can be due to a number of reasons. First, we used cross-validation for the graphical lasso to select the regularization parameter λ. This led to much denser precision matrices and, hence, different functional connectivity estimates. Second, the graphical lasso employs shrinkage, which can explain in part the observation that graphical lasso partial correlation estimates are typically lower than those obtained with the Bayesian approach. As shown in Figure 4.4C, an alternative to cross-validation is to set the regularization parameter to a fixed value, achieving a desired network density. However, in this case, many connections that are implied by the structural graph and achieve relatively high partial correlations under the Bayesian approach will be set

to zero. Furthermore, the shrinkage effect becomes stronger, thereby further underestimating partial correlation values. By comparing the model likelihood on hold-out data for either the G-Wishart approach, the cross-validated graphical lasso and the graphical lasso with a fixed density, we find that the graphical lasso algorithm with fixed density performs best, followed by the G-Wishart approach. Note however, that the fixed lasso density was determined by the structure G and that crossvalidation, which would be a standard way to estimate the lasso shrinkage, performed worse. Most likely, improvements in the structural estimate will increase the G-Wishart performance.

In order to constrain the estimation problem, BFC analysis makes use of a structural graph as estimated from diffusion imaging data. Drawn conclusions therefore critically depend on the quality of the employed structural graph. As shown in Figure 4.1C and Figure 4.1D, errors in the estimated structural matrix can produce serious biases as connections between regions that show high partial correlation can be completely ruled out from the analysis. In other words, false negatives in the structural estimate strongly (negatively) influence the usefulness of the G-Wishart approach. For example, we observed that the potential connection between posterior cingulate cortices (which was found to have a high partial correlation in two subjects when using the graphical lasso) was absent in our estimates. In order to prevent the exclusion of important connections due to biases associated with probabilistic streamlining [Dauguet et al., 2007; Li et al., 2012], a more lenient threshold may be used to estimate the structural graph, preventing false negatives at the expense of false positives. In addition, connections known to be absent may simply be ruled out in the prior. Notwithstanding these caveats, the proposed methodology is theoretically sound and may take advantage of various future developments in diffusion imaging [Dell'Acqua and Catani, 2012].

A potential alternative to the multimodal approach followed in this chapter would be to dispense with diffusion imaging data altogether and use functional data to estimate not only a posterior density over partial correlation matrices but also over structural graphs [Atay-Kayis and Massam, 2005]. In order to solve this inference problem, sophisticated methods have been developed [Wang and Li, 2012]. However, at present, these methods do not scale well with problem size, prohibiting a straightforward application to functional connectivity analysis.

Notwithstanding the strengths and weaknesses of different approaches to functional connectivity analysis, we maintain that, in order to draw valid conclusions about functional connectivity, one needs to employ methods that quantify the uncertainty in our estimates. This holds especially when inferences depend on a small amount of data. One approach would be to use bootstrap procedures together with the graphical lasso [Hastie et al., 2009]. Another approach, as demonstrated in this chapter, is to use a Bayesian approach. While both structural connectivity and effective connectivity have been tackled using Bayesian approaches [Jbabdi et al., 2007; Daunizeau et al., 2011], research into how whole brain functional connectivity can be estimated using Bayesian approaches has remained scarce. Some notable exceptions are the approach by Venkataraman et al. [2010] who used a forward model in which fMRI and DWI data were combined and the approach by [Marrelec et al., 2006], who used a Bayesian approach to estimate a group partial correlation matrix. As advocated in this chapter, we propose that a

generative model consisting of a G-Wishart prior and a multivariate Gaussian likelihood term serves as an elegant new approach for Bayesian functional connectivity analysis.

5

UNCONSTRAINED FUNCTIONAL CONNECTIVITY

A key objective in many areas of science is to uncover the interactions amongst a large number of variables based on a limited amount of data. Examples include gene regulatory networks, where one wants to identify the interactions amongst DNA segments, market basket analysis where the relations are studied between customers based on their purchase behavior, or neuroscience where the connections between segregated neuronal populations are linked to cognitive ability and impairment. One way to estimate these relations is to employ Gaussian graphical models, where the non-zero entries in the off-diagonal of a precision matrix correspond to the edges in a conditional independence graph [Dempster, 1972]. However, fully Bayesian estimation of the posterior of a Gaussian graphical model has proven to be notoriously hard.

To allow Bayesian inference of the Gaussian graphical model, a conjugate prior [Diaconis and Ylvisaker, 1979] on a precision matrix restricted by the conditional independence graph G was constructed for decomposable graphs [Dawid and Lauritzen, 1993], and later generalized to arbitrary graphs [Roverato, 2002]. Subsequent work coined this distribution the G-Wishart distribution [Atay-Kayis and Massam, 2005]. A number of Monte Carlo algorithms for model estimation using the G-Wishart distribution have been developed [Piccioni, 2000; Mitsakakis et al., 2011; Dobra and Lenkoski, 2011; Wang and Li, 2012], but each of these algorithms required substantial computational resources due to difficulty with sampling from the G-Wishart distribution. To address this bottleneck, a recent study proposed an efficient way to directly sample from the G-Wishart distribution [Lenkoski, 2013] by scaling samples from a regular Wishart distribution to fit the required dependency structure [Hastie et al., 2009]. Even with the direct sampler, approximating the Gaussian graphical model remained difficult because of the doubly intractable partition function of the G-Wishart distribution. However, by combining features of the exchange algorithm [Murray et al., 2006] with reversible jump sampling [Green, 1995], calculating the partition function may be circumvented [Lenkoski, 2013]. The algorithm that

This chapter is based on: M. Hinne, A. Lenkoski, T. Heskes and M.A.J. van Gerven, 2014. "Efficient sampling of Gaussian graphical models using conditional Bayes factors." *Stat* 3, pp. 326-336.

implements this idea, named the double reversible jump algorithm, provides substantial computational gains compared to earlier approaches [Lenkoski, 2013].

Although the double reversible jump algorithm enables model selection in a more efficient manner than previously possible, computational costs remain a limiting factor in practical applications with a large number of variables. In this chapter, we propose two novel, faster, algorithms for Bayesian estimation of the Gaussian graphical model. In the first algorithm, we combine the direct sampler [Lenkoski, 2013] with an efficient representation of the conditional Bayes factor [Cheng and Lenkoski, 2012], which results in an elegant Metropolis-Hastings algorithm to which we will refer as the double conditional Bayes factor sampler. In the second algorithm, we cast the double conditional Bayes factors algorithm in a birth-death MCMC setting [Mohammadi and Wit, 2015]. Here, rather than accepting or rejecting a new state with an edge added or removed, we associate with these changes birth and death events, respectively. These events occur with such rates that their equilibrium coincides with the posterior of interest [Stephens, 2000]. Both algorithms provide substantial speed improvement over the status quo, as we show in simulations.

We also provide an application of our algorithms by estimating structural and functional connectivity between subcortical structures using resting-state fMRI. It is a major goal in cognitive neuroscience to understand how spatially segregated neural populations are coupled, using indirect measures of neural activity such as functional magnetic resonance imaging [Smith et al., 2013; Salinas and Sejnowski, 2001]. In this context, the anatomical pathways between neural populations are referred to as structural connectivity whereas correlated activity patterns between these populations are referred to as functional connectivity [Friston, 2011]. Both forms of connectivity may be estimated simultaneously using Gaussian graphical models. Here, the precision matrix captures the functional interactions between variables and the associated conditional independence graph represents the direct connections between variables. Bayesian estimation of Gaussian graphical models is particularly relevant since the posterior over precision matrices provides complete information about the strength of functional interactions and the posterior over conditional independence graphs allows one to associate a probability with a putative direct connection between variables of interest.

5.1 GAUSSIAN GRAPHICAL MODELS

Preliminaries

Let observed data $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_n)^T$ consist of n independent draws from a p-dimensional multivariate Gaussian distribution $\mathcal{N}(\mathbf{0}, \mathbf{K}^{-1})$, with zero mean and precision (inverse covariance) matrix $\mathbf{K}$. Here, $\mathbf{K} \in \mathbb{P}_p$, with $\mathbb{P}_p$ the space of positive definite $p \times p$ matrices. The likelihood of $\mathbf{K}$ is given by

$$p(\mathbf{X} \mid \mathbf{K}) = \prod_{i=1}^{n} \mathcal{N}(\mathbf{x}_i \mid \mathbf{0}, \mathbf{K}^{-1}) \propto |\mathbf{K}|^{n/2} \exp\left[-\frac{1}{2}\left\langle \mathbf{K}, \hat{\boldsymbol{\Sigma}} \right\rangle\right] , \tag{5.1}$$

where $\hat{\mathbf{\Sigma}} = \mathbf{X}^\mathsf{T}\mathbf{X}$ is the empirical scatter and $\langle \cdot, \cdot \rangle$ the trace inner product operator. The precision matrix has the important property that zero elements correspond to conditional independencies. In other words, (5.1) specifies a Gaussian Markov random field with respect to a graph $G = (V, E)$, with $V = \{1, \ldots, p\}$ and $E \subset V \times V$, in which the absence of a connection indicates independence, i.e. $(i, j) \notin E \rightarrow k_{ij} = 0$. For convenience, throughout this chapter we slightly abuse notation and use $(i, j) \in G$ to indicate that the edge (i, j) is present in E.

The dependency graph may be used to specify a prior distribution on the precision matrix, which is known as the G-Wishart distribution [Roverato, 2002]:

$$P(\mathbf{K} \mid G, b, \mathbf{D}) = \mathcal{W}_G(b, \mathbf{D}) = \frac{|\mathbf{K}|^{(b-2)/2}}{Z_G(b, \mathbf{D})} \exp\left[-\frac{1}{2}\langle \mathbf{K}, \mathbf{D} \rangle\right] \mathbf{1}_{\mathbf{K} \in \mathbb{P}_G} \ , \tag{5.2}$$

in which $\mathbb{P}_G$ is the space of positive definite $p \times p$ matrices that have zero elements wherever $(i, j) \notin G$, b is the prior degrees of freedom, $\mathbf{D}$ is the prior scaling matrix and $\mathbf{1}_x$ evaluates to 1 if and only if x holds and to 0 otherwise. The G-Wishart distribution is conjugate to the multivariate Gaussian likelihood in (5.1), so that

$$\begin{aligned} P(\mathbf{K} \mid G, b, \mathbf{D}, \mathbf{X}) &= \mathcal{W}_G(b + n, \mathbf{D} + \hat{\mathbf{\Sigma}}) \\ &= \frac{|\mathbf{K}|^{(n+b-2)/2}}{Z_G(b + n, \mathbf{D} + \hat{\mathbf{\Sigma}})} \exp\left[-\frac{1}{2}\left\langle \mathbf{K}, \mathbf{D} + \hat{\mathbf{\Sigma}} \right\rangle\right] \ . \end{aligned} \tag{5.3}$$

Note that the Wishart distribution is a special case of the G-Wishart distribution, with which it coincides if G is a fully connected graph. Importantly, the partition function $Z_G(b, \mathbf{D})$ depends on G, which makes the G-Wishart a doubly intractable distribution. We return to the implications of this fact later on.

Central to this work is that we wish to perform model selection in Gaussian graphical models, which revolves around the joint posterior

$$P(G, \mathbf{K} \mid \mathbf{X}) \propto P(\mathbf{X} \mid \mathbf{K}) P(\mathbf{K} \mid G) P(G) \ . \tag{5.4}$$

In the remainder, we outline several algorithms to approximate this distribution.

Direct samples from the G-Wishart distribution

Since the prior $P(\mathbf{K} \mid G)$ is $\mathcal{W}_G(b, \mathbf{D})$, we need a way to draw samples from the G-Wishart distribution. Up until recently, this was achieved using a block Gibbs sampler that updates $\mathbf{K}$ according to either the edges of G [Wang and Li, 2012] or its clique decomposition [Piccioni, 2000]. Although this enables model inference of $P(G, \mathbf{K} \mid \mathbf{X})$, as desired, both approaches require substantial computational effort, making them prohibitive for use in contexts with a large number of variables. An alternative method was proposed that is more efficient [Lenkoski, 2013], which is an adaptation of an algorithm for estimating the mode $\hat{\mathbf{K}}$ of the G-Wishart distribution [Hastie et al., 2009; Moghaddam et al., 2009]. The algorithm is as follows:

1. Sample $\mathbf{K}^* \sim \mathcal{W}(b, \mathbf{D})$.
2. Let $\mathbf{C} = (\mathbf{K}^*)^{-1}$ and $\boldsymbol{W} = \mathbf{C}$.
3. Repeat for $j = 1, 2, \ldots, p$ until convergence:
   a) Let $N_j \subset V$ be the set of variables that are connected to j in G. Form $\boldsymbol{W}_{N_j}$ and $\mathbf{C}_{N_j,j}$ and solve $\hat{\beta}_j^* = \boldsymbol{W}_{N_j}^{-1}\mathbf{C}_{N_j,j}$.
   b) Form $\hat{\beta}_j \in \mathbb{R}^{p-1}$ by copying the elements of $\hat{\beta}_j^*$ to the appropriate locations and imputing zeros in those locations not connected to j in G.
   c) Replace $\boldsymbol{W}_{j,-j}$ and $\boldsymbol{W}_{-j,j}$ with $\boldsymbol{W}_{-j,-j}\hat{\beta}_j$.
4. Return $\mathbf{K} = \boldsymbol{W}^{-1}$.

Conceptually, the algorithm draws a sample from a Wishart distribution, which is then iteratively scaled according to the dependence structure in G. In practice, we observe that convergence (see step 3) is typically reached within a handful of iterations, even for moderate to large p.

5.2 SAMPLING ALGORITHMS

The direct sampler paves the way for novel inference algorithms. Here, we introduce two novel algorithms for approximation of the joint posterior in (5.4).

Double reversible jump sampler

As a baseline for comparison, we use the double reversible jump (DRJ) sampler [Lenkoski, 2013]. This algorithm was shown to be more efficient as previously used approaches and may be considered state of the art. It builds upon the reversible jump sampler discussed in [Dobra et al., 2011]. The key idea offered by this approach is that it introduces an auxiliary variable $\mathbf{K}^0 \sim \mathcal{W}_G(b, \mathbf{D})$, as in the exchange algorithm [Murray et al., 2006], that is efficiently sampled using the direct sampler discussed above. Because of the way this auxiliary variable is constructed, the doubly intractable partition functions of the G-Wishart distribution are canceled out in the calculation of the acceptance ratios of newly proposed graphs.

Direct double conditional Bayes factor sampler

The double reversible jump algorithm provides a substantial improvement over previous algorithms, as it avoids the need to approximate the ratio of partition functions or invoke the Gibbs sampling algorithm for drawing samples from the G-Wishart distribution. Nonetheless, the algorithm can be simplified. In [Cheng and Lenkoski, 2012] it is shown that if G and $\tilde{G}$ differ only

in the edge $e = (p-1, p)$ and $G \subset \tilde{G}$, the odds ratio of these two models may be expressed as

$$\frac{P(\mathbf{X} \mid \tilde{G}, \mathbf{K}, \mathbf{D})}{P(\mathbf{X} \mid G, \mathbf{K}, \mathbf{D})} = N(\mathbf{K}, \mathbf{D} + \hat{\mathbf{\Sigma}}) \frac{Z_G(b, \mathbf{D})}{Z_{\tilde{G}}(b, \mathbf{D})} \tag{5.5}$$

with

$$N(\mathbf{K}, \mathbf{U}) \equiv \phi_{p-1,p-1} \left(\frac{2\pi}{u_{pp}}\right)^{1/2} \times \exp\left[\frac{1}{2} u_{pp} \left(\frac{\phi_{p-1,p-1} u_{p-1,p}}{u_{pp}} - \frac{\sum_{l=1}^{p-2} \phi_{lp-1} \phi_{lp}}{\phi_{p-1,p-1}}\right)^2\right], \tag{5.6}$$

where $\mathbf{K} = \Phi^T \Phi$, with Φ the Cholesky decomposition of $\mathbf{K}$. The term in (5.5) can be considered the conditional Bayes factor of the comparison between G and $\tilde{G}$. Similar to the double reversible jump approach, [Cheng and Lenkoski, 2012] propose to augment the sampling process with an auxiliary variable $\tilde{\mathbf{K}}^0 \sim \mathcal{W}_G(b, \mathbf{D})$. This results in a convenient acceptance ratio for the addition of an edge to G

$$\alpha = \frac{N(\mathbf{K}, \mathbf{D} + \hat{\mathbf{\Sigma}})}{N(\tilde{\mathbf{K}}^0, \mathbf{D})} \frac{P(\tilde{G})}{P(G)}, \tag{5.7}$$

where the ratio is inverted if the edge is removed from G instead. Note that the variables G, $\mathbf{K}$, $\mathbf{U}$ and $\mathbf{D}$ must be permuted for each edge flip to place the particular edge under consideration in the position $(p-1, p)$.

The algorithm described in [Cheng and Lenkoski, 2012] employs the block Gibbs sampler to sample from the G-Wishart distribution. Instead, here we propose to make use of the direct sampler explained in Section 5.1 to arrive at the following procedure for estimation of the Gaussian graphical model:

1. Let $G = G^{[s]}$ be the current graph and let $\mathbf{K} = \mathbf{K}^{[s]} \sim \mathcal{W}_G(b + n, \mathbf{D} + \hat{\mathbf{\Sigma}})$.
2. For each edge $(i, j) \in G$, do:
   a) Create a permutation of the variables so that $(i, j) \to (p-1, p)$. Permute G, $\mathbf{K}$, $\mathbf{D}$ and $\hat{\mathbf{\Sigma}}$ accordingly.
   b) Let $\tilde{G} = G \cup (p-1, p)$ if $(p-1, p) \notin G$ or $\tilde{G} = G \setminus (p-1, p)$ if $(p-1, p) \in G$.
   c) Draw $\tilde{\mathbf{K}}^0 \sim \mathcal{W}_{\tilde{G}}(b, \mathbf{D})$.
   d) Accept the move from G to $\tilde{G}$ with probability α as in (5.7).
   e) Restore the original ordering of G, $\mathbf{K}$, $\mathbf{D}$ and $\hat{\mathbf{\Sigma}}$ and draw $\tilde{\mathbf{K}} \sim \mathcal{W}_{\tilde{G}}(b + n, \mathbf{D} + \hat{\mathbf{\Sigma}})$
3. Set $G^{[s+1]} = \tilde{G}$ and $\mathbf{K}^{[s+1]} \sim \mathcal{W}_{\tilde{G}}(b + n, \mathbf{D} + \hat{\mathbf{\Sigma}})$.

The usage of the direct sampler instead of the block Gibbs updates makes this direct double conditional Bayes factors (DCBF) algorithm computationally much more efficient [Liang, 2010].

Double continuous time sampler

A downside of the usage of an auxiliary variable scheme is that it decreases the acceptance probability of proposals, as essentially two moves have to be accepted at once. This hampers mixing of the Markov chain, so that multimodal distributions are approximated poorly. To improve acceptance, [Mohammadi and Wit, 2015] introduce a continuous-time Markov process [Cappé et al., 2003] for Gaussian graphical models. Rather than accepting the addition or removal of an edge, [Mohammadi and Wit, 2015] associates birth and death events with these changes, respectively. Each edge dies independently of all others as a Poisson process with death rate $d_e(G, \mathbf{K})$. Because the edges are independent, the overall death rate at a particular pair of graph G and precision $\mathbf{K}$ is $d(\mathbf{K}) = \sum_e d_e(G, \mathbf{K})$. Birth rates $b(\mathbf{K})$ are defined similarly, but for non-edges instead.

Because the birth and death processes are independent Poisson processes, the expected time between two events is $1/(d(\mathbf{K}) + b(\mathbf{K}))$. This time can be considered the process spends at any particular instance of $(G, \mathbf{K})$. The probability of the death event of edge $e \in G$ is

$$P(\text{death of edge } e) = \frac{d(G, \mathbf{K})}{b(G, \mathbf{K}) + d(G, \mathbf{K})} \ , \tag{5.8}$$

with again an analogous definition for the birth event for a non-edge.

Mohammadi and Wit [Mohammadi and Wit, 2015] show that the birth-death process has the posterior $P(G, \mathbf{K} \mid \mathbf{X})$ as stationary distribution, if for all edges and non-edges e

$$d_e(\tilde{G}, \tilde{\mathbf{K}}) P(\tilde{G}, \tilde{\mathbf{K}} \mid \mathbf{X}) = b_e(G, \mathbf{K}) P(G, \mathbf{K} \mid \mathbf{X}) \ , \tag{5.9}$$

for $\tilde{G} = G \cup e$. The birth and death rates may be chosen accordingly as

$$b_e(G, \mathbf{K}) = \frac{P(\tilde{G}, \tilde{\mathbf{K}} \mid \mathbf{X})}{P(G, \mathbf{K} \mid \mathbf{X})} \quad \text{for } e \notin G \ , \tag{5.10}$$

and

$$d_e(G, \mathbf{K}) = \frac{P(G, \mathbf{K} \mid \mathbf{X})}{P(\tilde{G}, \tilde{\mathbf{K}} \mid \mathbf{X})} \quad \text{for } e \in G \ , \tag{5.11}$$

with again $\tilde{G} = G \cup e$.

The key observation is now that these birth-death rates can be computed using the double conditional Bayes factors as in (5.7). Here again we make use of the exchange framework by introducing the auxiliary variable $\mathbf{K}^0$, such that explicit evaluation of the partition functions

is circumvented. This leads to a novel approach that we will refer to as the double continuous time (DCT) sampler, given by:

1. Let $G = G^{[s]}$ be the current graph and let $\mathbf{K} = \mathbf{K}^{[s]} \sim \mathcal{W}_G(b + n, \mathbf{D} + \hat{\mathbf{\Sigma}})$.
2. For each non-edge $e \notin G$:
   a) Create a random permutation of the variables so that $(i, j) \to (p - 1, p)$. Permute G, $\mathbf{K}$, $\mathbf{D}$ and $\hat{\mathbf{\Sigma}}$ accordingly.
   b) Let $\tilde{G} = G \cup e$. Draw $\mathbf{K}^0 \sim \mathcal{W}_{\tilde{G}}(b, \mathbf{D})$
   c) Compute the birth rate $b_e(G, \mathbf{K})$ using (5.10).
3. Compute the total birth rate of the current state $b(G, \mathbf{K})$.
4. For each edge $e \in G$:
   a) Create a random permutation of the variables so that $(i, j) \to (p - 1, p)$. Permute G, $\mathbf{K}$, $\mathbf{D}$ and $\hat{\mathbf{\Sigma}}$ accordingly.
   b) Let $\tilde{G} = G \setminus e$. Draw $\mathbf{K}^0 \sim \mathcal{W}_{\tilde{G}}(b, \mathbf{D})$
   c) Compute the death rate $d_e(G, \mathbf{K})$ using (5.10).
5. Compute the total death rate of the current state $d(G, \mathbf{K})$ and the waiting time between events $w(G, \mathbf{K}) = 1/(d(\mathbf{K}) + b(\mathbf{K}))$.
6. Create a birth or death event according to the probabilities of death events (5.8) and birth events, and set $G^{[s+1]} = \tilde{G}$ and $\mathbf{K}^{[s+1]} \sim \mathcal{W}_{\tilde{G}}(b + n, \mathbf{D} + \hat{\mathbf{\Sigma}})$.

5.3 EXPERIMENTS

In this section we first analyze the validity of the two proposed methods using an example with a known precision matrix. Subsequently we apply the algorithms in an explorative study to identify structural and functional connectivity between subcortical brain structures.

Simulation

We compared the performance of the double reversible jump algorithm and the two novel algorithms using a simulation proposed in [Wang and Li, 2012]. In this example, we have $p = 6$ and $n = 18$. Furthermore, the precision matrix $\mathbf{K}$ is given by $k_{ii} = 1$ for $i = 1, \ldots, p$, $k_{i,i+1} = k_{i+1,i} = 0.5$ for $i = 1, \ldots, p - 1$ and finally $k_{1p} = k_{p1} = 0.4$. The associated conditional independence graph G follows as $(i, j) \in G \leftrightarrow k_{ij} \neq 0$. The scatter matrix is then constructed as $\hat{\mathbf{\Sigma}} = \mathbf{X}^T\mathbf{X} = n\mathbf{K}^{-1}$, which corresponds to n independent observations

of $\mathcal{N}(0, K^{-1})$. Through exhaustive enumeration of all 32 768 possible graphs of size p, [Wang and Li, 2012] shows that the posterior edge probabilities are

$$P((i,j) \in G \mid X) = \begin{pmatrix} 1 & 0.969 & 0.106 & 0.085 & 0.113 & 0.850 \\ 0.969 & 1 & 0.980 & 0.098 & 0.081 & 0.115 \\ 0.106 & 0.980 & 1 & 0.982 & 0.098 & 0.086 \\ 0.085 & 0.098 & 0.982 & 1 & 0.980 & 0.106 \\ 0.113 & 0.081 & 0.98 & 0.980 & 1 & 0.970 \\ 0.850 & 0.115 & 0.086 & 0.106 & 0.970 & 1 \end{pmatrix} \tag{5.12}$$

and the expectation of K is

$$\mathbb{E}(K \mid X) = \begin{pmatrix} 1.139 & 0.569 & -0.011 & 0.006 & -0.013 & 0.403 \\ 0.569 & 1.175 & 0.574 & -0.008 & 0.005 & -0.014 \\ -0.011 & 0.574 & 1.176 & 0.574 & -0.008 & 0.006 \\ 0.006 & -0.008 & 0.574 & 1.175 & 0.573 & -0.011 \\ -0.013 & 0.005 & -0.008 & 0.573 & 1.175 & 0.569 \\ 0.403 & -0.014 & 0.006 & -0.011 & 0.569 & 1.138 \end{pmatrix} . \tag{5.13}$$

We approximate this ground truth using the three different algorithms, each implemented in Matlab. Throughout, we use vague priors in the form of $P(G) \propto 1$ for G and $P(K \mid G) = \mathcal{W}_G(3, I_p)$. The algorithms are each executed for 100 000 iterations, of which the first 50 000 are discarded as burn-in. Conditional expectations for edges (i.e. edge probabilities) and precision matrices are then calculated as

$$\mathbb{E}((i,j) \in G \mid X) = \frac{1}{T} \sum_{t=1}^{T} 1_{(i,j) \in G_t} \quad \text{and} \quad \mathbb{E}(K \mid X) = \frac{1}{T} \sum_{t=1}^{T} K_t \tag{5.14}$$

for the double reversible jump and the double conditional Bayes factor algorithms, with T the number of samples. For the double continuous time algorithm, these expectations are calculated as

$$\mathbb{E}((i,j) \in G \mid X) = \frac{1}{W} \sum_{t=1}^{T} w_t 1_{(i,j) \in G_t} \quad \text{and} \quad \mathbb{E}(K \mid X) = \frac{1}{W} \sum_{t=1}^{T} w_t K_t \ , \tag{5.15}$$

with $W = \sum_{t=1}^{T} w_t$. It is easy to see that this idea generalizes the discrete time MCMC approach by assuming $w_t = 1$ for all t.

We quantify the approximation accuracy of the three algorithms in a number of ways. First, the accuracy of the edge probabilities is expressed using the mean squared error with respect to the true probabilities in (5.12). Second, we compute the Kullback-Leibler divergence [Kullback and Leibler, 1951] between the precision matrix obtained in [Wang and Li, 2012] as de-

Table 5.1: Results for the comparison between the three described samplers on a simulated example, averaged over 10 simulations. Standard errors are indicated in parentheses. Shown are the mean squared error (MSE) of edge probabilities relative to (5.12), the Kullback-Leibler divergence (KL) between the expected precision matrix and (5.13), the number of unique models visited, the marginal posterior probability of the true graph $P(G \mid \hat{\boldsymbol{\Sigma}})$ and the relative speed of the algorithms compared to the double reversible jump baseline.

Method	MSE	KL	#models	$P(G \mid \hat{\boldsymbol{\Sigma}})$	Rel. speed
DRJ	5e-04 (4e-05)	1e-04 (2e-05)	1 299 (31)	0.37 (1e-3)	1 (0)
DCBF	5e-04 (2e-05)	1e-04 (1e-05)	1 472 (23)	0.38 (4e-3)	3.6 (1e-01)
DCT	1e-03 (1e-05)	7e-04 (3e-04)	1 187 (35)	0.43 (1e-3)	3.8 (1e-02)

fined in (5.13) and $\hat{\mathbf{K}} \equiv \mathbb{E}(\mathbf{K} \mid \mathbf{X})$ using either of the algorithms. We also count the number of unique models that each algorithm considers to express mixing behavior. Next, we compute the marginal posterior probability of the true graph. Finally, we compute the relative computational speeds of the algorithms. The results of the comparison are shown in Table 5.1. The algorithms have similar performance in approximating the desired posterior distribution and each obtains the true graph as the mode of the approximated distribution. Contrary to [Mohammadi and Wit, 2015], we do not find the continuous time algorithm to have the best mixing. In fact, of the three considered models, the continuous time MCMC approach finds the smallest number of unique models. Note that the continuous time approach may converge faster [Rao and Teh, 2012], but this is not apparent in this simulation. Finally, the efficiency of our way of computing the conditional Bayes factor (see (5.5)) is demonstrated by a substantial speed increase, as the DCBF algorithm is 3.57 times faster than the DRJ sampler, and the DCT algorithm is 3.80 times faster than the DRJ algorithm, whereas the algorithm in [Mohammadi and Wit, 2015] is 1.79 times slower than the DRJ sampler.

Subcortical brain connectivity

As an explorative example, we estimate structural and functional connectivity in a fully Bayesian setting. In the previous chapter, functional connectivity has been estimated under the assumption that the underlying structural connectivity was known. Here, we address the more challenging problem of simultaneously estimating the posterior distribution of both structural and functional connectivity.

Empirical data

The data consist of resting-state functional MRI data collected for one subject. We refer the reader to [van Oort et al., 2014] for details of the acquisition protocol. Preprocessing was performed using FSL 5.0 [Jenkinson et al., 2012] and consisted of the following steps. T1 images were linearly registered to MNI-152 space. Multi-echo volumes at each TR were combined [Poser

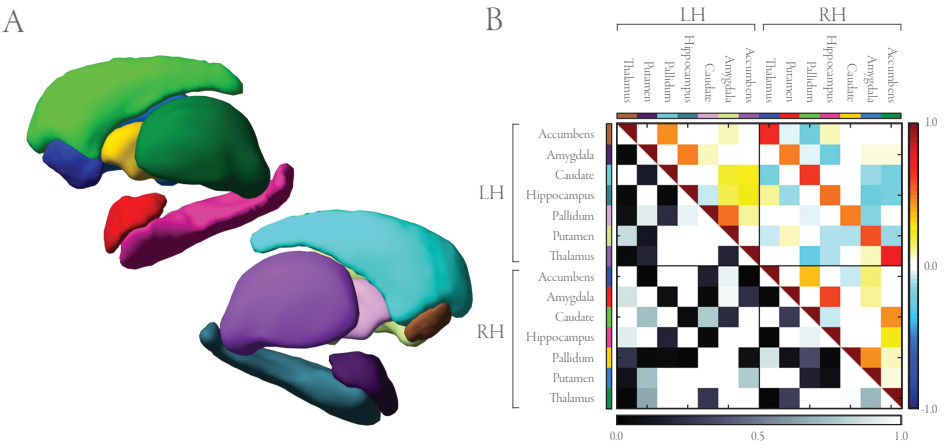


FIGURE 5.1: Subcortical connectivity. A. Subcortical structures, consisting of bilateral accumbens, amygdala, caudate, hippocampus, pallidum, putamen and thalamus. B. Posterior probabilities of structural connectivity (lower triangle) and expected partial correlations between these structures (upper triangle). LH and RH indicate left hemisphere, right hemisphere, respectively.

et al., 2006]. Motion correction was performed using MCFLIRT and estimated motion parameters were regressed out together with their temporal derivatives and mean time courses for both WM and CSF. Finally, data were high-pass filtered at 0.001 Hz. Subcortical structures were segmented using FSL FIRST [Patenaude et al., 2011], resulting in data for a total of 14 regions, consisting of bilateral accumbens, amygdala, caudate, hippocampus, pallidum, putamen and thalamus (see Figure 5.1A). For each of these regions the signal was averaged over all voxels in that region and subsequently standardized to have zero mean and unit variance.

Bayesian structural and functional connectivity estimation

The human brain can be viewed as a complex dynamical system where ongoing changes in neuronal dynamics produce adaptive behavior [Bullmore and Sporns, 2009]. These dynamics can be expressed in terms of interactions between brain regions, which is commonly referred to as functional connectivity. At the same time, direct functional interactions presuppose anatomical links between brain regions, known as structural connectivity. For this reason, structural and functional connectivity must be intimately related [Akil et al., 2011].

Functional connectivity is most easily expressed using a covariance matrix that, in the case of standardized data, provides the correlation structure between different brain regions. However, this approach suffers from the drawback that it cannot distinguish between direct and indirect connections. Alternatively, one may use partial correlations that capture only direct effects, in the absence of confounding factors. The matrix of partial correlations $\mathbf{R}$ may be obtained from a precision matrix using $r_{ij} = 1$ if $i = j$ and $r_{ij} = -k_{ij}/\sqrt{k_{ii}k_{jj}}$ otherwise. Because functional coupling must be accompanied by an anatomical connection, partial corre-

lations between brain regions not only reveal the strength of these couplings, but also indicate which regions are physically connected. In other words, the joint posterior in (5.4) becomes a distribution over functional connectivity $\mathbf{K}$ (or, equivalently, $\mathbf{R}$) and structural connectivity G.

We proceed by approximating the joint posterior using both the DCBF algorithm as well as the DCT sampler. Both algorithms were executed for 100 000 iterations, of which the first 50 000 were discarded as burn in. Once again, we set $P(G) \propto 1$ and $P(\mathbf{K} \mid G) = \mathcal{W}_G(3, \mathbf{I}_p)$. The algorithms yield almost identical results, as shown by an MSE of edge probabilities of 0.0006 and a symmetrized Kullback-Leibler divergence of 0.0002.

Figure 5.1B shows the posterior edge probabilities and partial correlations produced by the DCT algorithm. The structural connectivity estimate shows that the majority of edges is associated with either very high or very low edge probabilities. The functional connectivity estimate shows that functional homologues in left and right hemispheres are associated with high partial correlations (expected partial correlations $\langle r \rangle$ in the range $[0.48, 0.73]$), indicating that these functional homologues have similar functional roles. Within a cortical hemisphere, the most salient functional interactions (highest expected partial correlations with $\langle r \rangle$ in the range $[0.23, 0.61]$) are given bilaterally by the connections between amygdala and hippocampus, pallidum and putamen, accumbens and caudate, caudate and thalamus, and finally hippocampus and thalamus. These functional interactions can be explained by direct pathways as well as unobserved common inputs that induce a high partial correlation. Interestingly, edges with high posterior probability (edge probability higher than 0.999) can be associated with weak absolute partial correlations (with $\langle r \rangle$ as low as 0.1). This indicates that there exist weakly coupled regions (from the linear correlation point of view) that cannot be explained away by other functional interactions.

5.4 Discussion

We have proposed two novel algorithms for Bayesian model selection in a Gaussian graphical model. The first algorithm combines a direct manner to sample G-Wishart variates [Lenkoski, 2013] with an efficient way of computing conditional Bayes factors when comparing two different models [Cheng and Lenkoski, 2012], resulting in an improved Metropolis-Hastings approach. The second approach integrates the direct sampler within a birth-death continuous time Markov process [Mohammadi and Wit, 2015]. Both algorithms provide accurate estimates of the posterior graphs and precision matrices and are substantially faster (up to a factor of 3.80) than previously available alternatives. We demonstrate the use of the algorithms by estimating, for the first time, both structural and functional connectivity simultaneously using fMRI data.

In future work we aim to improve mixing of the samplers by introducing moves between graphs that differ by more than a single edge. Similarly, one may conceive events other than births and deaths of edges. In either case, the corresponding conditional Bayes factors must be derived, and these should be more efficient to compute than a series of consecutive edge additions and removals. We expect that this will further contribute to efficient estimation of Gaussian graphical models.

6

SIMULTANEOUS ESTIMATION OF STRUCTURAL AND FUNCTIONAL CONNECTIVITY

In the early days of neuroscience much attention was devoted to identifying the functional specialization of different brain areas [Friston, 2011]. More recently, this focus has shifted towards revealing how these areas are organized into networks and how these networks, rather than their individual constituents, relate to cognition [Bressler and Menon, 2010; Baronchelli et al., 2013; Sporns, 2014] and neurological or psychological pathology [Catani, 2007; Craddock et al., 2013; Fornito and Bullmore, 2012]. The increasing interest in neuronal connectivity sprouted its own subdiscipline known as *connectomics* [Sporns et al., 2005; Hagmann, 2005; Bullmore and Sporns, 2009]. Within connectomics, one distinguishes between structural connectivity and functional connectivity. Structural connectivity is concerned with the anatomical white-matter fiber bundles that connect remote regions of the brain. It may be estimated in vivo by diffusion weighted MRI (dMRI), which measures the fractional anisotropy of the diffusion of water molecules [Le Bihan et al., 2001]. Functional connectivity in turn expresses the (degree of) dependency between the neuronal activity of separate brain regions [Friston, 1994; Craddock et al., 2013] and is typically measured non-invasively via either functional MRI, electro- or magnetoencephalography (fMRI, EEG and MEG, respectively) [Fornito and Bullmore, 2014].

Several measures to quantify (the degree of) functional coupling exist [Smith et al., 2011; Fiecas et al., 2013], of which the most prevalent is covariance. When the activity signal is normalized to have zero mean and unit variance, covariance coincides with Pearson correlation. As the correlation matrix is easy to compute, it has become the de facto standard in operationalizing functional connectivity. It does however have an important drawback: it is unable to differentiate between direct and indirect effects. For example, if regions A and B are correlated, and similarly B and C show correlation, then correlation between A and C is induced [Varoquaux

This chapter is based on: M. Hinne, R.J.Janssen, T. Heskes and M.A.J. van Gerven, 2014. "Bayesian estimation of conditional independence graphs improves functional connectivity estimates." *PLoS Computational Biology* 11(11), pp. e1004534.

and Craddock, 2013; Smith, 2012]. This poses a problem for functional connectomics, as it introduces type 1 errors. The problem may be remedied to some extent by using *partial* correlations instead. Its interpretation is similar to Pearson correlation, but it captures only direct effects as the influence from other regions is partialled out. In practical terms, the matrix of partial correlations may be obtained by taking the inverse of the covariance matrix, known as the precision matrix, and rescaling this. Assuming the data is normally distributed, both the precision matrix and the partial correlation matrix capture the conditional independence structure of the considered variables, i.e. when two regions are conditionally independent given all other regions, their precision and partial correlation are zero.

Ideally, the partial correlation matrix would reflect the functional connectivity that generated the observed data. If this matrix is sparse, the corresponding conditional independence graph provides an intuitive representation of the interaction between different regions. In practice however, the obtained partial correlation matrices are not sparse, which makes the estimated connectivity more difficult to interpret. In addition, if the number of samples is small and the number of regions large, there is no unique inverse of the covariance matrix and consequently no unique matrix of partial correlations. Even when these conditions are met, the maximum likelihood solution is often ill-behaved, in which case the solution must be regularized [Pourahmadi, 2011]. A popular approximation of the precision matrix is acquired via the graphical LASSO (Least Absolute Shrinkage and Selection Operator), which regularizes the elements of the precision matrix using the ℓ_1-norm [Smith et al., 2011; Varoquaux and Craddock, 2013; Varoquaux et al., 2010]. This approach shrinks the partial correlations towards zero so as to create sparse solutions, which are easier to interpret. Although the graphical LASSO was found to be one of the must accurate methods in identifying connectivity in a comparative study [Smith et al., 2011], it introduces a bias that underestimates functional connectivity, thus creating type 2 errors (see Chapter 4). In addition, both the original maximum likelihood solution as well as the LASSO estimate provide point estimates that do not quantify the reliability of their outcome. In earlier work, we have proposed a Bayesian alternative to the graphical LASSO that uses the G-Wishart distribution to restrict the partial correlation estimates to a previously defined conditional independence graph. We showed that structural connectivity provides an elegant candidate for this graph, and that this approach was able to outperform the graphical LASSO on simulated data. Importantly however, we assumed that the conditional independence graph was available a priori. In the current contribution we take this line of reasoning a critical step forwards and learn both functional connectivity as well as its conditional independence structure simultaneously. Apart from estimating the degree to which two regions have correlated activity, we can now also express the probability of these regions being conditionally independent. As we will show, this results in a more effective approach to regularization than the graphical LASSO, while retaining the additional benefits of the Bayesian framework.

At the foundation of this contribution lies a probabilistic generative model that describes how a particular independence structure generates partial correlations that in turn generate observable data. Using a neurologically plausible simulation with several different conditions, as described by Smith et al. [Smith et al., 2011], we show that in many cases our Gaussian graphical model approach is favorable to both the maximum likelihood alternative and graphical LASSO

regularized solutions. Subsequently, we apply the model to estimate functional connectivity between bilateral accumbens, amygdala, caudate, hippocampus, pallidum, putamen and thalamus using their blood-oxygenation level dependent (BOLD) signal time courses, measured using resting-state fMRI. Finally, we demonstrate how the advantages of a Bayesian approach can be put to practice by showing two extensions to our connectivity model. First, we show how the problem of data fusion for connectivity studies [Zhu et al., 2013; Rykhlevskaia et al., 2008] may be tackled by simply providing multiple likelihood terms; one for each imaging modality. This is demonstrated empirically by combining the fMRI time series with dMRI probabilistic tractography results. Second, we describe how further background knowledge on putative connections may be used to both constrain and inform functional connectivity.

6.1 METHODS

Functional connectivity as a Gaussian graphical model

From a methodological perspective, finding functional connectivity is often seen as a covariance selection problem. This boils down to finding a sparse partial correlation matrix associated with the time series (activity) of a set of variables (brain regions), a problem known as covariance selection. Here, the problem is approached using a Gaussian graphical model (GGM), where we assume that the data $\mathbf{X} = (x_1, \ldots, x_n)^T$ consist of n independent draws from a p-dimensional multivariate Gaussian distribution $\mathcal{N}(0, \mathbf{K}^{-1})$, with zero mean and precision (inverse covariance) matrix $\mathbf{K}$. Here, $\mathbf{K} \in \mathbb{P}_p$, with $\mathbb{P}_p$ the space of positive definite $p \times p$ matrices. The likelihood of $\mathbf{K}$ is given by

$$P(\mathbf{X} \mid \mathbf{K}) = \prod_{i=1}^{n} \mathcal{N}(x_i \mid 0, \mathbf{K}^{-1}) \propto |\mathbf{K}|^{n/2} \exp\left[-\frac{1}{2}\langle \mathbf{K}, \mathbf{\Sigma}\rangle\right] \quad , \tag{6.1}$$

where $\mathbf{\Sigma} = \mathbf{X}^T\mathbf{X}$ and $\langle \cdot, \cdot \rangle$ the trace inner product operator. The assumption of Gaussianity is justified empirically, as BOLD data has been shown to follow a Gaussian distribution [Hlinka et al., 2011].

The precision matrix has the important property that zero elements correspond to conditional independencies, provided the data is normally distributed. In other words, (6.1) specifies a Gaussian Markov random field with respect to a graph $G = (V, E)$, with $V = \{1, \ldots, p\}$ and $E \subset V \times V$, in which the absence of a connection indicates conditional independence, i.e. $(i, j) \notin E \rightarrow k_{ij} = 0$ [Lauritzen, 1996; Whittaker, 2009].

In order to estimate the precision matrix $\mathbf{K}$ of a zero-mean multivariate Gaussian density from data $\mathbf{X}$ one may maximize the log-likelihood which gives the maximum likelihood estimate (MLE):

$$\hat{\mathbf{K}} = \underset{\mathbf{K} \in \mathbb{P}_p}{\operatorname{argmax}} (\log |\mathbf{K}| - \langle \mathbf{\Sigma}\mathbf{K} \rangle) \tag{6.2}$$

where the maximization is constrained to precision matrices in the family of $p \times p$ positive definite matrices $\mathbb{P}_p$. If $\mathbf{\Sigma}$ is positive-definite, there exists a unique solution to (6.2) in the form of $\mathbf{\Sigma}^{-1}$. However, if the number of samples is small compared to the number of variables, the solution does not exist, and even if $n > p$, the maximum likelihood estimate is often ill-behaved and requires regularization [Pourahmadi, 2011]. A frequently used method of regularization is called the graphical LASSO [Friedman et al., 2008], which penalizes the magnitude of the elements of $\mathbf{K}$. The LASSO approach gives the following MLE:

$$\hat{\mathbf{K}} = \underset{\mathbf{K} \in \mathbb{P}_p}{\operatorname{argmax}} \left[\log |\mathbf{K}| - \langle \mathbf{\Sigma K} \rangle - \lambda \|\mathbf{K}\|_1 \right] \quad , \tag{6.3}$$

in which the shrinkage parameter λ determines the amount of penalization that is applied. Several studies have applied the graphical LASSO in order to estimate functional connectivity [Smith et al., 2011; Varoquaux and Craddock, 2013; Varoquaux et al., 2010]. Alternative regularization schemes are available [Valdés-Sosa et al., 2005], such as ridge regression or elastic net [Ryali et al., 2012], but we will not consider these methods in detail here. Rather, we emphasize that each of these regularization approaches provides only a point estimate, instead of a posterior distribution over $\mathbf{K}$. This makes it impossible to quantify the uncertainty associated with the estimate, which can lead to incorrect conclusions about functional connectivity in light of finite data. Moreover, it has been shown that the graphical lasso is not guaranteed to find the true graph even in the limit of infinite data [Meinshausen, 2008]. In addition, solutions obtained through regularization tend to underestimate functional connectivity, as seen in Chapter 4.

Recently, extensions of the (graphical) LASSO approach have been proposed that allow for statistical inference. For example, [Lockhart et al., 2014] introduce a significance test that can be applied to LASSO estimates while [van de Geer et al., 2014; Javanmard and Montanari, 2014] describe a *desparsified* LASSO that attempts to de-bias the results using a projection onto the residual space. However, these approaches make assumptions on the sparsity of $\mathbf{K}$, which may not be warranted.

Alternatively, a Bayesian approach can be applied to the covariance selection problem, which requires that we specify a prior distribution on $\mathbf{K}$. As we hope to identify conditional independencies between the considered variables, a convenient prior distribution arises in the form of the G-Wishart distribution [Roverato, 2002]:

$$P(\mathbf{K} \mid G, \delta, \mathbf{D}) = \mathcal{W}_G(\delta, \mathbf{D}) = \frac{|\mathbf{K}|^{(\delta-2)/2}}{Z_G(\delta, \mathbf{D})} \exp\left[-\frac{1}{2}\langle \mathbf{K}, \mathbf{D} \rangle\right] \mathbb{1}_{\mathbf{K} \in \mathbb{P}_G} \quad , \tag{6.4}$$

in which $\mathbb{P}_G$ is the space of positive definite $p \times p$ matrices that have zero elements wherever $(i, j) \notin G$, δ is the degrees of freedom parameter, $\mathbf{D}$ is the prior scaling matrix and $\mathbb{1}_x$ evaluates

to 1 if and only if x holds and to 0 otherwise. The G-Wishart distribution is conjugate to the multivariate Gaussian likelihood in (6.1), so that

$$\begin{aligned} P(\mathbf{K} \mid G, \delta, \mathbf{D}, \mathbf{X}) &= \mathcal{W}_G(\delta + n, \mathbf{D} + \mathbf{\Sigma}) \\ &= \frac{|\mathbf{K}|^{(n+\delta-2)/2}}{Z_G(\delta + n, \mathbf{D} + \mathbf{\Sigma})} \exp\left[-\frac{1}{2}\langle \mathbf{K}, \mathbf{D} + \mathbf{\Sigma}\rangle\right] \quad . \end{aligned} \tag{6.5}$$

Note that the Wishart distribution is a special case of the G-Wishart distribution, with which it coincides if G is a fully connected graph.

It should be pointed out that in the limit of $n \to \infty$, any prior will be fully dominated by the data. In theory, even when the true precision matrix $\mathbf{K}$ contains very small elements, the probability of a corresponding edge will go to 1 in the limit of an infinite amount of data. The interesting question is what happens if the magnitude of these elements scales as a function of n, e.g., as $1/n$. Where asymptotic analyses have been successfully applied to better understand the behavior of regularization approaches such as the graphical LASSO [Zhao and Yu, 2006; Meinshausen and Bühlmann, 2010], such analyses of Bayesian procedures are complex and may lead to counterintuitive results [Ritov et al., 2014]. For the G-Wishart prior in particular, similar analyses have, to the best of our knowledge, not yet been pursued.

The preliminaries described above allow us to specify the distribution that is central to this work, i.e. the joint posterior over both the conditional independence graph and the precision matrix (an illustration of the graphical model is provided in Figure 6.1A):

$$P(G, \mathbf{K} \mid \mathbf{X}) \propto P(\mathbf{X} \mid \mathbf{K}) P(\mathbf{K} \mid G) P(G) \quad . \tag{6.6}$$

Note that the necessary hyperparameters are typically omitted for clarity. In practice, functional connectivity is more intuitively understood in terms of partial correlations. The partial correlation matrix $\mathbf{R}$ may be obtained from the precision matrix by applying the transformation

$$r_{ij} = \begin{cases} 1 & \text{if } i = j, \\ -\frac{k_{ij}}{\sqrt{k_{ii} k_{jj}}} & \text{otherwise.} \end{cases} \tag{6.7}$$

By transforming each element of $\mathbf{K}$ in (6.6), the distribution $P(G, \mathbf{R} \mid \mathbf{X})$ is constructed. When discussing our experimental results, we will focus on partial correlations rather than precision values, unless explicitly stated otherwise. Note that the relation between the dependency structure G and the precision matrix $\mathbf{K}$, as discussed above, also holds between G and the partial correlations $\mathbf{R}$. That is, absence of a connection in $(i, j) \in G$ implies $r_{ij} = 0$.

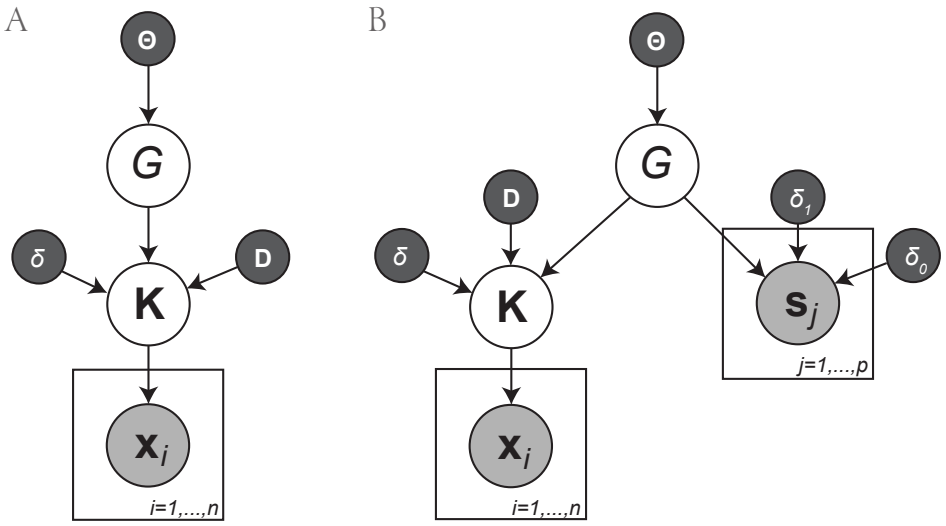


FIGURE 6.1: A The generative model for the conditional dependencies graph and precision matrix with hyperparameters $\boldsymbol{\Theta}$, δ and $\mathbf{D}$. B The generative model for structural connectivity and the precision matrix, based on both BOLD time series $\mathbf{X}$ and probabilistic streamline counts $\mathbf{N}$ as well as hyperparameters $\boldsymbol{\Theta}$ for the dependency structure, δ and $\mathbf{D}$ for the functional connectivity strengths and δ_1 and δ_0 for the structural data.

The Bayesian generative model must be completed by specifying a prior distribution to draw G from. Here, we assume that a priori all edges are marginally independent and each have probability θ. That is, we have

$$P(G \mid \boldsymbol{\Theta}) = \prod_{i<j} \theta_{ij}^{g_{ij}} (1-\theta_{ij})^{1-g_{ij}} \quad , \tag{6.8}$$

with $g_{ij} \in \{0,1\}$, $g_{ij} = 1 \leftrightarrow (i,j) \in G$ and $\Theta = \{\theta_{ij}\}_{i<j}$. Initially we use $\theta_{ij} = 0.5\, \forall_{i,j}$ to indicate that we have no a priori preference for either a dependence or an independence. The impact of different values for θ_{ij} on the posterior estimates is discussed in Appendix C.3, where it is shown that the prior is to a large extent dominated by the likelihood.

Functional connectivity variants

One of the benefits of the Bayesian framework is that extensions to the generative model are straightforward to implement. In this section we use the distribution given in (6.6) to provide two illustrations of such extensions for analyzing connectivity.

Integrating additional modalities

In the ideal case where the complete neural system is considered (i.e. there are no hidden variables that may explain away some conditional dependencies), the conditional independence graph almost entirely coincides with the structural connectome as each functional relation must be facilitated by an anatomical connection [Raj and Chen, 2011]. In other words, G now represents both the conditional independence graph as well as structural connectivity. In this case, functional connectivity may be estimated more accurately by incorporating additional imaging modalities that inform the conditional independence structure. To do so, we must employ an additional likelihood term describing how the data from the extra imaging modality is generated by G. The posterior distribution of connectivity is then given by

$$P(G, \mathbf{K} \mid \mathcal{F}) \propto P(\mathbf{K} \mid G) P(G) \prod_{\mathbf{F} \in \mathcal{F}} P(\mathbf{F} \mid \mathbf{K}, G) \ , \tag{6.9}$$

with $\mathcal{F}$ the collection of data sets to be combined. The result of this mathematically straightforward exercise provides an elegant way to obtain data fusion. While several techniques have been proposed to achieve this (see e.g. [Sui et al., 2011; Rykhlevskaia et al., 2008] for reviews on this topic), these typically rely on ad-hoc strategies instead of a generative model. Although the choice for specific probability distributions may be subject to change, the generative modeling approach serves as a generic way to link structural and functional connectivity and the different modalities that provide data regarding them.

Here, we use the model for structural connectivity based on probabilistic tractography that was proposed in Chapter 2, which we repeat here for convenience. The matrix $\mathbf{S}$ is assumed to contain probabilistic streamline counts [Behrens et al., 2003a] that run from region j to all other $\{1, \ldots, p\} \setminus j$ regions. It is generated from existing anatomical connections, i.e. structural connectivity, through

$$\begin{aligned} P(\mathbf{S} \mid \mathbf{G}, \delta_1, \delta_0) &= \prod_j P(\mathbf{s}_j \mid \mathbf{g}_j, \delta_1, \delta_0) \\ &= \prod_j \mathrm{DirMult}(\delta_1 \mathbf{g}_j + \delta_0(\mathbf{1} - \mathbf{g}_j)) \ , \end{aligned} \tag{6.10}$$

wherein δ_1 and δ_0 are hyperparameters that govern the distributions of streamlines over existing and absent connections, respectively. Integration with the Gaussian graphical model is achieved by incorporating (6.10) into (6.9):

$$P(G, \mathbf{K} \mid \mathbf{X}, \mathbf{S}) \propto P(\mathbf{X} \mid \mathbf{K}) P(\mathbf{K} \mid G) P(\mathbf{S} \mid G) P(G) \ . \tag{6.11}$$

A visual representation of the generative model is shown in Figure 6.1B. Throughout this chapter we refer to our method as the Bayesian Gaussian graphical model (BGGM) approach.

Informative prior

The assumption that the prior probability of connections is the same for all region pairs (see (6.8)) is rather crude, and may be replaced depending on available background information. To illustrate this, we describe an additional approach to connectivity based on the assumption that homotopic regions in different hemispheres are directly connected, but that other interhemispheric connections do not exist. Within either hemisphere, we remain agnostic about connectivity. This intuition is easily formalized by

$$\theta_{ij} = \begin{cases} 0.5 & \text{for } i \text{ and } j \text{ in the same hemisphere,} \\ 1 & \text{for } i \text{ and } j \text{ homotopic regions and} \\ 0 & \text{otherwise.} \end{cases} \tag{6.12}$$

Clearly, this prior is more restrictive than a homogeneous prior, as most of the elements corresponding to cross-hemisphere connections are now excluded. In addition, the restrictive zero probability of some of the interhemispheric connections is an extreme choice. However, we use it here to provide an example of how information regarding the absence of connections (e.g. in the case of a white-matter lesion) affects the estimates of the present connections

Simulation

To analyze the performance of the Gaussian graphical model approach to functional connectivity, we compare our results to those presented in [Smith et al., 2011]. Here, realistic BOLD time series are generated according to the dynamic causal modeling (DCM) fMRI forward model [Friston et al., 2003], that makes use of the nonlinear balloon model [Buxton et al., 1998], based on a known constructed network as its starting point. In total, 28 simulations with different parameters such as number of nodes, number of generated samples, sampling frequency and noise levels were constructed. For each simulation, 50 different time series are generated. For the full details of the approach as well as the different simulation parameters, we refer to the original description in [Smith et al., 2011] as well as the corresponding web page where the simulation may be downloaded[1]. In the simulation study, it was shown that using partial correlation (both maximum likelihood as well as lasso regularized point estimates) resulted in the best (undirected) reconstructions of the ground truth. As these methods performed best, and are closely related to our approach, we use these to compare our results with.

The evaluation procedure is as follows: For each run of each of the 28 different simulations, the time series $\mathbf{X}$ of that run are used to compute $P(G, \mathbf{R} \mid \mathbf{X})$. In addition, for each run the maximum likelihood estimate (MLE) is computed, as well as the graphical lasso regularized point estimate using the same regularization as in [Smith et al., 2011] (i.e. $\lambda \in \{5, 100\}$). The quality of the reconstruction of the ground truth is quantified in three ways. Let $\mathbf{R}^*$ be the ground truth functional connectivity that we are trying to recover. Then $\mathbf{\Gamma} = |\mathbf{R}^* - \mathbf{R}|$ gives

[1] http://www.fmrib.ox.ac.uk/analysis/netsim/

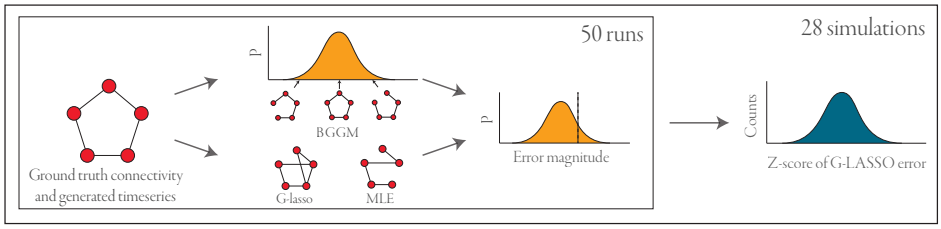


FIGURE 6.2: The evaluation procedure of the simulated fMRI data. First, both the posterior distribution $P(G, \mathbf{R} \mid \mathbf{X})$ and the point estimates (for the graphical lasso or maximum likelihood estimate) are determined. Subsequently the error compared to the ground truth is computed for all samples in the approximated distribution as well as for the point estimates (see text for this procedure). These results are summarized by computing the z-score for the point estimate error relative to the distribution of errors obtained from the Bayesian approach. Finally, the z-scores are aggregated across the runs, resulting in a histogram of error z-scores for each simulation.

the reconstruction error (where $\mathbf{R}$ is either a sample from $P(G, \mathbf{R})$, or a point estimate). The total reconstruction error is $\eta(\Gamma) = \frac{2}{p(p-1)} \sum_{i<j} \gamma_{ij}$, the true positive error is $\eta_{tp}(\Gamma) = \frac{1}{N_{tp}} \sum_{i<j} \gamma_{ij} \delta_{t_{ij} \neq 0}$, where N_{tp} is the number of nonzero elements in the ground truth $\mathbf{R}^*$, i.e. the number of true present connections, and finally the true negative error is given by $\eta_{tn}(\Gamma) = \frac{1}{N_{tn}} \sum_{i<j} \gamma_{ij} \delta_{t_{ij}=0}$, where N_{tn} is the number of zero elements in the ground truth $\mathbf{R}^*$, i.e. the number of true absent connections. The indicator function δ_x evaluates to 1 if and only if its argument x holds true, and to 0 otherwise.

In [Smith et al., 2011], a null distribution is computed for each of the different methods, by randomly permuting the node labels in the different runs (to remove any influence between the different nodes), which is subsequently used to derive a z-score for an error measure similar to η. However, in the case of Bayesian functional connectivity, a distribution characterizing the uncertainty of the results is already available in the form of $P(G, \mathbf{R})$. By applying η to each of the samples of this distribution, we obtain $P(\eta)$. As we find that for each pair of simulations and runs, $P(\eta)$ is is closely approximated by a Gaussian distribution (Kolmogorov-Smirnov test, $p < 1e-5$), the z-scores of a point estimate $\mathbf{R}$ relative to the BGGM distribution may be computed as $z(\mathbf{R}) = (\eta(\mathbf{R}^*, \mathbf{R}) - \mu)/\sigma$, in which μ and σ are the mean and standard deviation of the distribution, respectively. The procedure is illustrated in Figure 6.2.

Uncertainty in connectivity distributions

The Bayesian formulation of the model allows us to describe and compare the shapes of the different posterior distributions. We compute the entropy of the posterior distributions as

$$H = -\sum_{G} \left[P(G, \mathbf{K} \mid \mathbf{X}) \log_2 P(G, \mathbf{K} \mid \mathbf{X}) \right] \quad , \tag{6.13}$$

to indicate the diversity of models that have been encountered in the Markov chains. In addition, the posterior probability of the maximum a posteriori sample is derived, i.e.

$$P(\hat{G}, \mathbf{K} \mid \mathbf{X}) = \max_{G} P(G, \mathbf{K} \mid \mathbf{X}) \ , \tag{6.14}$$

to quantify how much of the posterior distribution is dominated by its mode.

Approximate inference

The Markov chain Monte Carlo (MCMC) scheme as described in Appendix C.1 was used to approximate the posterior distributions of interest for each subject using either the simulated BOLD signal time series, the BOLD time series data for the fourteen subcortical regions (see (6.6)), the combination of time series data and tractography output for the subcortical regions (see (6.11)) or finally the BOLD time series data in combination with the informed prior. Throughout, a vague prior on the precision is used: $P(\mathbf{K} \mid G) = \mathcal{W}_G(3, \mathbf{I}_p)$, cf. [Moghaddam et al., 2009]. The parameters of the probabilistic streamline model are set to $(\delta_1, \delta_0) = (1, 0.5)$, which expresses that high streamline counts are most likely associated with a structural connection, while still allowing for tractography noise [Janssen et al., 2014]. Once convergence was established (see Appendix C.2), the approximated distributions were uniformly thinned to $T = 1\,000$ samples, to make subsequent analyses more manageable and to have an equal number of samples for all different settings.

Materials

The acquired data consist of a T1 anatomical scan, resting-state functional data and diffusion-weighted images (DWI), collected for each subject. We refer the reader to [van Oort et al., 2014] for details of the acquisition protocol. Preprocessing steps were performed using FSL 5.0 [Jenkinson et al., 2012] with default settings unless otherwise specified.

Preprocessing of the resting-state functional MRI data consisted of the following steps. T1 images were linearly registered to MNI-152 space. Multi-echo volumes at each TR were combined [Poser et al., 2006]. Motion correction was performed using MCFLIRT and estimated motion parameters were regressed out together with their temporal derivatives and mean time courses for both WM and CSF. Finally, data were high-pass filtered at 0.001 Hz. Note that we did not apply global signal regression, as this step is known to introduce artifactual negative correlations [Murphy et al., 2009; van Dijk et al., 2010].

Preprocessing of the DWI data was conducted using FSL FDT [Behrens et al., 2003a] and consisted of motion correction, correction for eddy currents and estimation of the diffusion parameters. To obtain a measure of white-matter connectivity, FDT Probtrackx 2.0 [Behrens et al., 2003a, 2007] was used with seed voxel to target voxel tracking. Structural scans were segmented using FAST [Zhang et al., 2001] and FIRST [Patenaude et al., 2011] to generate seed and target voxels. Seed voxels were those voxels in the cortical gray matter mask with a non-zero

white-matter partial volume estimate and the outermost voxels of the subcortical masks. The remainder of the cortical and subcortical voxels served as target voxels. In addition, streamlines were terminated once they hit the target mask. This prevents polysynaptic connections being erroneously interpreted as direct connections.

Finally, subcortical structures were segmented using FSL FIRST [Patenaude et al., 2011], resulting in data for a total of fourteen regions, consisting of bilateral accumbens, amygdala, caudate, hippocampus, pallidum, putamen and thalamus. For the functional data, for each of these regions the signal was averaged over all voxels in that region and subsequently standardized to have zero mean and unit variance. For the streamline data, all streamline counts were aggregated over pairs of voxels in pairs of regions, resulting in a 14×14 matrix of streamline counts.

6.2 RESULTS

Below we discuss both the simulation results as well as the connectivity estimates obtained on empirical data. For readability, we refer to the probability of conditional dependence as 'connection probability' and to a pair of regions that are conditionally independent or not-independent, conditioned on all other variables, simply as an 'independent' (or 'disconnected') or 'dependent' (or 'connected') region pair, respectively.

Simulation results

Figure 6.3 shows the (smoothed) histograms of z-scores aggregated over the 50 runs per simulation, for the graphical lasso approach with $\lambda = 100$ (the results for $\lambda = 5$ and the MLE are almost identical; the MLE results are shown in Supplementary Figure C.2). In this figure, distributions of errors with high z-scores have substantially larger errors than the errors from the BGGM approach, while distributions with low z-scores have smaller errors. The significance threshold at $p < 0.01$ is indicated by the red dotted lines. The first row of Figure 6.3 shows the total scores (both true positives and true negatives) for each simulation, while the second and the third row split this score into the contributions for true positive connections and true negative connections, respectively. These results indicate that in terms of true positives, the lasso approach typically has an equal to slightly better performance than our Bayesian alternative. However, the BGGM approach identifies true negatives at least as well as G-lasso, and in several cases significantly outperforms it. On the whole, the proposed method is up to par with the graphical lasso (for $\lambda \in \{5, 100\}$) and the MLE, while at times outperforming them greatly.

To obtain insight in the behavior that creates these results, we take a closer look at some of the simulation results. As an example, Figure 6.4A shows the ground truth network and the reconstruction by the graphical lasso, as well as the expectation (i.e. posterior mean of the samples) using the BGGM approach. In addition, the figure shows for three different connections the estimated partial correlation in detail. The first, between nodes 1 and 5, is present in the ground truth network. Our approximation is (correctly) confident that this node pair is not independent, and assigns a posterior partial correlation distribution close to the ground truth.

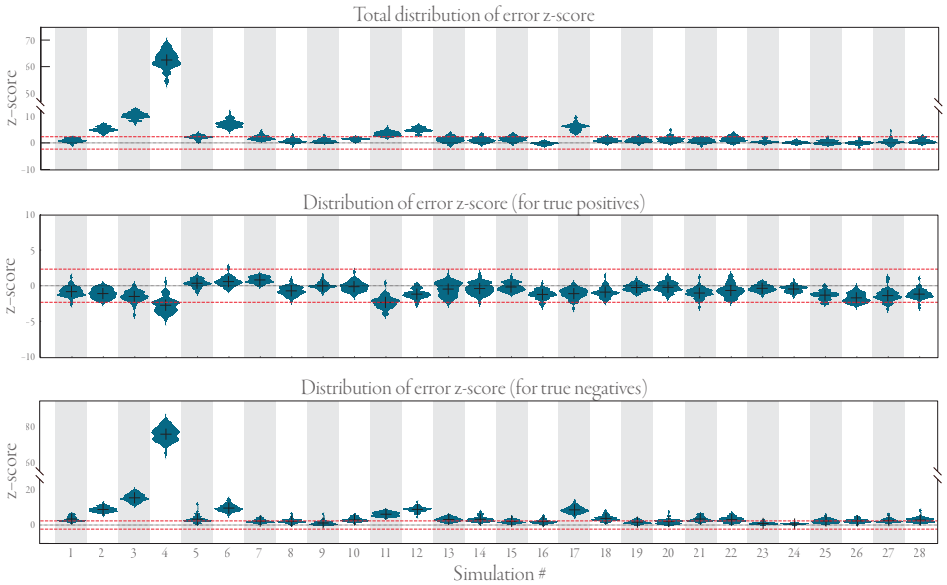


FIGURE 6.3: The histograms for each of the 28 different simulations. Positive error z-scores indicate that the point estimate was less effective in recovering the ground truth than the Gaussian graphical model, while the reverse is true for negative error z-scores. The red dashed lines indicate the interval outside of which the difference in performance is significant ($p < 0.01$, z-test). Note the different ordinate axes.

The graphical lasso estimate is slightly closer to the ground truth than the mode of the distribution. For the second node pair, between nodes 3 and 5, a connection should be absent, but because of the limited number of data samples the signals of these nodes have become correlated. This time, the BGGM approach shows a bimodal distribution. The first mode is centered close to the graphical lasso estimate, but the second mode is at zero, as there is non-negligible evidence for this pair of nodes being disconnected. This means that on the whole (i.e. the entire distribution), the BGGM approach correctly estimates this connection strength lower than the graphical lasso. A similar observation can be made for the third node pair, between nodes 1 and 4, of which the BGGM estimate is fairly certain about their independence. Because of this, most of the partial correlation mass is at zero, rather than at the value indicated by the graphical lasso estimate.

These results beg the question: what if we regularize the graphical lasso even more? Although Smith et al. report no further improvement after $\lambda = 100$ [Smith et al., 2011], it is possible that more regularization brings the G-lasso estimate closer to the BGGM results. In Figure 6.4B, the same visualization is provided, but this time for $\lambda = 10\,000$. This time, we see that indeed the graphical lasso estimate is closer to the BGGM expectation than before. In particular for the connection between nodes 1 and 4, the G-lasso now correctly estimates the absence of this connection. However, for the connection between nodes 3 and 5, the results hardly change, which means that the BGGM estimate is closer to the ground truth still, as, conditioned on the absent

connection, the estimated partial correlation is zero. Finally, for the true positive connection between nodes 1 and 5, we see that the strong regularization causes the graphical lasso to underestimate the connection, which will only become worse when we increment λ even further.

The pattern of simulations in which the BGGM outperforms the graphical lasso is not random. In [Smith et al., 2011], each of the simulations is based on a network consisting of 5 nodes, except for simulations 2, 3, 4, 6, 11, 12 and 17, which consist of networks of 10, 15, 50, 10, 10 and 10 nodes, respectively. Precisely these simulations benefit the most from the BGGM approach, as can be seen in Figure 6.3. As for these simulations the ratio N/p is smallest, it is here that the most improvement can be obtained from regularization, e.g. by the graphical lasso [Smith et al., 2011]. As we have shown above, the BGGM provides further improvement still, because this approach conditions on conditional independencies.

We further analyzed the effect of sample size on recovery of the ground truth by taking the simulation with the most available samples (simulation 7 in [Smith et al., 2011]) and attempting to recover the ground truth using increasingly smaller subsets of the samples. We compared the BGGM results with the graphical lasso with $\lambda \in \{5, 100, 1\,000, 10\,000\}$. The outcome of this experiment is shown in Figure 6.5, once again split into the total error, error in recovery of true positives and error in recovery of true negatives. The results indicate that for small sample size, the BGGM approach already outperforms the graphical lasso in total error, although the differences become more pronounced as more samples are considered. Extremely strong regularization (i.e. $\lambda = 10\,000$) does result in better estimation of absent connections (by simply forcing almost all connections to zero), but this comes at the cost of excluding connections that should be present. For weak regularization (i.e. $\lambda = 5$), small sample size appears to be somewhat beneficial in recovery of true positive connections, as here the performance of the graphical lasso is similar to our approach. However, this effect diminishes as more samples are acquired (inducing more spurious connections). In terms of true negatives, weak regularization is clearly outperformed by the BGGM approach.

In addition, we analyzed the effect of small sample sizes on the estimates. We used simulation 3 (with $p = 15$) and repeated the procedure as before, but this time the number of samples was varied $n \in \{5, 10, \ldots, 45, 50\}$, so that situations of $n < p$ were included. The results of this experiment are shown in Figure 6.6. They show that, unsurprisingly, weak regularization (i.e. $\lambda = 5$) is insufficient to recover the ground truth when few samples are available. Strong shrinkage (i.e. $\lambda = 10\,000$) results in a low recovery error, but this comes at the expense of significantly underestimating true positive connections. In general, the BGGM approach performs approximately equal to the graphical LASSO for small to moderate regularization, given this limited sample size scenario.

Empirical results

Below we discuss the connectivity estimates we obtained on the empirical data, for the original BGGM model, the data fusion variant and the effect of incorporating background information.

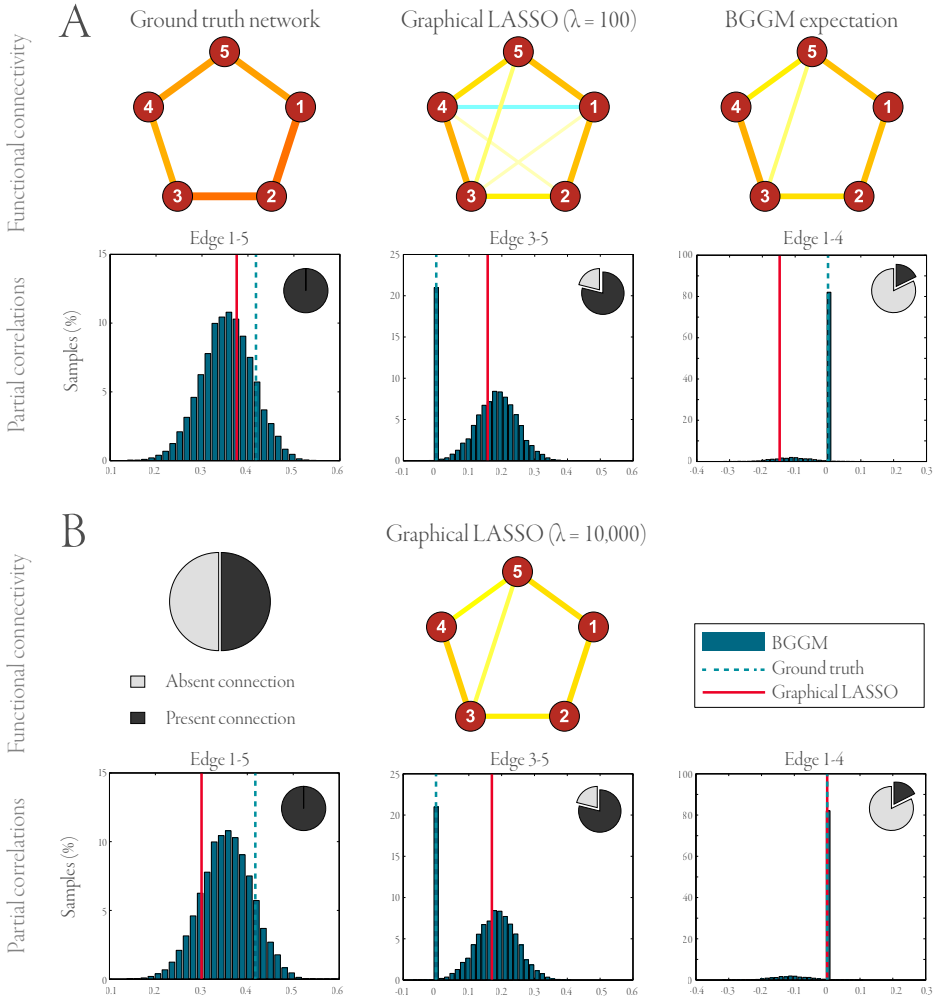


FIGURE 6.4: A. Simulation details. First row: the ground truth connectivity of one run of simulation 1, as well as the constructions by the graphical lasso ($\lambda = 100$) and the expectation of the Gaussian graphical model approach. Second row: estimated partial correlation for a true positive connection, a true negative connection with strong empirical correlation, and a true negative connection with weak empirical correlation. B. The same, but with stronger regularization for the graphical lasso ($\lambda = 10\,000$). This time, the G-lasso estimate is similar to the BGGM expectation for connection 1–4, but over-regularizes the true positive connection 1–5.

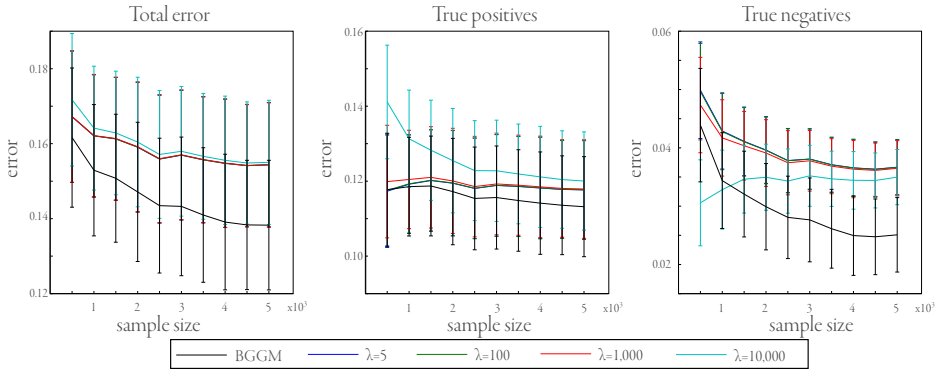


FIGURE 6.5: Effect of different sample sizes in recovery of ground truth connectivity, for the BGGM approach as well as for the graphical lasso with $\lambda \in \{5, 100, 1\,000, 10\,000\}$. Error bars indicate one standard deviation over the 50 runs. For the BGGM approach, the error bars indicate one standard deviation over the expectations of the runs.

Functional connectivity distributions

For all twenty subjects, functional connectivity was estimated as the posterior distribution over conditional independence graphs and partial correlation structures. We find that there is minor inter-subject variability in the number of identified non-independencies, as indicated by a small standard deviation of the mean expected density across subjects, of 0.62 ($SD = 0.04$).

For the subject with the sparsest dependency structure, its mean posterior conditional independence graph as well as its mean posterior partial correlations are shown as adjacency matrices in Figure 6.7. The conditional independence graph for this subject has a mean density of 0.55 ($SD = 0.03$). From Figure 6.7 it can be seen that a number of connections are present, while supporting a partial correlation close to zero. Most likely, these connections support dependencies that are induced by noise in the data, rather than true connections between subcortical regions. This is further supported by looking at the (variance of the) group-averaged results: Supplementary Figure C.3A shows the group-average of the mean posterior connectivity estimates for all subjects. This reveals that no pairs of regions can consistently be marked as independent. However, a stable backbone of connections that are clearly dependent exists within both hemispheres, consisting bilaterally of accumbens — caudate, amygdala — hippocampus, pallidum — putamen, caudate — thalamus and hippocampus — thalamus, that each have a mean posterior connection probability of $\geqslant 0.94$ and partial correlations in the range $[0.15, 0.58]$. Similarly, a number of connections appear stable between hemispheres. Interhemispheric connectivity consists predominantly of connections between functionally homologous regions, which have mean posterior connection probabilities of $\geqslant 0.95$ and partial correlations in the range $[0.35, 0.73]$. Other strong interhemispheric connections with probability $\geqslant 0.90$ consist of left amygdala — right hippocampus, left caudate — right thalamus, left hippocampus — right putamen, left accumbens — right caudate and left caudate — right accumbens, all with nega-

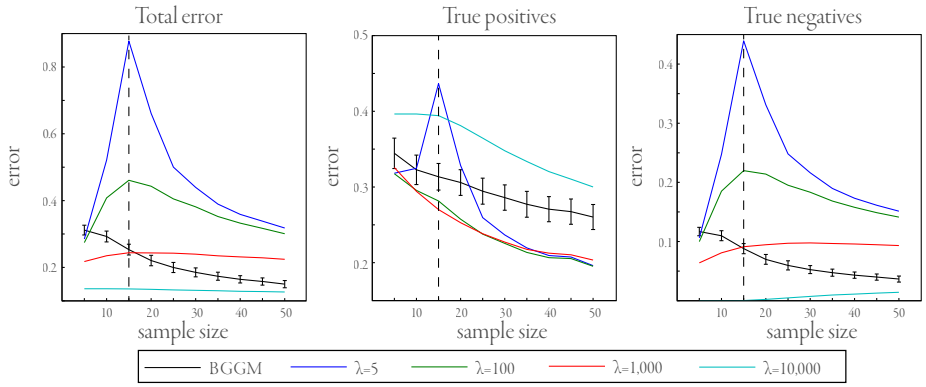


FIGURE 6.6: Effect of small sample sizes, including $n < p$, in recovery of ground truth connectivity, for the BGGM approach as well as for the graphical LASSO with $\lambda \in \{5, 100, 1\,000, 10\,000\}$. Error bars indicate one standard deviation over the 50 runs. For the BGGM approach, the error bars indicate one standard deviation over the expectations of the runs.

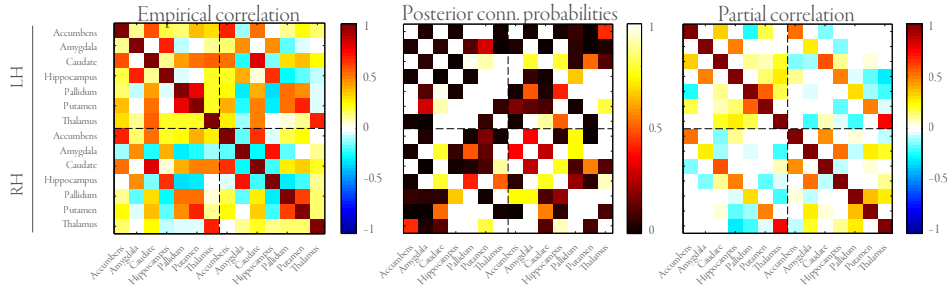


FIGURE 6.7: Subcortical connectivity for one subject. From left to right: the empirical correlation matrix, the mean posterior connection probability matrix and the mean posterior partial correlation matrix. The connections for the left hemisphere (LH) and the right hemisphere (RH) are separated by the dashed lines.

tive partial correlations in the range $[-0.23, -0.16]$, mimicking the structure found within the hemispheres.

The between-subject standard deviation of the mean posterior estimates, as shown in Supplementary Figure C.3A, shows that although there is quite some between-subject variability in terms of conditional independencies, the partial correlation structures are very stable. This indicates that the Bayesian Gaussian graphical model approach explores many dependencies in the data, which can vary across subjects but contribute little to the overall partial correlation structure as they correspond to small partial correlations.

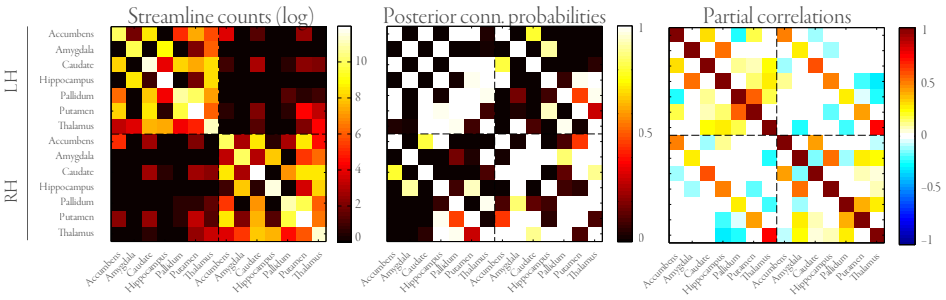


FIGURE 6.8: Subcortical connectivity for one subject using the data fusion model. From left to right: the empirical streamline log-counts, the mean posterior connection probability matrix and the mean posterior partial correlation matrix. Note the reduction in connectivity, in particular between the hemispheres, compared to Figure 6.7. The connections for the left hemisphere (LH) and the right hemisphere (RH) are separated by the dashed lines.

Bayesian data fusion

Similar to the previous section, functional connectivity was again estimated for all twenty subjects, but this time using the data fusion approach. This implies that the conditional independence graph is now interpreted as an estimate of structural connectivity, informed by both resting state fMRI as well as probabilistic tractography. In Figure 6.8, the adjacency matrices of the mean posterior estimates are shown for the same subject as used previously. Overall, the same backbone of functional connectivity is visible as when using only the fMRI data. However, there are a number of differences. In particular, adding information from probabilistic streamlines leads to substantially sparser mean network density: for this subject the density drops to 0.46 (SD $=$ 0.02). In addition, particular connections change from predominantly absent to predominantly present, and vice versa. Figure 6.9 shows for this subject some of the connections with the largest difference in mean posterior partial correlation. This indicates that the addition of tractography data can both add and remove connections. In general however, we see that the dependencies that are removed due to the addition of tractography data, are those that supported small partial correlations.

In Supplementary Figure C.3B, the aggregated connectivity results are shown for all twenty subjects, as well as the standard deviations of these estimates. This reveals that the uncertainty about the retrieved connectivity decreases by adding the tractography data. Interestingly, although the expectations of the partial correlation estimates hardly change compared to the previous model (compare e.g. Figures 6.7 and 6.8), the variance of these estimates does decrease. Most likely, this is due to the fact that the bimodal behavior of partial correlations (as was observed in the simulation, where one mode is present for $g_{ij} = 1$ and one for $g_{ij} = 0$) becomes unimodal as the tractography data gives more stringent estimates of G.

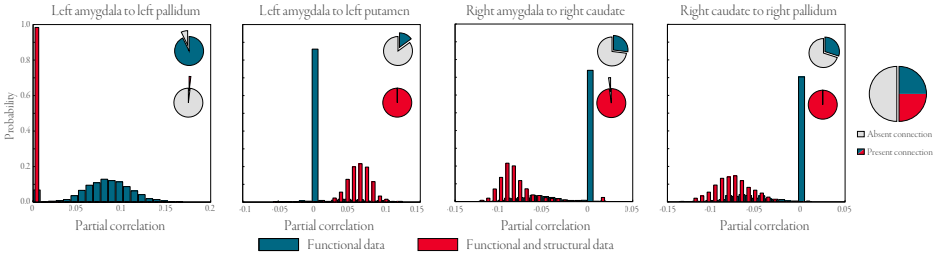


FIGURE 6.9: Examples of differences in partial correlation estimates between the BGGM estimates and the data fusion approach.

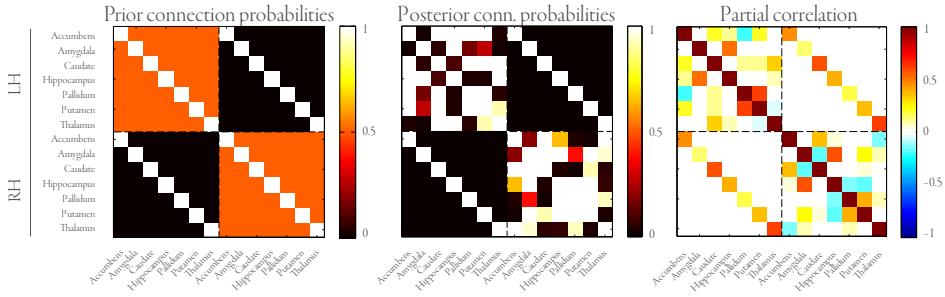


FIGURE 6.10: Subcortical connectivity for one subject using the informative prior. From left to right: the prior probability of a non-independence, the mean posterior connection probability matrix and the mean posterior partial correlation matrix. The connections for the left hemisphere (LH) and the right hemisphere (RH) are separated by the dashed lines.

Incorporating background knowledge

Here we discuss the effects of assuming a priori that interhemispheric connectivity must follow the connections between the functionally homologous regions. As this prior restricts interhemispheric connections even more than the data fusion model, the network densities decrease even further. For the subject that was used as an example earlier, the density now drops to 0.34 (SD = 0.02). Of course, this follows directly from the definition of the prior, that simply excludes a number of connections. Because of this absence of interhemispheric connections, dependencies between regions in different hemispheres must now follow a longer path via the homotopic connection. As a consequence, some of the intrahemispheric connections have an increased probability of being dependent, which we quantify by considering the density within hemispheres only. Aggregated over all subjects, we find that using the prior results in a mean density within hemispheres of 0.66 (SD = 0.07), slightly higher than for the initial model that has a mean density within hemispheres of 0.63 (SD = 0.05). The aggregated results as well as their standard deviations are shown in Supplementary Figure C.3C. This further shows that, similar to the data fusion model results, the variance of the elements *within* hemispheres is decreased as well, by restricting the connectivity *between* hemispheres.

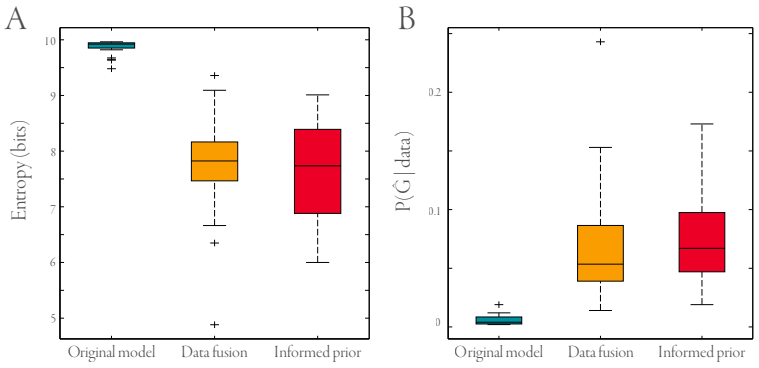


FIGURE 6.11: Differences in posterior distribution shapes. A. Entropy of the posterior distribution. B. Posterior probability of the mode $\hat{G}$. We refer to the prior distribution defined in (6.8) as the vague prior.

Comparing the different distributions

Both the data fusion model as well as the usage of the informed prior pose restrictions on the posterior distribution of connectivity. This effect is illustrated by computing the entropy of the different approaches, as shown in Figure 6.11. Whereas for the original model the posterior distribution appears very broad, both alternative specifications decrease this uncertainty. In particular for the data fusion approach, one subject has a maximum a posteriori estimate with probability as high as 0.24, compared to only 0.02 when using only fMRI data. A similar picture arises by counting the fraction of unique models in each of the distributions. Here, we see that the original model has its probability density spread across many independency structures ($96\% \pm 5$ of the visited samples are unique), while the extended models are more peaked around a few high probability samples ($46\% \pm 13$ and $45\% \pm 13$ of the samples are unique).

The differences between the three approaches to connectivity are further illustrated by the scatter plot in Figure 6.12. Here, for all connections across all subjects the expectations of the original model compared to the two extensions are shown. The first row of Figure 6.12 shows that data fusion results in decreased connectivity between hemispheres. The latter connections may be less likely in the alternative model, but are not forced to zero. Partial correlations remain largely unaffected, as shown in Figure 6.12C, except for a few interhemispheric connections that become excluded by the tractography data and therefore are assigned zero partial correlation. The informed prior puts all interhemispheric connections to zero, as seen in Figure 6.12B, except for the homologous connections that have probability close to one in both models. Out of the two extensions, this approach has the most influence on the partial correlation results, as evidenced by Figure 6.12D. Here, not only are the interhemispheric partial correlations that do not correspond to homotopic connectivity set to zero, most other connections have lower partial correlations. This suggests that the partial correlations that are present in the original approach must be compensated by other, stronger, connections, which is no longer necessary with this prior.

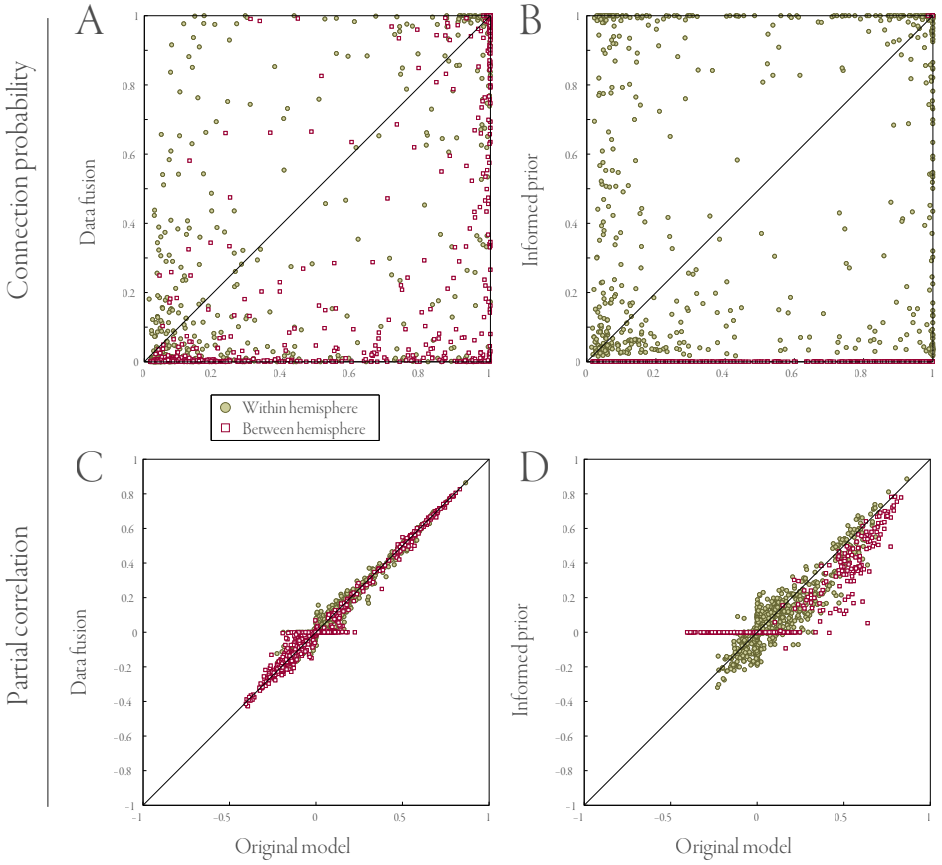


FIGURE 6.12: Scatter plot of the expectations of connection probabilities and partial correlations. The top row shows the connection probabilities for the two model extensions versus the original model. The bottom row is the same, but for partial correlations.

In Figure 6.13, the variance of the connection probabilities and partial correlations is shown. In the data fusion approach, some of the connections and partial correlations become much more precise, as shown by a lower variance (typically those connections for which no streamline data is present and which, as a result are excluded). Simultaneously, some partial correlations in fact have a larger variance (see Figure 6.13C), which indicates that for these connections the BOLD time series and the probabilistic streamlines contradict one another. Lastly, the informed prior obviously decreases the variance for interhemispheric connections, both in connectivity and partial correlations. For the intrahemispheric connections (about which the prior is the same as in the original model), the variance of both connectivity and partial correlations appears to remain largely unaffected. The variance of partial correlations for the connections between

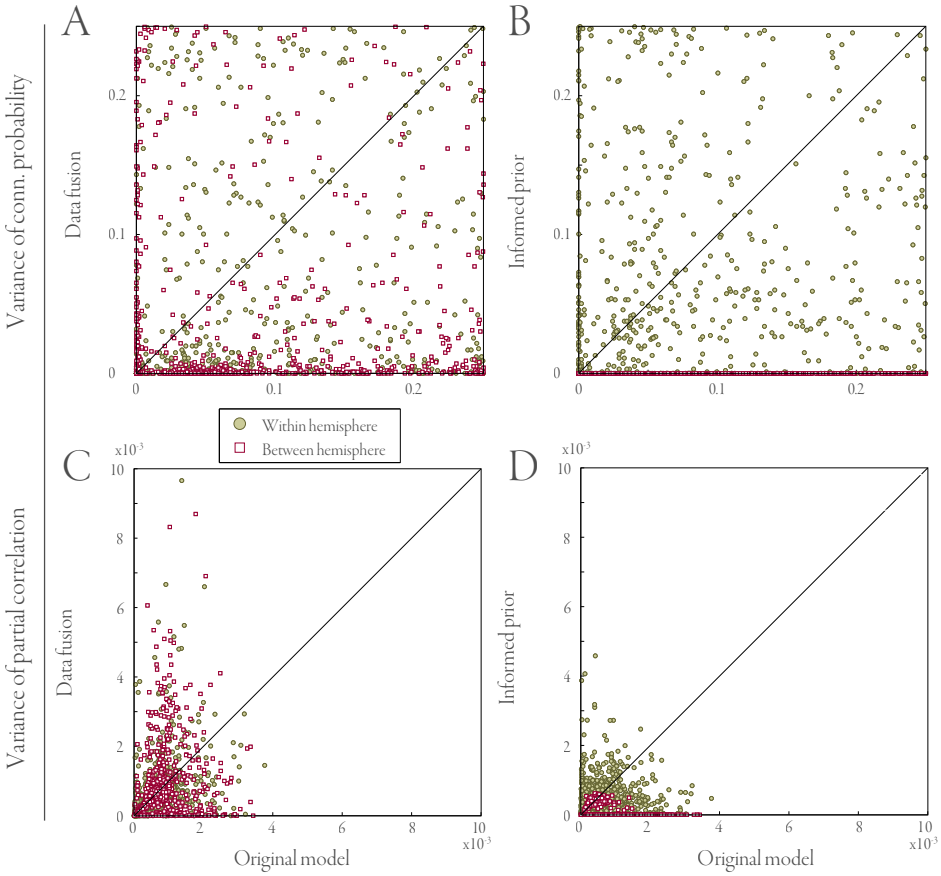


FIGURE 6.13: Scatter plot of the variances of connectivity and partial correlations. The top row shows the variance of connections for the two model extensions versus the original model. The bottom row is the same, but for the variance of partial correlations.

functional homologues decreases marginally, as shown by a mean variance of 4.0e—4 compared to 4.9e—4 for the original functional connectivity model.

6.3 DISCUSSION

Functional connectivity may be quantified using different metrics. The most obvious approach is to use Pearson correlation, but this metric is sensitive to polysynaptic influences. An alternative that does not suffer from this drawback is partial correlation, which was further advocated for its ability to retrieve true connections and its capacity to deal with noise [Smith et al., 2011]. Partial correlation between two variables may be interpreted as Pearson correlation conditioned

on all other variables. In practice, partial correlation can be computed by applying a simple transformation to the precision matrix of a multivariate Gaussian distribution. The precision matrix and, consequently, the matrix of partial correlations, has the interesting property that conditional independence between variables, given all other variables, appears as a zero value in the corresponding matrix element [Marrelec et al., 2006], which may conveniently be collected in a conditional independence graph. Typically, this graph is mostly ignored, while the precision or partial correlation matrix is considered the quantity of interest. In this chapter, we have provided a Bayesian generative model for functional connectivity in which the conditional independence graph plays a central role, as it is assumed to generate the precision matrix and thus functional connectivity. As opposed to regularized maximum likelihood estimates for the precision matrix, our approach characterizes the full posterior distribution of both conditional (in)dependencies and partial correlations. In addition to this model, we described a number of model variants that address specific issues with, and conceptual extensions to, connectivity.

We subjected our approach to the simulations that were presented in [Smith et al., 2011], and compared its performance to the maximum likelihood estimate as well as to the graphical lasso. The latter of these two has been shown to be the most successful in recovering connectivity in these simulations [Smith et al., 2011]. The results of the simulation are encouraging. Although we observe that for true positive connections, our approach occasionally underestimates connections, it more than compensates for this in correctly estimating true negatives (i.e. the sparsity structure of the network). When true positives and true negatives are both taken into account, corrected for their respective numbers of occurrence, we find that our approach performs at least as well as the graphical lasso, and significantly better for simulations with small sample size compared to the number of nodes in the network. A closer look at these results shows that when estimating partial correlations, conditioning on the presence or absence of a connection provides a considerable advantage over shrinkage. In particular for connections with a moderate probability of independence our method yields a bimodal distribution of partial correlations, differentiating between the conditionally dependent and independent node pairs.

In addition to our simulation results, we used our approach to approximate the posterior distribution of functional connectivity between subcortical areas for twenty participants. This allowed us to identify a connectivity backbone that consists of strong connections and partial correlations. At the same time, we see that a number of connections are strongly dependent, but foster only weak partial correlations. This emphasizes that a richer picture of connectivity is obtained by looking at both the structure of conditional independence, as well as the strength of these connections in terms of partial correlation.

Partial-correlation based methods are susceptible to common input effects that may induce spurious connections if they are not accounted for, for example when variables (i.e. brain regions) are missing [Woolrich and Stephan, 2013; Zalesky et al., 2012] from the analysis. If instead the full neural system is observed, it is straightforward that direct functional connections presuppose anatomical connections between the corresponding regions. This allows us to combine the generative model for functional connectivity with a similar model for structural connectivity from Chapter 2 using probabilistic tractography obtained from diffusion weighted MRI. Conceptually, this results in a data fusion model in which an underlying model of anatomy

drives both the observations for functional interactions, as well as for estimates of structural fibres. Compared to alternatives that, for example, weigh a regularization parameter by the strength of structural connectivity [Sui et al., 2011; Rykhlevskaia et al., 2008; Ng et al., 2011; Pineda-Pardo et al., 2014; Calamante et al., 2013; Bowman et al., 2012], our approach is based on a generative model in which data fusion is made possible by the use of different likelihood terms. Furthermore, in our model both sources of data affect both types of connectivity; structural connectivity regularizes functional connectivity and simultaneously functional dependencies influence the probability of structural connections. On empirical data the data fusion approach leads to sparser connectivity, in particular between hemispheres. However, some connections are conditionally dependent to such a degree that the model infers a connection regardless of the lack of support by the tractography data. This is helpful in estimating structural connectivity, as it is well known that structural connectivity based on diffusion weighted imaging suffers from a large number of false negatives [Jbabdi et al., 2007]. In addition, data fusion lowers the variance for many of the partial correlations, indicating that combining both imaging modalities leads to more robust estimates [van Dijk et al., 2010; Jbabdi et al., 2007; Damoiseaux and Greicius, 2009; Biessmann et al., 2011]. However, for a number of connections the data for functional and structural connectivity appear to contradict each other, which actually results in increased variance. Note that our data fusion approach has similarities to linked ICA [Groves et al., 2011], which also uses a Bayesian generative model to integrate different data modalities. However, whereas linked ICA assumes that each data modality may be decomposed into a number of (shared) components, our model assumes that anatomical connectivity is the variable that is shared across modalities.

Our final model variant uses an informative prior which encodes the assumption that between-hemisphere connections are restricted to those between functionally homologous regions (cf. for example [Gloor et al., 1993]). This is only one of many prior distributions that, depending on the research question and available background information, may be used to inform the connectivity estimates. As expected, the prior removes the negative partial correlations that are visible for contralateral connections in the other model variants. Indirectly, the prior also affects the partial correlations within hemispheres, as they become slightly lower in magnitude across the board. These results touch upon an unresolved issue in connectomics concerning the interpretation of negative (partial) correlations. It has been suggested that a substantial number of negative partial correlations are due to global signal regression and are therefore artifactual in nature rather than biological [Liang et al., 2012; Carbonell et al., 2014; Schwarz and McGonigle, 2011]. On the other hand, it has been shown that even without global signal regression, negative connections exist and these may even have biological meaning [Chen et al., 2011]. Although it is outside the scope of this chapter to resolve this matter, we have shown that an informed prior may be used to encode such assumptions or correct for biases.

The shapes of the posterior distributions reveal that none of the model variants are dominated by their mode. In particular for the original model the distributions are very broad and contain many unique models. Although a number of connections is consistently present, the conditional independence graphs vary substantially across subjects. In contrast, the data fusion approach and the informed prior result in distributions that are more tightly centered around

the maximum a posteriori connectivity, yet even here there remains substantial support for alternative models. This has important implications for connectomics studies. These are typically aimed at obtaining a point estimate (which can often be interpreted as the mode of an implicit posterior distribution), so a substantial number of connections with significant support from the data will be excluded and spurious connections will be suggested. The widths of the posterior distributions strongly advocate a Bayesian approach, or at the very least point-estimated connectivity results should be treated with great care, e.g. by applying a bootstrapping procedure [Hastie et al., 2009].

The main limitation of our study is one of scale. Bayesian inference has the drawback of being computationally demanding in approximating the posterior distributions, and although state-of-the-art machinery has been applied to make this process efficient, it remains impossible to apply the same methods to a large number of variables. Applying the models to large-scale datasets requires either more efficient implementations, e.g. by using GPU programming, or additional efficiency gains in the field of Gaussian graphical models.

Finally, a fundamental assumption in Gaussian graphical model estimation is that the functional data is normally distributed. In case this assumption fails, it may prove difficult to interpret the estimated connectivity. However, as discussed by [Hlinka et al., 2011], BOLD time series do tend to be mostly Gaussian.

The most pressing issue for future work is, as mentioned above, improving the methodology to handle a larger number of variables. However, a number of interesting research questions may be addressed even with a limited number of regions. For example, a model may be constructed that defines the BOLD time series to be generated by a mixture of partial correlation matrices, instead of a single one. By applying appropriate constraints, such as that consecutive datapoints are likely to be generated by the same connectivity matrix, this setup can be applied to differentiate experimental conditions based on their connectivity distributions [Schmidt and Mørup, 2013]. Similarly, subjects may be assigned to either patients or healthy controls by defining a shared conditional independence graph for either group.

The data fusion approach may be extended to incorporate any number of imaging modalities, provided that a forward model can be constructed that shares at least one variable with the other modalities. For example, structural connectivity may inform functional connectivity estimated from MEG instead of or in addition to fMRI data [Pineda-Pardo et al., 2014].

Additional information may also be incorporated into the prior. This may be explicit evidence for (the absence of) a connection, e.g. tracer studies that reveal the presence of a fiber bundle can make particular connections more likely or, conversely, knowledge about white-matter lesions may preclude connections. Alternatively one could construct a prior in which the probability of a connection is a function of the distance between the corresponding end points.

In conclusion, the proposed Bayesian approach to functional connectivity has demonstrated that connectivity may be meaningfully divided into structure and strength. Several model variants have been discussed, each with their own characteristics. Application of the models has shown convincingly that multiple unique structures are possible given the same data. This il-

lustrates the advantages of a Bayesian approach to connectivity, and provides a word of caution for traditional (regularized) maximum likelihood estimators.

7

DISCUSSION

The goal of this thesis was to apply Bayesian probability theory in order to explicitly quantify the uncertainty associated with estimates of brain connectivity. This approach was applied to structural connectivity in Chapters 2 and 3, subsequently to functional connectivity in Chapters 4 and 5 and finally to the simultaneous estimation of structural and functional connectivity in Chapter 6. Here, I summarize the scientific contributions of the individual chapters, discuss the limitations of my work and provide an outlook of the field of connectomics.

7.1 SUMMARY OF THE PRESENTED WORK

Chapter 2: Bayesian inference of structural connectivity

Determining structural connectivity using (probabilistic) tractography on diffusion-weighted MRI data is often performed in an ad-hoc manner: By applying a threshold to the number of streamlines that run between two regions of interest, the existence of a connection for this region pair is determined. The result is a single, binary, network that describes which pairs of regions may be considered anatomically connected. In Chapter 2, we propose a Bayesian alternative solution. The solution consists of a generative model which postulates that probabilistic streamlines emanating from a region-of-interest are distributed according to a multinomial distribution. The corresponding prior is a Dirichlet distribution, whose hyperparameters are modulated by the presence or absence of a structural connection. Specifically, the hyperparameters capture the prior belief that we are agnostic about streamline count distributions for connected regions[1], while we expect few streamlines when a connection is absent (in this case streamlines may be considered false positives).

Two different prior distributions on connectivity are considered to complete the generative model. The first expresses the prior belief that structural networks have a particular density, i.e.

[1]Note that the number or fraction of streamlines connecting pairs of regions provides little information regarding the *strength* of a connection, so existing connections may support widely varying streamline counts [Jones et al., 2013; O'Donnell and Pasternak, 2014; Janssen et al., 2014].

number of edges. The second prior distribution is an approximation of an hierarchical model, where it is assumed that different subjects in a population have similar connectivity.

The posterior distributions representing the support for different networks, given the data and our prior beliefs, are approximated using a Markov chain Monte Carlo (MCMC) sampling scheme. The resulting distributions of connectivity are validated by comparing the structural estimates with the conditional independencies found in resting-state functional connectivity for the same subjects. Our results show that the model with a prior on network density performs similarly to thresholding approaches. The approximated hierarchical prior consistently outperforms the first model as well as thresholding approaches, showing that the additional information contained in the hierarchical coupling between subjects improves the estimates of structural connectivity.

By capturing the posterior distributions of connectivity, rather than a point estimate, we quantify the uncertainty of our estimates within individual subjects. We show how this can be used to, for example, compute marginal connection probabilities. Additional demonstration of how uncertainty propagates to measures from graph theory is provided in [Janssen et al., 2014].

Chapter 3: A priori clustered structural connectivity

Structural connectivity is often used to inform a connectivity-based parcellation. However, possible errors in connectivity estimates propagate and induce errors in these parcellations. Furthermore, connectivity-based clustering algorithms have a number of implicit assumptions that influence the resulting parcellations. In Chapter 3 we describe a Bayesian procedure for connectivity-based parcellation, in which estimation of connectivity and the clustering thereof are intertwined. Using probabilistic streamline data, we recover the posterior distribution over the connectivity parameters, the cluster assignments of individual nodes, the link probabilities between clusters and lastly the number of clusters. The result is a probabilistic interpretation of parcellation based on connectivity. This allows us to compute, among other things, the marginal probability that two nodes belong in the same cluster.

No explicit assumptions were made about the connectivity behavior of the different clusters, nor about the number of clusters that our approach should select. Instead, the proposed model was able to extract the number of clusters from the data. Our results identified two different types of clusters: densely intra-connected clusters corresponding predominantly to cortical lobes, as well as small clusters containing hub nodes, that interconnect the other clusters. Interestingly, the regions in the latter type of clusters have each been pointed out as 'rich-club' regions [van den Heuvel and Sporns, 2011]; regions that form the backbone of the connectome. The performance of the approach as a parcellation strategy is validated by comparing the between-subject reproducibility of the parcellations with those obtained from the K-means algorithm as well as from Infomap. We found that our model is up to par with Infomap on single-subject parcellation, while surpassing the results from K-means.

Our model does not result in a single point-estimate of a parcellation, but instead describes the posterior distribution of supported parcellations. This reveals that some regions are difficult to reliably assign to a particular cluster. Moreover, some regions display a gradual change in

cluster-assignment probability. This indicates that the boundaries between clusters are not that clear-cut, so that caution is advised when parcellating with deterministic approaches.

Chapter 4: Functional connectivity constrained by structure

While the previous two chapters have been concerned with structural connectivity, in Chapter 4 we consider functional connectivity instead. We quantify functional connections using partial correlations, which have been shown to be successful in recovering functional connectivity from fMRI BOLD signal time series [Smith et al., 2011; Schmittmann et al., 2015]. Unfortunately, partial correlations are difficult to estimate when the number of nodes in the network surpasses the number of observations. And even when there are more samples available than nodes in the network, the estimation error of the partial correlations may still be large. Several regularization techniques have been proposed to counteract these difficulties [Friedman et al., 2008], but these tend to be biased towards weaker connection strengths. Also, these approaches all compute a single most likely estimate, which fails to capture the associated uncertainty of the estimation.

In Chapter 4, we again propose a Bayesian alternative that a) provides the means to compute partial correlation without shrinkage bias and b) describes the posterior distribution of functional connectivity rather than a point estimate. The generative model uses the G-Wishart distribution as a prior distribution on partial correlations [Atay-Kayis and Massam, 2005; Dobra and Lenkoski, 2011]. Here, the graph G describes the pairs of regions for which functional connectivity is to be estimated. For connections not in G, the partial correlation is defined to be zero. As a result, the number of free parameters is reduced. In the context of brain connectivity, structural connectivity serves as an obvious candidate for G. We show that, using a simple estimate for G determined by probabilistic streamline counts, we are able to efficiently compute the posterior distribution over partial correlations.

Using simulated data, we evaluated and discussed the strengths and weaknesses of the proposed approach. We found that the model is highly effective in recovering functional connectivity, provided the supplied structural constraint G is a good representation of the independencies in the functional data. When applied to empirical data, we found functional connectivity that was confirmed by literature, although it must be acknowledged that our structural constraint G used here was suboptimal. On hold-out data we found that the likelihood of our connectivity estimates was higher than the likelihood obtained by cross-validated graphical lasso, but that the graphical lasso tuned to a network density similar to our estimates performed even better.

Chapter 5: Unconstrained functional connectivity using Gaussian graphical models

The previous chapter showed that the G-Wishart distribution provides a useful way of learning partial correlations for functional connectivity, but that the success of this approach is highly dependent on the quality of the structural constraint G. This inspired the work presented in Chapter 5, in which G is learned from the data simultaneously. In other words, instead of ap-

proximately inferring the distribution of $P(\mathbf{R} \mid G, \mathbf{X})$, i.e. the partial correlations $\mathbf{R}$ given a constraint G and data $\mathbf{X}$, we were now interested in the joint distribution $P(\mathbf{R}, G \mid \mathbf{X})$, i.e. both functional connectivity as well as its constraint, given data.

While straightforward to define, constructing a MCMC sampler to approximate the joint distribution is a very challenging endeavour. Central to this difficulty is that the G-Wishart distribution is *doubly intractable*. That is, the partition function of this distribution is dependent on one of its parameters. When the MCMC algorithm proposes a new sample by updating a connection in G, the partition function of the proposed sample does not cancel out the partition function of the current sample, as is typically the case in Metropolis-Hastings MCMC. The way to deal with this problem is to introduce a reversible-jump step in the MCMC sampler [Green, 1995], which compensates for the differences in normalization. Unfortunately, reversible-jump sampling results in a low acceptance rate, which implies that the algorithm takes long to converge. In Chapter 5, we focussed on integrating several improvements in state-of-the-art sampling algorithms to create a reasonably efficient MCMC procedure for estimation of the joint distribution. The algorithm was validated using a simulation and subsequently applied to estimate functional connectivity between subcortical areas. Although the example at the end of this chapter is geared towards functional brain connectivity, it should be noted that the proposed methodology is suited for wide applications in the field of Gaussian graphical models.

Chapter 6: Multi-modal data fusion for joint estimation of structural and functional connectivity

Most brain imaging studies are aimed at analyzing one particular data modality at a time. Yet when multiple modalities are combined, the result may be more than the sum of its parts [Sui et al., 2011; Rykhlevskaia et al., 2008]. In Chapter 6 we combine the models from previous chapters in order to inform both structural as well as functional connectivity using both probabilistic streamline data and BOLD time series. The crucial step in integrating the two modalities (and their respective generative models) is to recognize that the conditional (in)dependence graph G that is used in the G-Wishart context may be equated with anatomical wiring under some modest assumptions. That is, the conditional independencies in functional data give insight in structural connectivity, because direct functional coupling must be mediated by a structural connection. At the same time, evidence for a structural connection makes conditional independence between the corresponding regions unlikely.

The chapter is organized into two parts. In the first part, we compare our results with those obtained using the graphical lasso [Friedman et al., 2008], using simulated data of fMRI BOLD response time series [Smith et al., 2011], with known ground truth. The results of this comparison showed that our approach typically performs as least as well as the graphical lasso, while at times performing substantially better. By zooming in, we found that this is due to the fact that partial correlations and conditional (in)dependencies are now considered as separate parameters. Each element of the posterior distribution is a tuple consisting of a conditional independence graph and the corresponding partial correlations. There is no shrinkage effect in these tuples, but instead each different model has a different sparsity structure. The result is for each

connection a probability of conditional dependence, as well as the distribution of supported partial correlations.

In the second part of the chapter, we estimated connectivity between fourteen subcortical structures, using either the model from Chapter 5, the newly introduced data fusion approach, or a model in which we showed how prior information can be used to guide the connectivity estimation procedure. We observed that the data fusion approach decreased the variance of partial correlation estimates, indicating that data fusion results in more certain estimates. Furthermore, we found that when explicit prior information is available (for example when a particular connection is known to be absent, e.g. through a lesion), this can be used to inform not only the connection in question, but its effects also propagate to the estimates of other connections.

7.2 Limitations and suggestions for future research

There is no such thing as a free lunch. While I have shown how Bayesian methods can be applied to quantify uncertainty in connectivity estimates, this typically comes at the cost of computationally expensive MCMC procedures for approximate inference. Below, I discuss these and other limitations to my work. In addition, I provide suggestions as to how they may be overcome and how the current models may be extended.

Efficient approximate inference

The posterior distribution of the parameters is rarely available in closed form. As a consequence, methods such as MCMC are necessary to approximate this quantity instead. Although MCMC is arguably one of the most important algorithms invented in the previous century [Cipra, 2000], it is also notoriously slow in high-dimensional settings [Gilks et al., 1996]. Because of this drawback, my methods have been applied to relatively small networks, ranging from 14 up to 162 nodes. As the field of connectomics is progressing towards analysis of large-scale networks [Van Essen and Ugurbil, 2012; Schmittmann et al., 2015], it is crucial to develop more efficient sampling techniques that enable approximate computation of the posterior in reasonable time. An example of a promising approach is no-U-turn sampling (NUTS) [Hoffman and Gelman, 2014]. The NUTS approach simulates Hamiltonian dynamics, enabling much bigger jumps between successive states of the Markov chain. As a result, mixing is improved greatly compared to Metropolis-Hastings sampling, which was mostly used throughout this thesis. A downside of the NUTS approach is that it requires evaluation of the gradient of the likelihood. As a result, the method is unsuitable for use with discrete parameters (such as a binary connectivity matrix, or cluster assignments), so that appropriate workarounds must be constructed. Other strategies that might be particularly useful in the context of high-dimensional models include sequential sampling, in which the data is processed in batches instead of all at once [Korattikara et al., 2015], or trading some of the accuracy of MCMC methods for the speed of variational inference [Salimans et al., 2015].

The work presented in Chapters 5 and 6 may be improved by recent developments in the field of Gaussian graphical models. For example, Uhler et al. [2014] describe how the ratio of posterior probability between two graphical models can be computed exactly, without using a reversible-jump sampling scheme as described in Chapter 5. This results in a speed increase for two reasons: first there is the benefit of less computational demand per sample, and second the acceptance rate of the sampler improves by not having to perform a reversible-jump, which results in better mixing. This enables application of the method to networks with a much larger number of nodes.

Alternative generative models

The model for structural connectivity that was proposed in Chapter 2 assumed, for the sake of simplicity, that structural connectivity could be represented as a binary network. However, actual white-matter bundles have many varying microstructural properties [Assaf et al., 2013] and are therefore better captured using a weighted network instead [Jbabdi et al., 2015]. While it is straightforward to rephrase the forward model so it captures weighted connections, it is an open problem in dMRI-based connectivity studies what tractography-related quantity properly reflects connectivity weights [Jones, 2010; Jones et al., 2013; O'Donnell and Pasternak, 2014].

A data fusion approach as in Chapter 6 may be used to integrate several different sources of data on structural connections in order to either inform or validate weighted estimates [Jbabdi and Johansen-Berg, 2011]. Further progression in Bayesian connectivity models for data fusion can be achieved by incorporating copula distributions [Nelsen, 1999]. A copula is a multivariate probability distribution with the property that its marginals are uniform over the $[0, 1]$ interval. Copulas can be used to determine conditional (in)dependence between non-Gaussian variables, which is particularly useful when working with for example ordinal or nominal variables [Dobra and Lenkoski, 2011; Mohammadi and Wit, 2015].

Chapter 6 describes an intuitive way of specifying a constraining prior distribution for functional connectivity. While posed here as a proof-of-concept, a natural extension of this idea is to gather data from literature, and combine these into an anatomical prior that can be used as a constraint for functional connectivity. For example, the geometrical constraints that brain connectivity is subject to [Roberts et al., 2015], may be captured using an appropriate prior. Other information may be obtained from tracer studies in nonhuman primates [Van Essen and Ugurbil, 2012]. As several studies have shown that there is a high degree of correspondence between for example macaque and human structural connectivity [Miranda-Dominguez et al., 2014; Jbabdi et al., 2013; Rushworth et al., 2009; Goulas et al., 2014], this type of data provides unique insights that are otherwise impossible to obtain for human subjects.

Combining data from different species results in another challenge for which Bayesian models are particularly suited: While on the one hand we would like to move towards analysis of larger networks, on the other hand tracer studies are often restricted to coarse-grained parcellations containing a few dozen regions [Markov et al., 2014; Felleman and Van Essen, 1991]. Using a Bayesian hierarchical model, we can specify how the coarse information from primate tracer studies informs the finer-grained human connectome estimates. An example of how this may

be accomplished is by learning a stochastic block model as in Chapter 3 at the level of tracer connectivity, which in turns forms a prior distribution for structural connectivity at a more fine-grained scale.

Many functional connectivity studies consider a brain network to be static during the time data is acquired. However, recent studies have pointed out that dynamics of (functional) connectivity may be more insightful than static snapshots. For example, one could specifically take into account the temporal dependence between successive observations [Tank et al., 2015], or one could identify through which transitions a functional network moves during the performance of a particular task [Cribben et al., 2013; Hutchison et al., 2013; Calhoun et al., 2014; Kopell et al., 2014]. This can be incorporated in Bayesian connectomics in several ways. For example, one can construct a temporal mixture model of different G-Wishart distributions, each corresponding to a particular cognitive state. As the dynamics of structural connectivity can be ignored for the duration of a few cognitive tasks, the mixture can share a single structural constraint G. By supplying task-positive and task-negative data, the model can infer the different functional networks that are active during the phases of the experiment. This in turn can be used to generate hypotheses about how changes in cognitive state lead to different functional networks (see [Kucyi and Davis, 2014; Nielsen et al., 2015] for examples of a similar idea).

Further usage of the Bayesian framework

While the methods presented here do not rely on parameters such as a threshold that determines whether two regions are considered connected, or the number of clusters in a parcellation, my methods do depend on a number of hyperparameters as part of the priors in the described models. These hyperparameters were usually set to result in weakly informative priors that allowed the likelihood to dominate the posterior distribution, or to values that generated simulated data similar to the data observed empirically. However, these parameters could also be learned using empirical Bayes [Casella, 1985], replaced with a hierarchical model with uninformative priors at the highest level, could be optimized according to hold-out data or could be based on results from other studies. Although this was considered out of the scope of the thesis, a fully Bayesian treatment of these parameters should be addressed in future work.

The Bayesian framework has a further benefit that has not received attention in this thesis: Bayesian model selection. By computing the *Bayes factor* between two competing models one can quantify to which degree either model is favored over the other, considering the data. This forms the Bayesian alternative for hypothesis testing [Gelman et al., 2013]. Note that this is fundamentally different from classical hypothesis testing, where only evidence *against* the null hypothesis is considered (but never in favor of it) [Press, 2002]. This idea has been applied to determine effective (i.e. directed) connectivity in the dynamic causal modeling (DCM) framework [Friston et al., 2003; Stephan and Roebroeck, 2012], in particular for fMRI data. However, computing the Bayes factor between two models in requires integration over all parameters of the different models, which in general brings several computational challenges that must be overcome before these analyses become feasible. Once the computational hurdles of Bayesian

model selection have been addressed, it becomes possible to construct other generative models for connectivity and compare these in a fully Bayesian setting.

A caveat of Bayesian model selection is that our models are in practice only approximations of the true data generating process. That is, no model captures all details of the underlying physical processes and some modeling assumptions may be evidently wrong (e.g. assuming certain conditional independencies between variables to facilitate efficient computation). The true mechanism that describes the observed data is then not even in the support of the prior — we are simply not considering it a possibility. In spite of this Bayesian model selection can still be used, as it can be applied not only to find the (unattainable) true model, but also to find the model that best describes the data or the one that provides the best prediction of future observations [Liseo, 2000]. This is captured in the adage "all models are wrong, but some are useful" [Box and Draper, 1987], that suggests a pragmatic approach: we might be unable to find the true data generating process, but we can still find the most adequate model among the possibilities we consider. In order to construct robust models, i.e. models that are not (too) sensitive to small inconsistencies between the true mechanism and the model [Huber and Ronchetti, 2009], one can consider classes of priors and likelihoods instead of particular ones [Insua and Ruggeri, 2000]. There is still an active discussion about robustness of Bayesian analysis in general, for which we refer the reader to [Owhadi et al., 2015] and [Gelman et al., 2013].

Alternatives to Bayes

Besides Markov chain Monte Carlo, other methods to characterize uncertainty exist. For example, bootstrapping techniques [Efron and Tibshirani, 1993] randomly resample the data with replacement and construct estimates of the relevant parameters for each different set of samples, which can be thought of as a repeated maximum likelihood method. This approach has indeed been applied to brain connectivity studies (see for example [Damoiseaux et al., 2006; Raj and Chen, 2011; Doucet et al., 2011]). The straightforward implementation of bootstrapping makes it well-suited for use with regularization approaches such as the graphical lasso, so as to obtain a measure of uncertainty for this method. However, bootstrapping assumes that the original collected data samples are representative of the underlying distribution, and does not provide an easy way to integrate prior beliefs or hierarchical dependencies.

The usage of either MCMC or bootstrapping boils down to preference for the Bayesian or frequentist statistical framework, respectively, and this remains a matter of lively debate (see [Kass, 2011] for a noble attempt at discussing the issue without choosing sides) which we will not attempt to settle here.

7.3 OUTLOOK ON CONNECTOMICS

Throughout the previous paragraphs, the emphasis has been on adequate ways of *obtaining* connectivity estimates. But identifying which connections are present, and correlating them (or summary statistics of them) with phenotypical measures, should not be our end goal. In-

stead, connectomics is a tool that the neuroscientist may utilize to gain insight into how the brain is organized and, ultimately, produces the cognition and behavior we observe in daily life. However, if we are to properly put the claim 'we are our connectomes' [Seung, 2012] to the test and investigate how connectivity drives the workings of the brain, a lot of work remains to be done. Below I describe two — in my opinion — key ingredients that are needed to accomplish this: further analysis of how functional dynamics emerge from structural connectivity, and integration of different scales at which connectivity is studied.

The interplay between structure and function

Although there is ample evidence supporting the tight coupling between structure and function in brain connectivity [Greicius et al., 2009; Messé et al., 2015a,b,c; Shen et al., 2015; Hermundstad et al., 2013], as of yet we understand surprisingly little about how anatomy actually shapes the functional repertoire of the brain [Gu et al., 2015; Park and Friston, 2013]. While it is a crucial first step to observe that structural and functional connectivity show correspondence in network organization, the next step — showing how structural connectivity induces functional dynamics — will be even more revealing [Jirsa et al., 2010]. A number of studies have boldly accepted the challenge to move beyond a descriptive account of the correspondence between structure and function. For example, Gu et al. [2015] as well as Mišić et al. [2015] show how perturbations at particular nodes in the structural connectome propagate through the network and move the brain among different cognitive states. Other work suggests that functional dynamics can be explained using models of near-criticality, a regime of optimal balance between neuronal coupling and variability [Marinazzo et al., 2014; Hesse and Gross]. It has been shown that near-criticality leads to optimal neuronal communication, computation and information storage [Beggs and Timme]. In addition, Haimovici et al. [2013] show that the connectome provided by Hagmann et al. [2008] together with a simple rule for functional dynamics, is able to predict emerging resting-state connectivity, providing the network is at criticality.

Although the existence of a white-matter connection (i.e. an edge in a structural network) is a prerequisite for functional dependence [Honey et al., 2009], additional factors are relevant for functional connectivity. For example, neuronal activity is greatly affected at the molecular level by local neuromodulators that influence the dynamics of neuronal populations [Bargmann, 2012; Bargmann and Marder, 2013; Sporns, 2013]. These effects may be taken into account by constructing and comparing more biologically motivated generative mechanisms for connectivity [Betzel et al., 2016; Woolrich and Stephan, 2013].

Furthermore, realistic models describing the interaction between structure and function may incorporate anatomical constraints and evolutionary pressure [Budd and Kisvárday, 2012; Raj and Chen, 2011; Goulas et al., 2014; Vértes et al., 2012; Van Essen, 2013]. A noteworthy study that illustrates these ideas was conducted by Wedeen et al. [2012a], who identified that fibers estimated through tractography can be 'unfolded' to reveal a three-dimensional grid whose major directions followed the developmental axis. By combining these results with the tension-based morphogenesis theory of Van Essen [1997] (which implies that cortical folding is driven by axonal tension pulling different patches of cortex together), we obtain a glance at the orga-

nizational principles at the macroscale. Although the findings by Wedeen et al. [2012a] are not undisputed (see the response [Catani et al., 2012a] and its subsequent rebuttal [Wedeen et al., 2012b]), it suggests a rich source for hypotheses that may be further explored. For instance, studies of prenatal brain imaging may be used to provide insight in the relation between anatomy and connectivity during development [van den Heuvel et al., 2015; Jakab et al., 2014] or cross-species comparisons may be used to investigate evolutionary development and speciation [Reid et al., 2016].

Integrating macroscale and microscale connectivity

Presently, there is a divide between macroscale [Craddock et al., 2013] and microscale [Helmstaedter, 2013] studies of connectivity. At the macroscale, nodes contain many thousands of neurons. Despite networks at this scale having shown a wide range of applications, the explanatory power of these approaches about neuronal behavior is limited. At the other end of the spectrum, techniques such as serial block-face scanning electron and two-photon calcium imaging microscopy allow characterization of structural [Yatsenko et al., 2015] and functional connectivity [Stosiek et al., 2003] at the cellular level. It is at this level that notions about connectivity can be incorporated into models of neuronal coding [Harris and Mrsic-Flogel, 2013; Clopath et al., 2010]. While this provides a much more fine-grained description of neuronal connectivity, here we risk not seeing the forest for the trees; because of the sheer number of nodes the interpretability at this level becomes difficult and this is precisely why macroscale connectomics are so popular. So how do we obtain the best of both worlds and bridge this gap between macroscale and microscale connectivity? This forms one of the major challenges for connectomics in the years to come [DeFelipe, 2010; Johansen-Berg, 2013; Assaf et al., 2013]. It requires a simultaneous reductionist and emergentist approach [Kelso et al., 2013], combining structure and function at different scales, in order to learn how, for example, neuronal activity and Hebbian plasticity give rise to macroscale phenomena.

Improvements in our understanding of the structure-function relation, as well as in the interplay between different scales of analysis, will provide immediate benefit in application. In clinical diagnosis for example, it allows us to interpret changes in graph-theoretical measures in terms of their biophysical substrate [Johansen-Berg, 2013], providing a much more informative picture of what ails a patient.

7.4 CONCLUDING REMARKS

A decade has passed since the term 'connectome' was introduced [Sporns et al., 2005; Hagmann, 2005]. This network-based view of the brain has resulted in a surge of scientific output aiming to investigate brain networks and to relate them to cognitive performance or impairment [Fornito and Bullmore, 2012, 2014]. It can be expected that in the coming years the amount and quality of available neuroimaging data will increase rapidly, for example through programs such as the Human Connectome Project [Van Essen et al., 2012], the BRAIN initiative [Insel et al., 2013]

or the Blue Brain project [Markram et al., 2011]. Yet despite these vital efforts, there exists no analysis based on empirical data that gives us an absolute truth: statistical uncertainty will always remain. Throughout this thesis, I have advocated that this uncertainty should be made an explicit part of our studies of the connectome and I have demonstrated that the Bayesian framework provides the tools to accomplish this. Hopefully, my work contributes to a more nuanced understanding of brain connectivity.

A

APPENDIX TO CHAPTER 2

A.1 DERIVATION OF THE MULTI-SUBJECT PRIOR

Here we provide the derivation of the multi-subject prior based on the maximum likelihood estimates for previously seen subjects, as described in Section 2.1. The prior on $\mathbf{G}' \equiv \mathbf{G}^{(M+1)}$ is given by

$$\begin{aligned}
P\left(\mathbf{G}' \mid \mathcal{S}, \zeta\right) &= \sum_{\bar{\mathbf{G}}} P\left(\mathbf{G}' \mid \bar{\mathbf{G}}\right) P\left(\bar{\mathbf{G}} \mid \mathcal{S}, \zeta\right) \\
&\propto \sum_{\bar{\mathbf{G}}} P\left(\mathbf{G}' \mid \bar{\mathbf{G}}\right) P\left(\bar{\mathbf{G}} \mid \theta\right) \\
&\quad \times \sum_{\mathcal{G}} P\left(\mathcal{S} \mid \mathcal{G}, \delta_1, \delta_0\right) P\left(\mathcal{G} \mid \bar{\mathbf{G}}\right) \\
&\propto \sum_{\bar{\mathbf{G}}} P\left(\mathbf{G}' \mid \bar{\mathbf{G}}\right) P\left(\bar{\mathbf{G}} \mid \theta\right) \\
&\quad \times \prod_{m=1}^{M} \sum_{\mathbf{G}^{(m)}} P\left(\mathbf{S}^{(m)} \mid \mathbf{G}^{(m)}, \delta_1, \delta_0\right) \\
&\quad \times P\left(\mathbf{G}^{(m)} \mid \bar{\mathbf{G}}\right) \quad .
\end{aligned}$$

We approximate this quantity by assuming that the main contribution in the sum over $\mathbf{G}^{(m)}$ is due to the ML solution

$$\hat{\mathbf{G}}^{(m)} = \underset{\mathbf{G}^{(m)}}{\operatorname{argmax}}\, P\left(\mathbf{S}^{(m)} \mid \mathbf{G}^{(m)}, \delta_1, \delta_0\right) \quad .$$

Following the Erdős-Rényi model with $\theta = 0.5$ for $P\left(\bar{\mathbf{G}} \mid \theta\right)$ gives a flat prior on simple graphs. Up to irrelevant constants and keeping in mind that $\hat{\mathbf{G}}$ depends on $\mathbf{S}^{(m)}$, the prior is rewritten as

$$P\left(\mathbf{G}' \mid \mathcal{S}, \zeta\right) \approx \sum_{\bar{\mathbf{G}}} P\left(\mathbf{G}' \mid \bar{\mathbf{G}}\right) \prod_{m=1}^{M} P\left(\hat{\mathbf{G}}^{(m)} \mid \bar{\mathbf{G}}\right) \ ,$$

with $\hat{\mathcal{G}} = \{\hat{\mathbf{G}}^{(1)}, \ldots, \hat{\mathbf{G}}^{(M)}\}$ the different ML solutions. We assume that the prior factorizes into

$$P\left(\mathbf{G}^{(M+1)} \mid \mathcal{S}, \zeta\right) = \prod_{i<j} P\left(g_{ij}^{(M+1)} \mid \mathcal{M}, \zeta\right) \ . \tag{A.1}$$

Next, we define the probability that $g_{ij}^{(m)}$, $s \in (1, \ldots, M+1)$, inherits the connectivity from the parent network $\bar{g}_{ij}$ by

$$P\left(g_{ij}^{(m)} = 1 \mid \bar{g}_{ij} = 1\right) = P\left(g_{ij}^{(m)} = 0 \mid \bar{g}_{ij} = 0\right) \equiv q_{ij} \ ,$$

with q_{ij} close to 1. That is, each $g_{ij}^{(m)}$ is a copy of $\bar{g}_{ij}$ with unknown probability q_{ij}. The copying probabilities are assumed to be independent and have a flat prior. Estimating the prior probability for each edge is then nothing but an instance of Laplace's rule of succession. This says that, if we repeat an experiment that we know can result in a success (presence of an edge) or failure (absence of an edge) m times independently, and get $\sum_{m=1}^{M} \hat{g}_{ij}^{(m)}$ successes, then our best estimate of the probability that the next repetition $g_{ij}^{(M+1)}$ will be a success is:

$$P(g_{ij}^{(M+1)} = 1 \mid \hat{g}_{ij}^{(1)}, \ldots, \hat{g}_{ij}^{(M)}) = \frac{\sum_{m=1}^{M} \hat{g}_{ij}^{(m)} + 1}{M+2} \equiv \theta_{ij} \ .$$

Plugging this into (A.1), we obtain the prior

$$P\left(\mathbf{G}^{(M+1)} \mid \mathcal{S}, \zeta\right) = \prod_{i<j} \theta_{ij}^{g_{ij}^{(M+1)}} (1-\theta_{ij})^{(1-g_{ij}^{(M+1)})} \ .$$

A.2 MCMC SAMPLING

We derive here the acceptance rate γ of a sample $\mathbf{G}'$ in the sampling chain as a function of one edge flip in $\mathbf{G}$ (see Section 2.1). Note that each of the $2^{p(p-1)/2}$ possible networks $\mathbf{G}$ has a probability greater than zero of being constructed, which guarantees that the Markov chain is irreducible. The log acceptance rate of a suggested sample can be calculated as $\log \gamma = \Delta L_{kl} + \Delta P_{kl}$, with ΔL_{kl} and ΔP_{kl} the change in log-likelihood and log-prior respectively, after flipping edge g_{kl}. The sampling approach requires that we can efficiently update both

the likelihood and the prior for new samples in the Markov chain. The log-likelihood is given by

$$L \equiv \sum_i \left[\log \frac{S_i!}{\prod_j s_{ij}!} + \log \frac{\Gamma\left(\sum_j b_{ij}\right)}{\Gamma\left(\sum_j (b_{ij} + s_{ij})\right)} + \sum_j \log \frac{\Gamma\left(b_{ij} + s_{ij}\right)}{\Gamma\left(b_{ij}\right)} \right] \tag{A.2}$$

with $b_{ij} \equiv g_{ij}\delta_1 + (1 - g_{ij})\delta_0$. The change in log-likelihood as a consequence of flipping an edge g_{kl} is defined as

$$\Delta L_{kl} = \log P\left(\boldsymbol{S} \mid \boldsymbol{G}', \delta_1, \delta_0\right) - \log P\left(\boldsymbol{S} \mid \boldsymbol{G}, \delta_1, \delta_0\right) \quad , \tag{A.3}$$

with the sole difference that $g'_{kl} = g'_{lk} = (1 - g_{kl})$. Plugging (A.2) into (A.3) yields

$$\begin{aligned} \Delta L_{kl} = &\log \left[\frac{\Gamma\left(b'_{kl} + s_{kl}\right)}{\Gamma\left(b_{kl} + s_{kl}\right)}\right] + \log \left[\frac{\Gamma\left(b'_{lk} + s_{lk}\right)}{\Gamma\left(b_{lk} + s_{lk}\right)}\right] \\ &+ \log \left[\frac{\Gamma\left(\sum_j b'_{kj}\right)}{\Gamma\left(\sum_j b_{kj}\right)}\right] + \log \left[\frac{\Gamma\left(\sum_j b'_{lj}\right)}{\Gamma\left(\sum_j b_{lj}\right)}\right] \\ &- \log \left[\frac{\Gamma\left(\sum_j (b'_{kj} + s_{kj})\right)}{\Gamma\left(\sum_j (b_{kj} + s_{kj})\right)}\right] \\ &- \log \left[\frac{\Gamma\left(\sum_j (b'_{lj} + s_{lj})\right)}{\Gamma\left(\sum_j (b_{lj} + s_{lj})\right)}\right] - 2 \log \left[\frac{\Gamma\left(b'_{kl}\right)}{\Gamma\left(b_{kl}\right)}\right] \quad . \end{aligned} \tag{A.4}$$

The change in the log-prior as a consequence of flipping g_{kl} to $1 - g_{kl}$ for the prior follows from its definition in (2.6)

$$\begin{aligned} \Delta P_{kl} &= \log P\left(\boldsymbol{G}' \mid \mathcal{S}, \zeta\right) - \log P\left(\boldsymbol{G} \mid \mathcal{S}, \zeta\right) \\ &= (4g_{kl} - 2)\left[\log\left(\frac{\theta_{kl}}{1 - \theta_{kl}}\right) + \log\left(\frac{\theta_{lk}}{1 - \theta_{lk}}\right)\right] \quad . \end{aligned}$$

Here the edge probability θ_{kl} is the same for all edges in the case of the Erdős-Rényi model and estimated separately per edge in case of the multi-subject prior.

B

APPENDIX TO CHAPTER 3

B.1 MCMC AND GIBBS SAMPLING

In this section we elaborate on the MCMC approximations that we use to infer connectivity $\mathbf{G}$ and clustering $\mathbf{Z}$. In the single-subject case, connectivity and clustering are characterized by the joint posterior distribution

$$P(\mathbf{G}, \mathbf{Z} \mid \mathbf{S}, \delta_1, \delta_0, \alpha, \beta, \xi) \propto P(\mathbf{S} \mid \mathbf{G}, \delta_1, \delta_0)\, P(\mathbf{G} \mid \mathbf{Z}, \alpha, \beta)\, P(\mathbf{Z} \mid \xi) \quad . \tag{B.1}$$

First we note that in the definition of $P(\mathbf{G} \mid \mathbf{Z}, \alpha, \beta)$, α and β are used to draw a probability ρ from a beta distribution, which is then used to draw a connection g_{ij} from a Bernoulli distribution. As the beta distribution is conjugate to the Bernoulli, ρ may be integrated out, i.e.

$$\begin{aligned} P(\mathbf{G} \mid \mathbf{Z}, \alpha, \beta) &= \int P(\mathbf{G} \mid \boldsymbol{\rho})\, P(\boldsymbol{\rho} \mid \alpha, \beta)\, d\boldsymbol{\rho} \\ &= \prod_{a \geqslant b} \frac{\text{Beta}(B_+(a,b), B_-(a,b))}{\text{Beta}(\boldsymbol{\beta}_+(a,b), \boldsymbol{\beta}_-(a,b))} \end{aligned} \tag{B.2}$$

with a and b clusters and $B_+(a,b) = M_+(a,b) + \boldsymbol{\beta}_+(a,b)$ and $B_-(a,b) = M_-(a,b) + \boldsymbol{\beta}_+(a,b)$, where $M_+(a,b) = (1 - \frac{1}{2}\delta_{ab})\boldsymbol{z}_a^T \mathbf{G} \boldsymbol{z}_b$ is the number of edges between regions in cluster a and regions in cluster b, $M_-(a,b) = (1 - \frac{1}{2}\delta_{ab})\boldsymbol{z}_a^T(\boldsymbol{e}\boldsymbol{e}^T - \mathbf{I}_K)\boldsymbol{z}_b - M_+(a,b)$ is the number of non-edges between regions in clusters a and b, $\boldsymbol{e}$ is a vector of ones of size K, $\mathbf{I}_K$ is the $K \times K$ identity matrix and finally the Beta function defined as $\text{Beta}(a,b) =$

$\int_0^1 x^{a-1}(1-x)^{b-1}dx,$. The forward model for $\mathbf{S}$ can be simplified in a similar way, as the Dirichlet distribution is the conjugate prior for the multinomial, i.e.

$$\begin{aligned}
&P(\mathbf{S}\,|\,\mathbf{G},\delta_1,\delta_0)\\
&=\int P(\mathbf{S}\,|\,\mathbf{G},\mathbf{Y})\,P(\mathbf{Y}\,|\,\delta_1,\delta_0)\,d\mathbf{Y}\\
&=\prod_i\left[\frac{\left(\sum_j S_{ij}\right)!}{\prod_j S_{ij}!}\frac{\Gamma(\sum_j b_{ij})}{\Gamma\left(\sum_j(b_{ij}+S_{ij})\right)}\prod_j\frac{\Gamma(b_{ij}+S_{ij})}{\Gamma(b_{ij})}\right]\,,
\end{aligned}\tag{B.3}$$

with $b_{ij}=\delta_1 g_{ij}+\delta_0(1-g_{ij})$.

The posterior distribution can be approximated using Gibbs sampling. Conveniently, the different variables can be updated in blocks, which allows us to iteratively update the conditionals

$$\begin{aligned}
&P(\mathbf{G}\,|\,\mathbf{Z},\mathbf{S},\delta_1,\delta_0,\alpha,\beta)\\
&\propto P(\mathbf{S}\,|\,\mathbf{G},\delta_1,\delta_0)\,P(\mathbf{G}\,|\,\mathbf{Z},\alpha,\beta)\\
&=\prod_i\left[\frac{\left(\sum_j S_{ij}\right)!}{\prod_j S_{ij}!}\frac{\Gamma(\sum_j b_{ij})}{\Gamma\left(\sum_j(b_{ij}+S_{ij})\right)}\prod_j\frac{\Gamma(b_{ij}+S_{ij})}{\Gamma(b_{ij})}\right]\\
&\quad\times\prod_{a\geqslant b}\frac{\mathrm{Beta}\,(B_+(a,b),B_-(a,b))}{\mathrm{Beta}\,(\boldsymbol{\beta}_+(a,b),\boldsymbol{\beta}_-(a,b))}\,,
\end{aligned}\tag{B.4}$$

and

$$\begin{aligned}
&P(\mathbf{Z}\,|\,\mathbf{G},\mathbf{S},\alpha,\beta,\xi)\\
&\propto P(\mathbf{G}\,|\,\mathbf{Z},\alpha,\beta)\,P(\mathbf{Z}\,|\,\xi)\\
&=\left[\prod_{a\geqslant b}\frac{\mathrm{Beta}\,(B_+(a,b),B_-(a,b))}{\mathrm{Beta}\,(\boldsymbol{\beta}_+(a,b),\boldsymbol{\beta}_-(a,b))}\right]\xi^k\frac{\Gamma(\xi)}{\Gamma(\xi+K)}\prod_a\Gamma(n_a)\,,
\end{aligned}\tag{B.5}$$

in which n_a represents the number of regions in cluster a. The iterative block sampling approach is intuitively straightforward; we first update the connectivity variables given the clustering, then we update the clustering variables given the connectivity. To converge to the desired distribution it is essential that in each iteration the relevant variable is updated according to the most recent estimate of all other variables. We developed a Metropolis sampler for the first conditional [Hinne et al., 2013] and similarly [Mørup et al., 2010] developed a Gibbs sampler for the inference of the IRM, which implements a split-merge sampler to improve mixing [Jain and Neal, 2007]. Both samplers are implemented in Matlab.

For the hierarchical setting in which one cluster assignment $\mathbf{Z}$ is inferred for a group of subjects, let again $\mathcal{G} = (\mathbf{G}^{(1)}, \ldots, \mathbf{G}^{(M)})$ and $\mathcal{S} = (\mathbf{S}^{(1)}, \ldots, \mathbf{S}^{(M)})$. The posterior distribution that we are now interested in is given by

$$P(\mathcal{G}, \mathbf{Z} \mid \mathcal{S}, \delta_1, \delta_0, \alpha, \beta, \xi) \propto P(\mathcal{S} \mid \mathcal{G}, \delta_1, \delta_0)\, P(\mathcal{G} \mid \mathbf{Z}, \alpha, \beta)\, P(\mathbf{Z} \mid \xi) \quad . \tag{B.6}$$

The different subjects are assumed to be conditionally independent. Accordingly, the two conditionals factorize over subjects and become

$$\begin{aligned}
&P(\mathcal{G} \mid \mathbf{Z}, \mathcal{S}, \delta_1, \delta_0, \alpha, \beta) \\
&\propto P(\mathcal{S} \mid \mathcal{G}, \delta_1, \delta_0)\, P(\mathcal{G} \mid \mathbf{Z}, \alpha, \beta) \\
&= \prod_m P\left(\mathbf{S}^{(m)} \mid \mathbf{G}^{(m)}, \delta_1, \delta_0\right) P\left(\mathbf{G}^{(m)} \mid \mathbf{Z}, \alpha, \beta\right) \\
&= \prod_m \prod_i \left[\frac{\left(\sum_j S_{ij}^{(m)}\right)!}{\prod_j S_{ij}^{(m)}!} \frac{\Gamma(\sum_j b_{ij}^{(m)})}{\Gamma\left(\sum_j (b_{ij}^{(m)} + S_{ij}^{(m)})\right)} \right. \\
&\quad \left. \times \prod_j \frac{\Gamma(b_{ij}^{(m)} + S_{ij}^{(m)})}{\Gamma(b_{ij}^{(m)})} \right] \\
&\quad \times \prod_{a \geqslant b} \frac{\mathrm{Beta}\left(B_+^{(m)}(a,b), B_-^{(m)}(a,b)\right)}{\mathrm{Beta}\left(\boldsymbol{\beta}_+(a,b), \boldsymbol{\beta}_-(a,b)\right)} \quad ,
\end{aligned} \tag{B.7}$$

with $b_{ij}^{(m)} = \delta_1 g_{ij}^{(m)} + \delta_0 (1 - g_{ij}^{(m)})$, and

$$\begin{aligned}
&P(\mathbf{Z} \mid \mathcal{G}, \mathcal{S}, \alpha, \beta, \xi) \\
&\propto P(\mathcal{G} \mid \mathbf{Z}, \alpha, \beta)\, P(\mathbf{Z} \mid \xi) \\
&= \left[\prod_m P\left(\mathbf{G}^{(m)} \mid \mathbf{Z}, \alpha, \beta\right) \right] P(\mathbf{Z} \mid \xi) \\
&= \left[\prod_m \prod_{a \geqslant b} \frac{\mathrm{Beta}\left(B_+^{(m)}(a,b), B_-^{(m)}(a,b)\right)}{\mathrm{Beta}\left(\boldsymbol{\beta}_+(a,b), \boldsymbol{\beta}_-(a,b)\right)} \right] \\
&\quad \times \xi^k \frac{\Gamma(\xi)}{\Gamma(\xi + n)} \prod_a \Gamma(n_a) \quad .
\end{aligned} \tag{B.8}$$

To sample from the joint distribution in the hierarchical case, one simply iterates over updating the connectivity for each subject given the group-clustering and updating of the clustering given the product of the connectivity for each subject.

When approximating these posterior distributions using MCMC, one needs to assess whether the sampling chains have converged to the true distribution. We analyzed convergence by calculating the potential scale reduction factor (PSRF) [Gelman and Rubin, 1992], which can be understood as a ratio between intra-chain variance and inter-chain variance. Heuristically, we assumed convergence for connectivity when for each edge g_{ij} the PSRF dropped below 1.1. Convergence for clustering was expressed by deriving for each sample of $\mathbf{Z}$ the cluster probability matrix $\mathbf{M} = \mathbf{Z}^T\mathbf{Z}$. Similar to the convergence of connectivity, we assumed that the samples were converged to the target distribution when the PSRF for each m_{ij} was smaller than 1.1. To be able to calculate the PSRF, 10 independent sampling chains were executed, each starting from different, randomized initializations, consisting of a draw from the Chinese restaurant process for $\mathbf{Z}$ and an Erdős-Rényi random graph for $\mathbf{G}$ with a density of 20%. We found that 2,000 Gibbs iterations were sufficient to reach convergence. Note that for each iteration every element g_{ij} and z_{ir} was updated once. For the experiments that involved maximum a posteriori estimates instead of probability distributions, both the samplers were altered to perform simulated annealing with an exponential decay cooling schedule, which enforces convergence towards the mode of the distribution.

The computational complexity of one iteration of the bIRM sampler is $O(mMK^2)$ with $m \leqslant p^2$ the number of edges. Sampling connectivity for one subject for one iteration has a time complexity of $O(p^2)$. The sIRM approach therefore scales as $O(Mp^2K^2)$, or, if we assume $K \ll p$, as is the case in the parcellations we find, $O(Mp^2)$ instead. For comparison, the Infomap algorithm has a time complexity of $O(m)$ and K-means runs in $O(p^2K)$. This makes the proposed approach the slowest of the three algorithms. Note however that it provides the posterior distributions of $\mathbf{G}$, $\mathbf{Z}$ and ρ instead of a point estimate of $\mathbf{Z}$, and that the sampling process includes the determination of (the distribution of) K.

B.2 PARAMETER SELECTION

A number of parameters potentially influence the parcellations we obtain. These are

1. ξ, the concentration parameter of the Chinese restaurant prior,
2. α and β that determine the expected probability of a connection between two clusters via a beta distribution and
3. the two different Dirichlet hyperparameters δ_1 and δ_0 that influence how many streamlines are observed due to noise.

Throughout our experiments we used $\xi = \log p$, cf. [Mørup et al., 2010]. We observed that our results where very robust against different settings of ξ. In fact, only when ξ was increased to result in an expected number of clusters that was orders of magnitudes larger than the number of nodes (which makes it an impossible parcellation), did this parameter seem to have an effect. For the parameters α and β, the choice of $\alpha = \beta = 1$ was determined by our experimental setup, in which we did not want to make any prior assumptions on the cluster connectivity behavior in our parcellations. Finally, the parameters δ_1 and δ_0 determine the relation

between observed streamlines and structural connectivity. The chosen setting of $\delta_1 = 1$ and $\delta_0 = 0.1$ results in an uninformative prior for connected pairs of nodes, while enforcing that high streamline counts are unlikely to be explained by non-connections. This intuition was previously validated, by comparing the estimates for structural connectivity with functional independencies [Hinne et al., 2013].

B.3 K-means and Infomap

The K-means implementation that comes with the Matlab (MATLAB 7.7, The MathWorks Inc., Natick, MA, USA) Statistics Toolbox was used with the cosine similarity as the (inverse) distance metric. The Infomap algorithm [Rosvall and Bergstrom, 2008] was implemented using code from the Graph Cluster toolkit [van Laarhoven and Marchiori, 2013]. Both procedures were repeated 300 times with random initializations, of which the best was used as the final result.

C

APPENDIX TO CHAPTER 5

C.1 MARKOV CHAIN MONTE CARLO SAMPLING

We approximate the posterior distributions using Metropolis-Hastings Markov chain Monte Carlo. The procedure requires that we compute the acceptance ratio γ between the current state $(\mathrm{G}, \mathbf{K})$ and the proposed state $(\tilde{\mathrm{G}}, \tilde{\mathbf{K}})$. It is defined as the ratio of the posterior probabilities of the two states, i.e.

$$\gamma = \frac{\mathrm{P}(\tilde{\mathrm{G}}, \tilde{\mathbf{K}} \mid \mathbf{X})}{\mathrm{P}(\mathrm{G}, \mathbf{K} \mid \mathbf{X})} = \frac{\mathrm{P}(\mathbf{X} \mid \tilde{\mathrm{G}}, \tilde{\mathbf{K}})}{\mathrm{P}(\mathbf{X} \mid \mathrm{G}, \mathbf{K})} \frac{\mathrm{P}(\tilde{\mathrm{G}})}{\mathrm{P}(\mathrm{G})} \quad . \tag{C.1}$$

In general, the proposed state is accepted as a new sample with probability $\min(1, \gamma)$, otherwise the current state is stored instead. Until recently, this approach has proven impractical for this specific problem, as no efficient way was available to generate the proposed $\tilde{\mathbf{K}}$. Instead, a block Gibbs sampler that updates $\mathbf{K}$ according to either the edges of G [Wang and Li, 2012] or its clique decomposition [Piccioni, 2000] was used, but this is a computationally demanding solution. Significant improvement was obtained by the introduction of a direct sampler for the G-Wishart distribution [Lenkoski, 2013]. Together with a reversible-jump setup [Green, 1995], this resulted in a much faster way of computing (C.1), as described in [Lenkoski, 2013]. Additional efficiency was gained in recent work, by exploiting the analytical properties of the conditional Bayes factor $\mathrm{P}(\mathbf{X} \mid \mathrm{G}, \mathbf{K})$ to compute γ more efficiently [Hinne et al., 2014]. The corresponding algorithm is used in this paper.

When the data from probabilistic tractography are added, the acceptance ratio becomes

$$\gamma' = \frac{\mathrm{P}(\mathbf{X} \mid \tilde{\mathrm{G}}, \tilde{\mathbf{K}})}{\mathrm{P}(\mathbf{X} \mid \mathrm{G}, \mathbf{K})} \frac{\mathrm{P}(\mathbf{S} \mid \tilde{\mathrm{G}})}{\mathrm{P}(\mathbf{S} \mid \mathrm{G})} \frac{\mathrm{P}(\tilde{\mathrm{G}})}{\mathrm{P}(\mathrm{G})} \quad . \tag{C.2}$$

As $\mathbf{X}$ and $\mathbf{S}$ are conditionally independent given G, computing γ' consists of multiplying γ with the acceptance ratio of the structural model, which is described in detail in [Hinne et al., 2013].

C.2 CONVERGENCE

For each distribution, three parallel chains were executed until convergence. Convergence was assessed by computing the potential scale reduction factor (PSRF) [Gelman and Rubin, 1992], which can be understood as a ratio between intra-chain variance and inter-chain variance. Heuristically, we assumed convergence for connectivity when for each edge (i, j) and each element k_{ij} the PSRF dropped below 1.1. For the first model, convergence was attained within 20 000 samples. The multimodal approach proved substantially slower to converge, requiring between 50 000 and 200 000 samples. Four subjects contained edges and elements in the precision matrix that did not fully converge within reasonable time. However, for these subjects the chains did reach convergence in terms of the network density of G and the matrix determinant of $\mathbf{K}$, which motivated us to treat them the same as the other subjects. The last model with a more specific prior decreases the number of free parameters, which makes the convergence substantially faster; convergence was attained within 10 000 samples.

C.3 PRIOR INFLUENCE

Throughout our experiments we use a Bernoulli distribution with parameter θ to define the probability of an edge. In the prior, we consider all edges to be conditionally independent, so we have

$$P(G \mid \Theta) = \prod_{i<j} \theta_{ij}^{g_{ij}} (1 - \theta_{ij})^{1 - g_{ij}} \quad , \tag{C.3}$$

with $\Theta = (\theta_{ij})_{i<j}$. Conform [Hinne et al., 2013], we adopt $\forall_{i,j}[\theta_{ij} = 0.5]$. Note that $\mathbb{E}[d \mid \theta] = \theta$, with $d(G)$ the density of G, i.e. the fraction of present edges.

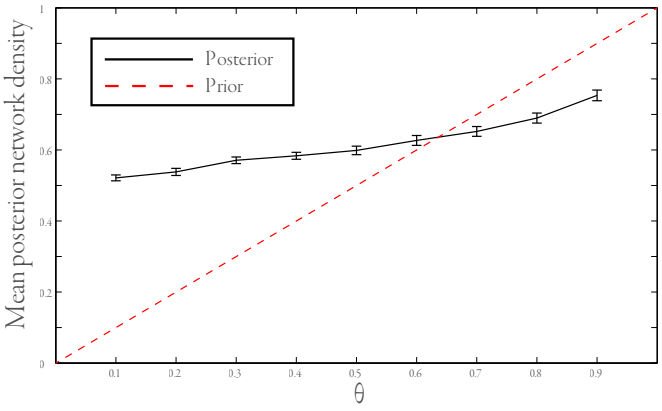


FIGURE C.1: The mean posterior network density as a function of the Bernoulli parameter θ. The prior network density is fully determined by θ and is indicated with the red dotted line. Error bars indicate one standard deviation of the posterior network density.

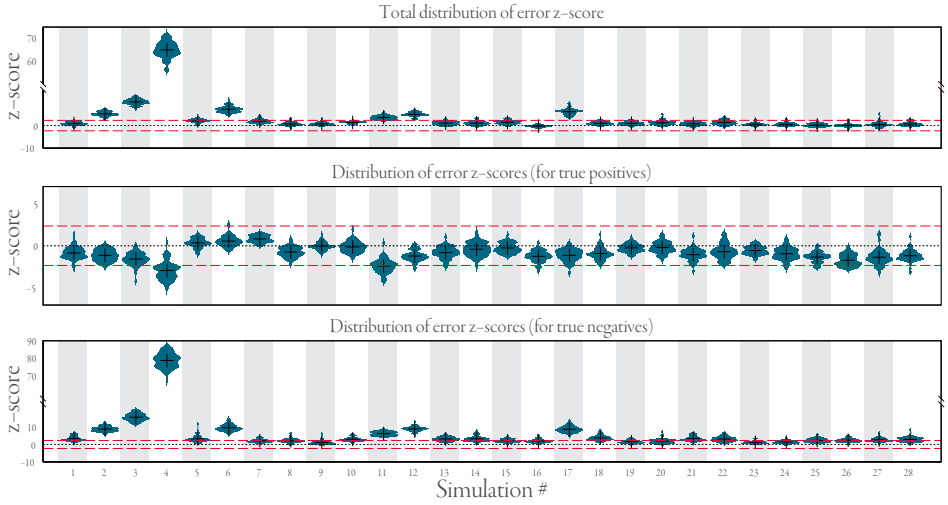


FIGURE C.2: The histograms for each of the 28 different simulations. Positive error z-scores indicate that the maximum likelihood estimate was less effective in recovering the ground truth than the Gaussian graphical model, while the reverse is true for negative error z-scores. The red dashed lines indicate the interval outside of which the difference in performance is significant ($p < 0.01$, z-test).

To analyze the influence of the prior, we repeated the connectivity estimation procedure for one subject, by collecting 1 000 samples with $\forall_{i,j}[\theta_{ij} = \{0.1, \ldots, 0.9\}$. The extreme values $\theta = \{0.0, 1.0\}$ have been excluded as in these cases the prior would fully determine the posterior outcome (see also the discussion of the informed prior in the main text). As shown in Figure C.1, the mean posterior network density is in the range $[0.52, 0.75]$, a much smaller range than the prior promotes. This indicates that although the effect of the prior cannot be ignored entirely, the posterior and prior network densities are far apart. This allows us to conclude that the posterior network density is dominated by the likelihood.

C.4 SUPPLEMENTARY FIGURES

Figure C.2 shows the results of the simulations where the BGGM approach is compared with the maximum likelihood estimates. Similar to the comparison with the graphical lasso results, the BGGM has in general a much smaller error in recovery of the true network structure than the MLE.

Figure C.3 shows the group-averaged posterior expectations for the functional connectivity model, the data fusion approach and the informed prior. The figure also shows the standard deviations across subjects. This indicates, for example, that the data fusion approach leads to a much lower variance for both the conditional dependencies, which are directly informed by probabilistic streamlines, but also for partial correlations.

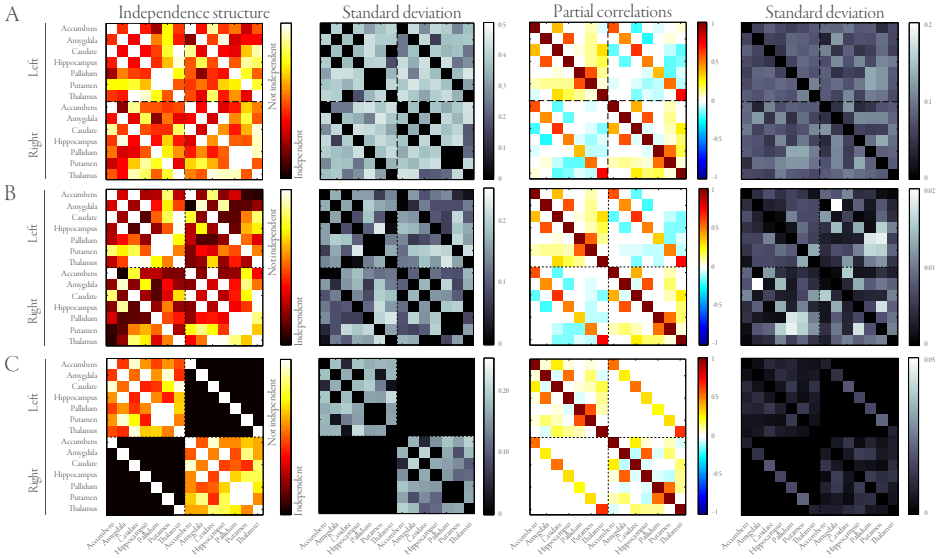


FIGURE C.3: Group-averaged functional connectivity, as average posterior expectations of the conditional independence graph (left) and partial correlations (right), as well as the standard deviations across twenty subjects. A. Functional connectivity model. B. Data fusion model. C. Informed prior model.

BIBLIOGRAPHY

Aicher C., Jacobs A. Z., and Clauset A. Learning latent block structure in weighted networks. *Journal of Complex Networks*, 3(2):221–248, 2015.

Airoldi E. M., Blei D. M., Fienberg S. E., and Xing E. P. Mixed membership stochastic blockmodels. In Koller D., Schuurmans D., Bengio Y., and Bottou L., editors, *Advances in Neural Information Processing Systems 21*, pages 33–40. Curran Associates, Inc., 2009.

Akil H., Martone M. E., and van Essen D. C. Challenges and opportunities in mining neuroscience data. *Science*, 331 (6018):708–712, 2011.

Aldous D. J. Exchangeability and related topics. In Hennequin P., editor, *École d'Été de Probabilités de Saint-Flour XIII – 1983*, volume 1117 of *Lecture Notes in Mathematics*, pages 1–198. Springer Berlin Heidelberg, 1985.

Andersen K. W., Madsen K. H., Siebner H. R., Schmidt M. N., Mørup M., and Hansen L. K. Non-parametric Bayesian graph models reveal community structure in resting state fMRI. *NeuroImage*, 100:301–315, 2014.

Anderson T. W. *An Introduction to Multivariate Statistical Analysis*. Wiley series in probability and mathematical statistics. Wiley, 1984.

Andrieu C., de Freitas N., Doucet A., and Jordan M. An introduction to MCMC for machine learning. *Machine Learning*, 50(1-2):5–43, 2003.

Anwander A., Tittgemeyer M., von Cramon D., Friederici A. D., and Knösche T. Connectivity-based parcellation of Broca's area. *Cerebral Cortex*, 17(4):816–825, 2006.

Assaf Y., Alexander D. C., Jones D. K., Bizzi A., Behrens T. E. J., Clark C. A., Cohen Y., Dyrby T. B., Huppi P. S., Knoesche T. R., LeBihan D., Parker G. J. M., and Poupon C. The CONNECT project: Combining macro- and micro-structure. *NeuroImage*, 80:273 – 282, 2013.

Atay-Kayis A. and Massam H. A Monte Carlo method for computing the marginal likelihood in nondecomposable Gaussian graphical models. *Biometrika*, 92(2):317–335, 2005.

Bai F., Shu N., Yuan Y., Shi Y., Yu H., Wu D., Wang J., Xia M., He Y., and Zhang Z. Topologically convergent and divergent structural connectivity patterns between patients with remitted geriatric depression and amnestic mild cognitive impairment. *Journal of Neuroscience*, 32(12):4307–4318, 2012.

Bargmann C. I. Beyond the connectome: how neuromodulators shape neural circuits. *BioEssays*, 34(6):458–465, 2012.

Bargmann C. I. and Marder E. From the connectome to brain function. *Nature Methods*, 10(6):483–490, 2013.

Baronchelli A., Ferrer-i-Cancho R., Pastor-Satorras R., Chater N., and Christiansen M. Networks in cognitive science. *Trends in Cognitive Sciences*, 17(7):348–360, 2013.

Bassett D. S., Brown J. A., Deshpande V., Carlson J. M., and Grafton S. T. Conserved and variable architecture of human white matter connectivity. *NeuroImage*, 54(2):1262–1279, 2011.

Beckmann C. F. and Smith S. M. Probabilistic independent component analysis for functional magnetic resonance imaging. *IEEE transactions on medical imaging*, 23(2):137–152, 2004.

Beckmann C. F., DeLuca M., Devlin J. T., and Smith S. M. Investigations into resting-state connectivity using independent component analysis. *Philosophical Transactions of the Royal Society: Series B*, 360(1457):1001–1013, 2005.

Beckmann M., Johansen-Berg H., and Rushworth M. F. S. Connectivity-based parcellation of human cingulate cortex and its relation to functional specialization. *Journal of Neuroscience*, 29(4):1175–1190, 2009.

Beggs J. M. and Timme N. *Frontiers in Physiology*, (163).

Behrens T. E. J., Woolrich M. W., Jenkinson M., Johansen-Berg H., Nunes R. G., Clare S., Matthews P. M., Brady J. M., and Smith S. M. Characterization and propagation of uncertainty in diffusion-weighted MR imaging. *Magnetic Resonance in Medicine*, 50(5):1077–1088, 2003a.

Behrens T. E. J., Woolrich M. W., Smith S. M., Boulby P. A., Barker G. J., Sillery E. L., Sheehan K., Ciccarelli O., Thompson A. J., Brady J. M., and Matthews P. M. Non-invasive mapping of connections between human thalamus and cortex using diffusion imaging. *Nature Neuroscience*, 6(7):750–757, 2003b.

Behrens T. E. J., Berg H. J., Jbabdi S., Rushworth M. F. S., and Woolrich M. W. Probabilistic diffusion tractography with multiple fibre orientations: What can we gain? *NeuroImage*, 34(1):144–155, 2007.

Betzel R. F., Avena-Koenigsberger A., Goñi J., He Y., de Reus M. A., Griffa A., Vértes P. A., Mišic B., Thiran J.-P., Hagmann P., van den Heuvel M., Zuo X.-N., Bullmore E. T., and Sporns O. Generative models of the human connectome. *NeuroImage*, 2016.

Biessmann F., Plis S., Meinecke F. C., Eichele T., and Muller K. Analysis of Multimodal Neuroimaging Data. *IEEE Reviews in Biomedical Engineering*, 4:26–58, 2011.

Blumensath T., Jbabdi S., Glasser M. F., Van Essen D. C., Ugurbil K., Behrens T. E. J., and Smith S. M. Spatially constrained hierarchical parcellation of the brain with resting-state fMRI. *NeuroImage*, 76:313–324, 2013.

Blundell C. and Teh Y. W. Bayesian hierarchical community discovery. In Burges C., Bottou L., Welling M., Ghahramani Z., and Weinberger K., editors, *Advances in Neural Information Processing Systems 26*, pages 1601–1609. Curran Associates, Inc., 2013.

Bolam J. P., Hanley J. J., Booth P. A., and Bevan M. D. Synaptic organisation of the basal ganglia. *Journal of Anatomy*, 196(4):527–542, 2000.

Bowman F. D. Brain imaging analysis. *Annual Review of Statistics and Its Application*, 1(1):61–85, 2014.

Bowman F. D., Guo Y., and Derado G. Statistical approaches to functional neuroimaging data. *Neuroimaging Clinics of North America*, 17(4):441–458, 2007.

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

Box G. E. P. and Draper N. R. *Empirical Model Building and Response Surfaces*. John Wiley & Sons, new York, NY, USA, 1st edition, 1987.

Bressler S. and Menon V. Large-scale brain networks in cognition: emerging methods and principles. *Trends in Cognitive Sciences*, 14(6):277–290, 2010.

Budd J. M. L. and Kisvárday Z. F. Communication and wiring in the cortical connectome. *Frontiers in Neuroanatomy*, 6(October), 2012.

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

Buxton R., Wong E., and Frank L. Dynamics of blood flow and oxygenation changes during brain activation: the balloon model. *Magnetic Resonance in Medicine*, 39:855–864, 1998.

Cabral J., Hugues E., Kringelbach M. L., and Deco G. Modeling the outcome of structural disconnection on resting-state functional connectivity. *NeuroImage*, 62:1342–1353, 2012.

Calamante F., Masterton R. A. J., Tournier J.-D., Smith R. E., Willats L., Raffelt D., and Connelly A. Track-weighted functional connectivity (TW-FC): a tool for characterizing the structural-functional connections in the brain. *NeuroImage*, 70:199–210, 2013.

Calhoun V. D., Miller R., Pearlson G., and Adali T. The chronnectome: Time-varying connectivity networks as the next frontier in fMRI data discovery. *Neuron*, 84(2):262–274, 2014.

Cappé O., Robert C. P., and Rydén T. Reversible jump, birth-and-death and more general continuous time Markov chain Monte Carlo samplers. *Journal of the Royal Statistical Society: Series B*, 65(3):679–700, 2003.

Carbonell F., Bellec P., and Shmuel A. Quantification of the impact of a confounding variable on functional connectivity confirms anti-correlated networks in the resting-state. *NeuroImage*, 86:343–353, 2014.

Carvalho C. M., Massam H., and West M. Simulation of hyper-inverse Wishart distributions in graphical models. *Biometrika*, 94(3):647–659, 2007.

Casella G. An introduction to empirical Bayes data analysis. *The American Statistician*, 39(2):83–87, 1985.

Caspers S., Geyer S., Schleicher A., Mohlberg H., Amunts K., and Zilles K. The human inferior parietal cortex: cytoarchitectonic parcellation and interindividual variability. *NeuroImage*, 33(2):430–448, 2006.

Castellanos F., Di Martino A., Craddock R., Mehta A., and Milham M. Clinical applications of the functional connectome. *NeuroImage*, 80:527–540, 2013.

Catani M. From hodology to function. *Brain*, 130(3):602–605, 2007.

Catani M., Bodi I., and Dell'Acqua F. Comment on "The Geometric Structure of the Brain Fiber Pathways". *Science*, 337(6102):1605–1605, 2012a.

Catani M., Dell'Acqua F., Bizzi A., Forkel S., Williams S., Simmons A., Murphy D., and Thiebaut de Schotten M. Beyond cortical localization in clinico-anatomical correlation. *Cortex*, 48(10):1262–1287, 2012b.

Catani M., Thiebaut de Schotten M., Slater D., and Dell'acqua F. Connectomic approaches before the connectome. *NeuroImage*, 80:2–13, 2013.

Cavanna A. E. and Trimble M. R. The precuneus: a review of its functional anatomy and behavioural correlates. *Brain*, 129(3):564–583, 2006.

Chen G., Chen G., Xie C., and Li S.-J. Negative functional connectivity and its dependence on the shortest path length of positive network in the resting-state human brain. *Brain connectivity*, 1(3):195–206, 2011.

Cheng Y. and Lenkoski A. Hierarchical Gaussian graphical models: Beyond reversible jump. *Electronic Journal of Statistics*, 6:2309–2331, 2012.

Chung H.-W., Chou M.-C., and Chen C.-Y. Principles and limitations of computational algorithms in clinical diffusion tensor MR tractography. *American Journal of Neuroradiology*, 32:3–13, 2010.

Chung M. K., Adluru N., Dalton K. M., Alexander A. L., and Davidson R. J. Scalable brain network construction on white matter fibers. In *SPIE Medical Imaging*, volume 7962, pages 1–6, 2011.

Churchland P. M. A new approach to human cognition and its significance for the philosophy of science. In Agazzi E. and Pauri M., editors, *The Reality of the Unobservable*, volume 215 of *Boston Studies in the Philosophy of Science*, pages 59–85. Springer Netherlands, 2000.

Churchland P. S. and Sejnowski T. J. *The Computational Brain*. MIT Press, Cambridge, MA, USA, 1st edition, 1994.

Cipra B. A. The Best of the 20th Century: Editors Name Top 10 Algorithms. *SIAM News*, 33(4), 2000.

Clopath C., Büsing L., Vasilaki E., and Gerstner W. Connectivity reflects coding: a model of voltage-based STDP with homeostasis. *Nature Neuroscience*, 13(3):344–352, 2010.

Cloutman L. L. and Lambon Ralph M. A. Connectivity-based structural and functional parcellation of the human cortex using diffusion imaging and tractography. *Frontiers in Neuroanatomy*, (34), 2012.

Conturo T. E., Lori N. F., Cull T. S., Akbudak E., Snyder A. Z., Shimony J. S., McKinstry R. C., Burton H., and Raichle M. E. Tracking neuronal fiber pathways in the living human brain. *Proceedings of the National Acaddemy of Sciences of the USA*, 96(18):10422–10427, 1999.

Cook P. A., Bai Y., Seunarine K. K., Hall M. G., Parker G. J., and Alexander D. C. Camino: open-source diffusion-MRI reconstruction and processing. In *14th Scientific Meeting of the International Society for Magnetic Resonance in Medicine*, volume 14, page 2759, Seattle, WA, USA, 2006.

Cook P. A., Symms M., Boulby P. A., and Alexander D. C. Optimal acquisition orders of diffusion-weighted MRI measurements. *Journal of Magnetic Resonance Imaging*, 25:1051–1058, 2007.

Cover T. M. and Thomas J. A. *Elements of Information Theory*. Wiley, New Jersey, 2nd edition, 2006.

Craddock R. C., James G. A., Holtzheimer P. E., Hu X. P., and Mayberg H. S. A whole brain fMRI atlas generated via spatially constrained spectral clustering. *Human Brain Mapping*, 33(8):1914–1928, 2012.

Craddock R. C., Jbabdi S., Yan C.-G., Vogelstein J. T., Castellanos F. X., Martino A. D., Kelly C., Heberlein K., Colcombe S., and Milham M. P. Imaging human connectomes at the macroscale. *Nature Methods*, 10(6):524–539, 2013.

Cribben I., Wager T. D., and Lindquist M. A. Detecting functional connectivity change points for single-subject fMRI data. *Frontiers in Computational Neuroscience*, 7:143, 2013.

Dahl J., Vandenberghe L., and Roychowdhury V. Covariance selection for non-chordal graphs via chordal embedding. *Optimization Methods and Software*, 23(4):501–520, 2008.

Damoiseaux J. S. and Greicius M. D. Greater than the sum of its parts: a review of studies combining structural connectivity and resting-state functional connectivity. *Brain Structure and Function*, 213(6):525–533, 2009.

Damoiseaux J. S., Rombouts S. A. R. B., Barkhof F., Scheltens P., Stam C. J., Smith S. M., and Beckmann C. F. Consistent resting-state networks across healthy subjects. *Proceedings of the National Acaddemy of Sciences of the USA*, 103(2):13848–13853, sep 2006.

Dauguet J., Peled S., Berezovskii V., Delzescaux T., Warfield S. K., Born R., and Westin C.-F. Comparison of fiber tracts derived from in-vivo DTI tractography with 3D histological neural tract tracer reconstruction on a macaque brain. *NeuroImage*, 37(2):530–538, 2007.

Daunizeau J., David O., and Stephan K. E. Dynamic causal modelling: A critical review of the biophysical and statistical foundations. *NeuroImage*, 58(2):312–322, September 2011.

Dawid A. P. and Lauritzen S. L. Hyper Markov laws in the statistical analysis of decomposable graphical models. *Annals of Statistics*, 21:1272–1317, 1993.

Deco G. and Kringelbach M. L. Great expectations: Using whole-brain computational connectomics for understanding neuropsychiatric disorders. *Neuron*, 84(5):892–905, 2014.

DeFelipe J. From the connectome to the synaptome: an epic love story. *Science*, 330(6008):1198–1201, 2010.

Deligianni F., Varoquaux G., Thirion B., Robinson E., Sharp D. J., Edwards A. D., and Rueckert D. A probabilistic framework to infer brain functional connectivity from anatomical connections. *Information Processing in Medical Imaging*, 22:296–307, 2011.

Dell'Acqua F. and Catani M. Structural human brain networks: hot topics in diffusion tractography. *Current Opinion in Neurology*, 25(4):375–383, 2012.

Dempster A. P. Covariance selection. *Biometrics*, 28:157–175, 1972.

Descartes R. De homine. *apud Franciscum Moyardum and Petrum Leffen.*, 1662.

Diaconis P. and Ylvisaker D. Conjugate priors for exponential families. *Annals of Statistics*, 7(2):269–281, 1979.

Dobra A. and Lenkoski A. Copula Gaussian graphical models and their application to modeling functional disability data. *Annals of Applied Statistics*, 5(2A):969–993, 2011.

Dobra A., Lenkoski A., and Rodriguez A. Bayesian inference for general Gaussian graphical models with application to multivariate lattice data. *Journal of the American Statistical Association*, 106:1418–1433, 2011.

Doucet G., Naveau M., Petit L., Delcroix N., Zago L., Crivello F., Jobard G., Tzourio-Mazoyer N., Mazoyer B., Mellet E., and Joliot M. Brain activity at rest: A multi-scale hierarchical functional organization. *Journal of Neurophysiology*, 105:2753–2763, 2011.

Efron B. and Tibshirani R. J. *An Introduction to the bootstrap*. Chapman & Hall, New York, 1993.

Ekman M., Derrfuss J., Tittgemeyer M., and Fiebach C. J. Predicting errors from reconfiguration patterns in human brain networks. *Proceedings of the National Academy of Sciences of the USA*, 109(41):16714–16719, 2012.

Erdős P. and Rényi A. On the evolution of random graphs. *Publications of the Mathematical Institute of the Hungarian Academy of Sciences*, 5:17–61, 1960.

Felleman D. J. and Van Essen D. C. Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex*, 1(1):1–47, 1991.

Fiecas M., Ombao H., van Lunen D., Baumgartner R., Coimbra A., and Feng D. Quantifying temporal correlations: a test-retest evaluation of functional connectivity in resting-state fMRI. *NeuroImage*, 65:231–241, 2013.

Finger S. *Origins of Neuroscience: A History of Explorations Into Brain Function*. Oxford University Press paperback. Oxford University Press, 2001.

Fisher R. A. Frequency distribution of the values of the correlation coefficient in samples of an indefinitely large population. *Biometrika*, 10(4):507–521, 1915.

Fornito A. and Bullmore E. T. Connectomic intermediate phenotypes for psychiatric disorders. *Frontiers in Psychiatry*, 2012.

Fornito A. and Bullmore E. T. Connectomics: A new paradigm for understanding brain disease. *European Neuropsychopharmacology*, 25(5):733–748, 2014.

Fornito A., Zalesky A., and Breakspear M. The connectomics of brain disorders. *Nature Reviews Neuroscience*, 16(3): 159–172, 2015.

Friedman J., Hastie T., and Tibshirani R. Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, 9 (3):432–441, 2008.

Friman O., Farnebäck G., and Westin C.-F. A Bayesian approach for stochastic white matter tractography. *IEEE Transactions in Medical Imaging*, 25(8):965–978, 2006.

Friston K. Functional and effective connectivity: a review. *Brain Connectivity*, 1(1):13–35, 2011.

Friston K. J., Harrison L., and Penny W. Dynamic causal modelling. *NeuroImage*, 19(4):1273–1302, 2003.

Friston K. Functional and effective connectivity in neuroimaging: A synthesis. *Human Brain Mapping*, 2:56–78, 1994.

Gelman A. and Rubin D. B. Inference from iterative simulation using multiple sequences. *Statistical Science*, 7(4): 457–472, 1992.

Gelman A., Carlin J. B., Stern H. S., Dunson D. B., Vehtari A., and Rubin D. B. *Bayesian Data Analysis, Third Edition*. Chapman & Hall/CRC Texts in Statistical Science. Taylor & Francis, 2013.

Gilks W., Richardson S., and Spiegelhalter D. *Markov chain Monte Carlo in practice*. Chapman & Hall/CRC, 1996.

Glickstein M. *Neuroscience: A Historical Introduction*. MIT Press, 2014.

Gloor P., Salanova V., Olivier A., and Quesney L. F. The human dorsal hippocampal commissure. *Brain*, 116(5):1249–1273, 1993.

Gong G., He Y., Concha L., Lebel C., Gross D., Evans A., and Beaulieu C. Mapping anatomical connectivity patterns of human cerebral cortex using in vivo diffusion tensor imaging tractography. *Cerebral Cortex*, 19(3):524–536, 2009.

Goulas A., Bastiani M., Bezgin G., Uylings H. B. M., Roebroeck A., and Stiers P. Comparative analysis of the macroscale structural connectivity in the macaque and human brain. *PLoS Computational Biology*, 10(3):e1003529, 2014.

Green P. J. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, 82:711–732, 1995.

Greicius M. D., Supekar K., Menon V., and Dougherty R. F. Resting-state functional connectivity reflects structural connectivity in the default mode network. *Cerebral Cortex*, 19(1):72–78, 2009.

Griffa A., Baumann P., Thiran J., and Hagmann P. Structural connectomics in brain diseases. *NeuroImage*, 80:515–526, 2013.

Groves A. R., Beckmann C. F., Smith S. M., and Woolrich M. W. Linked independent component analysis for multimodal data fusion. *NeuroImage*, 54(3):2198–2217, 2011.

Gu S., Pasqualetti F., Cieslak M., Telesford Q. K., Yu A. B., Kahn A. E., Medaglia J. D., Vettel J. M., Miller M. B., Grafton S. T., and Bassett D. S. Controllability of structural brain networks. *Nature Communications*, 6:8414, 2015.

Guimerà E. and Sales-Pardo M. Missing and spurious interactions and the reconstruction of complex networks. *Proceedings of the National Academy of Sciences of the USA*, 106(52):22073–22078, 2009.

Hagmann P. *From diffusion MRI to brain connectomics*. PhD thesis, École Polytechnique Fédérale de Lausanne, May 2005.

Hagmann P., Kurant M., Gigandet X., Thiran P., Wedeen V., Meuli R., and Thiran J. P. Mapping human whole-brain structural networks with diffusion MRI. *PLoS ONE*, 2(7):e597, 2007.

Hagmann P., Cammoun L., Gigandet X., Meuli R., Honey C. J., Wedeen V., and Sporns O. Mapping the structural core of human cerebral cortex. *PLoS Biology*, 6(7):e159, 2008.

Haimovici A., Tagliazucchi E., Balenzuela P., and Chialvo D. R. Brain organization into resting state networks emerges at criticality on a model of the human connectome. *Physical Review Letters*, 110(17):178101, 2013.

Haller S. and Bartsch A. J. Pitfalls in fMRI. *European Radiology*, 19(11):2689–2706, 2009.

Harris K. D. and Mrsic-Flogel T. D. Cortical connectivity and sensory coding. *Nature*, 503(7474):51–58, 2013.

Hastie T., Tibshirani R., and Friedman J. *The elements of statistical learning: data mining, inference and prediction*. Springer, 2 edition, 2009.

He Y., Chen Z., and Evans A. Small-world anatomical networks in the human brain revealed by cortical thickness from MRI. *Cerebral Cortex*, 17:2407–2419, 2008.

He Y., Dagher A., Chen Z., Charil A., Zijdenbos A., Worsley K., and Evans A. Impaired small-world efficiency in structural cortical networks in multiple sclerosis associated with white matter lesion load. *Brain*, 132(Pt 12):3366–3379, 2009.

Helmstaedter M. Cellular-resolution connectomics: challenges of dense neural circuit reconstruction. *Nature Methods*, 10(6):501–507, 2013.

Herculano-Houzel S. The human brain in numbers: A linearly scaled-up primate brain. *Frontiers in Human Neuroscience*, 3(31):1–11, 2009.

Hermundstad A. M., Bassett D. S., Brown K. S., Aminoff E. M., Clewett D., Freeman S., Frithsen A., Johnson A., Tipper C. M., Miller M. B., Grafton S. T., and Carlson J. M. Structural foundations of resting-state and task-based functional connectivity in the human brain. *Proceedings of the National Academy of Sciences of the USA*, 110(15):6169–6174, 2013.

Hesse J. and Gross T. *Frontiers in Systems Neuroscience*, (166).

Hilgetag C. and Kaiser M. Clustered organization of cortical connectivity. *Neuroinformatics*, 2(3):353–360, 2004.

Hinne M., Heskes T., Beckman C. F., and van Gerven M. A. J. Bayesian inference of structural brain networks. *NeuroImage*, 66:543–552, 2013.

Hinne M., Lenkoski A., Heskes T., and van Gerven M. A. J. Efficient sampling of Gaussian graphical models using conditional Bayes factors. *Stat*, 3:326–336, 2014.

Hlinka J., Palus M., Vejmelka M., Mantini D., and Corbetta M. Functional connectivity in resting-state fMRI: is linear correlation sufficient? *NeuroImage*, 54(3):2218–2225, 2011.

Hoffman M. and Gelman A. The no-U-turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research*, 15:30, 2014.

Honey C. J., Kötter R., Breakspear M., and Sporns O. Network structure of cerebral cortex shapes functional connectivity on multiple time scales. *Proceedings of the National Academy of Sciences of the USA*, (24):10240–10245, 2007.

Honey C. J., Sporns O., Cammoun L., Gigandet X., Thiran J. P., Meuli R., and Hagmann P. Predicting human resting-state functional connectivity from structural connectivity. *Proceedings of the National Academy of Sciences of the USA*, 106(6):2035–2040, 2009.

Honey C. J., Thivierge J.-P., and Sporns O. Can structure predict function in the human brain? *NeuroImage*, 52:766–776, 2010.

Horwitz B. and Poeppel D. How can EEG/MEG and fMRI/PET data be combined? *Human Brain Mapping*, 17(1): 1–3, 2002.

Huang S., Li J., Sun L., Ye J., Fleisher A., Wu T., Chen K., and Reiman E. Learning brain connectivity of Alzheimer's disease by sparse inverse covariance estimation. *NeuroImage*, 50(3):935–949, 2010.

Huber P. J. and Ronchetti E. M., editors. *Robust statistics*. Wiley, 2nd edition, 2009.

Hutchison R. M., Womelsdorf T., Allen E. A., Bandettini P. A., Calhoun V. D., Corbetta M., Della Penna S., Duyn J. H., Glover G. H., Gonzalez-Castillo J., Handwerker D. A., Keilholz S., Kiviniemi V., Leopold D. A., de Pasquale F., Sporns O., Walter M., and Chang C. Dynamic functional connectivity: promise, issues, and interpretations. *NeuroImage*, 80:360–378, 2013.

Insel T. R., Landis S. C., and Collins F. S. The NIH BRAIN initiative. *Science*, 340:687–688, 2013.

Insua R. D. and Ruggeri F., editors. *Robust Bayesian analysis*. Lecture notes in statistics. Springer, New York, 2000.

Iturria-Medina Y., Sotero R., Canales-Rodríguez E., Alemán-Gómez Y., and Melie-García L. Studying the human brain anatomical network via diffusion-weighted MRI and graph theory. *NeuroImage*, 40(3):1064–1076, 2008.

Iyer S. P., Shafran I., Grayson D., Gates K., Nigg J. T., and Fair D. A. Inferring functional connectivity in MRI using Bayesian network structure learning with a modified PC algorithm. *NeuroImage*, 75:165–175, 2013.

Jain S. and Neal R. M. Splitting and merging components of a nonconjugate Dirichlet process mixture model. *Bayesian Analysis*, 2(3):445–472, 2007.

Jakab A., Schwartz E., Kasprian G., Gruber G. M., Prayer D., Schöpf V., and Langs G. Fetal functional imaging portrays heterogeneous development of emerging human brain networks. *Frontiers in Human Neuroscience*, 8(852):852, 2014.

Janssen R. J., Hinne M., Heskes T., and van Gerven M A J . Quantifying uncertainty in brain network measures using Bayesian connectomics. *Frontiers in Computational Neuroscience*, 8(126), 2014.

Javanmard A. and Montanari A. Confidence intervals and hypothesis testing for high-dimensional regression. *Journal of Machine Learning Research*, 15:2869–2909, 2014.

Jbabdi S. and Johansen-Berg H. Tractography: Where do we go from here? *Brain Connectivity*, 1(3):169–183, 2011.

Jbabdi S., Woolrich M. W., Andersson J. L., and Behrens T. E. A Bayesian framework for global tractography. *NeuroImage*, 37(1):116–129, 2007.

Jbabdi S., Lehman J. F., Haber S. N., and Behrens T. E. J. Human and monkey ventral prefrontal fibers use the same organizational principles to reach their targets: Tracing versus tractography. *The Journal of Neuroscience*, 33(7): 3190–3201, 2013.

Jbabdi S., Sotiropoulos S. N., Haber S. N., van Essen D. C., and Behrens T. E. J. Measuring macroscopic brain connections in vivo. *Nature Neuroscience*, 18(11):1546–1555, 2015.

Jeffreys H. *Theory of Probability*. Oxford, Oxford, England, third edition, 1961.

Jenkinson M., Beckmann C. F., Behrens T. E., Woolrich M. W., and Smith S. M. FSL. *NeuroImage*, 62:782–790, 2012.

Jirsa V. K., Sporns O., Breakspear M., Deco G., and McIntosh A. R. Towards the virtual brain: Network modeling of the intact and the damaged brain. *Archives Italiennes de Biologie*, 148(3):189–205, 2010.

Johansen-Berg H. Human connectomics - What will the future demand? *NeuroImage*, 80:541–544, 2013.

Johansen-Berg H., Behrens T. E. J., Robson M. D., Drobnjak I., Rushworth M. F. S., Brady J. M., Smith S. M., Higham D. J., and Matthews P. M. Changes in connectivity profiles define functionally distinct regions in human medial frontal cortex. *Proceedings of the National Academy of Sciences of the USA*, 101(36):13335–13340, 2004.

Johansen-Berg H., Behrens T., Sillery E., Ciccarelli O., Thompson A., Smith S., and Matthews P. Functional-anatomical validation and individual variation of diffusion tractography-based segmentation of the human thalamus. *Cerebral Cortex*, 15(1):31–39, 2005.

Jones B., Carvalho C., Dobra A., Hans C., Carter C., and West M. Experiments in stochastic computation for high-dimensional graphical models. *Statistical Science*, 20(4):388–400, 2005.

Jones D. K. Challenges and limitations of quantifying brain connectivity in vivo with diffusion MRI. *Imaging in Medicine*, 2(3):341–355, 2010.

Jones D. K., Knösche T. R., and Turner R. White matter integrity, fiber count, and other fallacies: The do's and don'ts of diffusion MRI. *NeuroImage*, 73:239–254, 2013.

Kaden E., Knösche T. R., and Anwander A. Parametric spherical deconvolution: Inferring anatomical connectivity using diffusion MR imaging. *NeuroImage*, 37(2):474–488, 2007.

Kass R. E. Statistical inference: The big picture. *Statistical Science*, 26:1–9, 2011.

Kelso J. A. S., Dumas G., and Tognoli E. Outline of a general theory of behavior and brain coordination. *Neural Networks*, 37:120–131, 2013.

Kemp C., Tenenbaum J., Griffiths T., Yamada T., and Ueda N. Learning systems of concepts with an infinite relational model. In *Proceedings of the 21st national conference on Artificial intelligence*, volume 1 of *AAAI'06*, pages 381–388. AAAI Press, 2006.

Klein J. C., Behrens T. E. J., Robson M. D., Mackay C. E., Higham D. J., and Johansen-Berg H. Connectivity-based parcellation of human cortex using diffusion MRI: Establishing reproducibility, validity and observer independence in BA 44/45 and SMA/pre-SMA. *NeuroImage*, 34(1):204–211, 2007.

Knösche T. R. and Tittgemeyer M. The role of long-range connectivity for the characterization of the functional-anatomical organization of the cortex. *Frontiers in Systems Neuroscience*, 5(58):1–13, 2011.

Koch M., Norris D. G., and Hund-Georgiadis M. An investigation of functional and anatomical connectivity using magnetic resonance imaging. *NeuroImage*, 16(1):241–250, May 2002.

Kopell N. J., Gritton H. J., Whittington M. A., and Kramer M. A. Beyond the Connectome: The Dynome. *Neuron*, 83(6):1319–1328, 2014.

Korattikara A., Chen Y., and Welling M. Sequential tests for large-scale learning. *Neural Computing*, :1–26, 2015.

Kucyi A. and Davis K. D. Dynamic functional connectivity of the default mode network tracks daydreaming. *NeuroImage*, 100:471–480, 2014.

Kullback S. and Leibler R. A. On information and sufficiency. *Annals of Mathemathical Statistics*, 22(1):79–86, 1951.

Laplace P. S. *Essai philosophique sur les probabilités*. Courcier, 1814.

Lauritzen S. L. *Graphical models*. Oxford University Press, New York, 1996.

Le Bihan D., Mangin J.-F., Poupon C., Clark C., Pappata N., S. Molko, and Chabriat H. Diffusion tensor imaging: concepts and applications. *Journal of Magnetic Resonance Imaging*, 13(4):534–546, 2001.

Lenkoski A. A direct sampler for G-Wishart variates. *Stat*, 2(1):119–128, 2013.

Lenkoski A. and Dobra A. Computational aspects related to inference in Gaussian graphical models with the G-Wishart prior. *Journal of Computational and Graphical Statistics*, 20(1):140–157, 2011.

Li L., Rilling J. K., Preuss T. M., Glasser M. F., and Hu X. The effects of connection reconstruction method on the interregional connectivity of brain networks via diffusion tractography. *Human Brain Mapping*, 33(8):1894–1913, 2012.

Liang F. A double Metropolis-Hastings sampler for spatial models with intractable normalizing constants. *Journal of Statistical Computation and Simulation*, 80:1007–1022, 2010.

Liang Z., King J., and Zhang N. Anticorrelated resting-state functional connectivity in awake rat brain. *NeuroImage*, 59(2):1190–1199, 2012.

Lichtman J. W., Pfister H., and Shavit N. The big data challenges of connectomics. *Nature Neuroscience*, 17(11):1448–1454, 2014.

Liseo B. Robustness issues in bayesian model selection. In Insua D. R. and Ruggeri F., editors, *Robust Bayesian Analysis*, volume 152 of *Lecture Notes in Statistics*, pages 197–222. Springer New York, 2000.

Liu H., Qin W., Li W., Fan L., Wang J., Jiang T., and Yu C. Connectivity-based parcellation of the human frontal pole with diffusion tensor imaging. *Journal of Neuroscience*, 33(16):6782–6790, 2013.

Lo C.-Y., Wang P.-N., Chou K.-H., Wang J., He Y., and Lin C.-P. Diffusion tensor tractography reveals abnormal topological organization in structural cortical networks in Alzheimer's disease. *Journal of Neuroscience*, 30:16876–16885, 2010.

Lockhart R., Taylor J., Tibshirani R., and Tibshirani R. A significance test for the lasso. *Annals of Statistics*, 42:413–468, 2014.

Lowe M. J., Dzemidzic M., Lurito J. T., Mathews V. P., and Phillips M. D. Correlations in low-frequency BOLD fluctuations reflect cortico-cortical connections. *NeuroImage*, 12(5):582–587, 2000.

Lv J., Guo L., Hu X., Zhang T., Li K., Zhang D., Yang J., and Liu T. Fiber-centered analysis of brain connectivities using DTI and resting state fMRI data. *Medical image computing and computer-assisted intervention*, 13(Pt 2):143–150, 2010.

Madsen R. E., Kauchak D., and Elkan C. Modeling word burstiness using the Dirichlet distribution. In *Proceedings of the 22nd International Conference on Machine Learning*, ICML '05, pages 545–552, New York, NY, USA, 2005. ACM.

Marinazzo D., Pellicoro M., Wu G., Angelini L., Cortés J. M., and Stramaglia S. Information transfer and criticality in the ising model on the human connectome. *PLoS ONE*, 9(4):1–7, 2014.

Markov N. T., Ercsey-Ravasz M. M., Ribeiro Gomes A. R., Lamy C., Magrou L., Vezoli J., Misery P., Falchier A., Quilodran R., Gariel M. A., Sallet J., Gamanut R., Huissoud C., Clavagnier S., Giroud P., Sappey-Marinier D., Barone P., Dehay C., Toroczkai Z., Knoblauch K., Van Essen D. C., and Kennedy H. A weighted and directed interareal connectivity matrix for macaque cerebral cortex. *Cerebral Cortex*, 24(1):17–36, 2014.

Markram H., Meier K., Lippert T., Grillner S., Frackowiak R., Dehaene S., Knoll A., Sompolinsky H., Verstreken K., DeFelipe J., Grant S., Changeux J.-P., and Saria A. Introducing the human brain project. *Procedia Computer Science*, 7:39–42, 2011.

Marrelec G. and Benali H. A theoretical investigation of the relationship between structural equation modeling and partial correlation in functional MRI effective connectivity. *Computational Intelligence and Neuroscience*, 2009, 2009.

Marrelec G., Krainik A., Duffau H., Pélégrini-Issac M., Lehéricy S., Doyon J., and Benali H. Partial correlation for functional brain interactivity investigation in functional MRI. *NeuroImage*, 32(1):228 – 237, 2006.

Mars R. B., Jbabdi S., Sallet J., O'Reilly J. X., Croxson P. L., Olivier E., Noonan M.-A. P., Bergmann C., Mitchell A. S., Baxter M. G., Behrens T. E. J., Johansen-Berg H., Tomassini V., Miller K. L., and Rushworth M. F. S. Diffusion-weighted imaging tractography-based parcellation of the human parietal cortex and comparison with human and macaque resting-state functional connectivity. *Journal of Neuroscience*, 31(11):4087–4100, 2011.

McIntosh A. R. Moving between functional and effective connectivity. 2010.

Medina J. H. and Pazo J. H. Electrophysiological evidence for the existence of caudate-caudate connections. *International Journal of Neuroscience*, 15(1-2):99–101, 1981.

Meinshausen N. A note on the graphical lasso for graphical Gaussian model selection. *Statistics & probability letters*, 78(7):880–884, 2008.

Meinshausen N. and Bühlmann P. Stability selection. *Journal of the Royal Statistical Society: Series B*, 72:417–473, 2010.

Mensah P. and Deadwyler S. The caudate nucleus of the rat: cell types and the demonstration of a commissural system. *Journal of Anatomy*, 117(2):281–293, 1974.

Messé A., Benali H., and Marrelec G. Relating structural and functional connectivity in MRI: A simple model for a complex brain. *IEEE Transactions on Medical Imaging*, 34(1):27–37, 2015a.

Messé A., Hütt M.-T., König P., and Hilgetag C. C. A closer look at the apparent correlation of structural and functional connectivity in excitable neural networks. *Scientific Reports*, 5(7870):1–5, 2015b.

Messé A., Rudrauf D., Giron A., and Marrelec G. Predicting functional connectivity from structural connectivity via computational models using MRI: An extensive comparison study. *NeuroImage*, 111:65–75, 2015c.

Meunier D., Lambiotte R., Fornito A., Ersche K. D., and Bullmore E. Hierarchical modularity in human brain functional networks. *Frontiers in Neuroinformatics*, 3(37):1–12, 2009.

Milgram S. The small world problem. *Psychology Today*, 67(1):61–67, 1967.

Minka T. P. Estimating a Dirichlet distribution. Technical report, MIT, 2000.

Miranda-Dominguez O., Mills B. D., Grayson D., Woodall A., Grant K. A., Kroenke C. D., and Fair D. A. Bridging the gap between the human and macaque connectome: A quantitative comparison of global interspecies structure-function relationships and network topology. *The Journal of Neuroscience*, 34(16):5552–5563, 2014.

Mitsakakis N., Massam H., and Escobar M. D. A Metropolis-Hastings based method for sampling from the G-Wishart distribution in Gaussian graphical models. *Electronic Journal of Statistics*, 5:18–30, 2011.

Mišić B., Betzel R. F., Nematzadeh A., Goñi J., Griffa A., Hagmann P., Flammini A., Ahn Y.-Y., and Sporns O. Cooperative and Competitive Spreading Dynamics on the Human Connectome. *Neuron*, 86(6):1518–1529, 2015.

Moghaddam B., Marlin B., Khan E., and Murphy K. Accelerating Bayesian structural inference for non-decomposable Gaussian graphical models. In Bengio Y., Schuurmans D., Lafferty J., Williams C. K. I., and Culotta A., editors, *Advances in Neural Information Processing Systems 22*, volume 23 of *NIPS'10*, pages 1285–1293. Curran Associates Inc., 2009.

Mohammadi A. and Wit E. C. Bayesian structure learning in sparse Gaussian graphical models. *Bayesian Analysis*, 1 (10):109–138, 2015.

Mørup M., Madsen K. H., Dogonowski A.-M., Siebner H., and Hansen L. K. Infinite relational modeling of functional connectivity in resting state fMRI. In *Advances in Neural Information Processing Systems*, volume 23 of *NIPS'10*, pages 1750–1758. Curran Associates, Inc., 2010.

Mukherjee S. and Speed T. P. Network inference using informative priors. *Proceedings of the National Academy of Sciences of the USA*, 105(38):14313–14318, 2008.

Murphy K., Birn R. M., Handwerker D. A., Jones T. B., and Bandettini P. A. The impact of global signal regression on resting state correlations: are anti-correlated networks introduced? *NeuroImage*, 44(3):893–905, 2009.

Murray I., Ghahramani Z., and MacKay D. J. C. MCMC for doubly-intractable distributions. In *Proceedings of the 22nd Annual Conference in Uncertainty in Artificial Intelligence*, volume 22 of *UAI*, pages 359–366. AUAI Press, 2006.

Nanetti L., Cerliani L., Gazzola V., Renken R., and Keysers C. Group analyses of connectivity-based cortical parcellation using repeated k-means clustering. *NeuroImage*, 47(4):1666–1677, 2009.

Nelsen R. B. *An Introduction to Copulas*. Lecture notes in statistics. Springer, 1999.

Newman M. *Networks: an introduction*. Oxford University Press, Oxford; New York, 2010.

Ng B., Abugharbieh R., Varoquaux G., Poline J. B., and Thirion B. Connectivity-informed fMRI activation detection. In Fichtinger G., Martel A., and Peters T., editors, *Medical Image Computing and Computer-Assisted Intervention*, volume 6892 of *Lecture Notes in Computer Science*, pages 285–292. Springer Berlin Heidelberg, 2011.

Ng B., Varoquaux G., Poline J., and Thirion B. A novel sparse graphical approach for multimodel brain connectivity inference. *Medical Image Computing and Computer-Assisted Intervention*, 15(Pt 1):707–714, 2012.

Nielsen S. F. V., Madsen K. H., Røge R., Schmidt M. N., and Mørup M. *Nonparametric modeling of dynamic functional connectivity in fMRI data*. 2015.

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

O'Donnell L. J. and Pasternak O. Does diffusion MRI tell us anything about the white matter? An overview of methods and pitfalls. *Schizophrenia Research*, 161(1):133–141, 2014.

Ogawa S., Lee T. M., Kay A. R., and Tank D. W. Brain magnetic resonance imaging with contrast dependent on blood oxygenation. *Proceedings of the National Academy of Sciences of the USA*, 87(24):9868–9872, 1990.

O'Muircheartaigh J., Vollmar C., Traynor C., Barker G. J., Kumari V., Symms M. R., Thompson P., and et al. . Clustering probabilistic tractograms using independent component analysis applied to the thalamus. *NeuroImage*, 54 (3):2020–2032, 2011.

Owhadi H., Scovel C., and Sullivan T. Brittleness of bayesian inference under finite information in a continuous world. *Electronic Journal of Statistics*, 9(1):1–79, 2015.

Park C., Kim S., Kim Y., and Kim K. Comparison of the small-world topology between anatomical and functional connectivity in the human brain. *Physica A*, 387(23):5958–5962, 2008.

Park H.-J. and Friston K. Structural and functional brain networks: from connections to cognition. *Science*, 342(6158): 1238411, 2013.

Patenaude B., Smith S. M., Kennedy D. N., and Jenkinson M. A Bayesian model of shape and appearance for subcortical brain segmentation. *NeuroImage*, 56(3):907–922, 2011.

Pavlovic D., Vértes P., Bullmore E., Schafer W., and Nichols T. Stochastic blockmodeling of the modules and core of the Caenorhabditis elegans connectome. *PLoS ONE*, 9(7):e97584, 2014.

Penny W. D., Friston K. J., Ashburner J. T., Kiebel S. J., and Nichols T. E., editors. *Statistical Parametric Mapping: The Analysis of Functional Brain Images*. Academic Press Inc., 1st edition, 2006.

Piccioni M. Independence structure of natural conjugate densities to exponential families and the Gibbs sampler. *Scandinavian Journal of Statistics*, 27:111–127, 2000.

Pineda-Pardo J. A., Bruña R., Woolrich M., Marcos A., Nobre A. C., Maestú F., and Vidaurre D. Guiding functional connectivity estimation by structural connectivity in MEG: an application to discrimination of conditions of mild cognitive impairment. *NeuroImage*, 101:765–777, 2014.

Poser B. A., Versluis M. J., Hoogduin J. M., and Norris D. G. BOLD contrast sensitivity enhancement and artifact reduction with multiecho EPI: parallel-acquired inhomogeneity-desensitized fMRI. *Magnetic Resonance in Medicine*, 55(6):1227–1235, 2006.

Pourahmadi M. Covariance estimation: The GLM and regularization perspectives. *Statistical Science*, 26(3):369–387, 2011.

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., and Petersen S. E. Functional network organization of the human brain. *Neuron*, 72(4): 665–678, 2011.

Press S. J. *Bayesian Hypothesis Testing*, pages 217–232. John Wiley & Sons, Inc., 2002.

Raj A. and Chen Y.-H. The wiring economy principle: connectivity determines anatomy in the human brain. *PLoS ONE*, 6(9):e14832, 2011.

Rao V. and Teh Y. W. MCMC for continuous-time discrete-state systems. In Pereira F., Burges C. J. C., Bottou L., and Weinberger K. Q., editors, *Advances in Neural Information Processing 25*, volume 25 of *NIPS*, pages 701–709. Curran Associates, Inc., 2012.

Reid A. T., Lewis J., Bezgin G., Khundrakpamb B., Eickhoff S., McIntosh R., Bellec P., and Evans A. C. A cross-modal, cross-species comparison of connectivity analyses in the primate cortex. *NeuroImage*, 125:311–313, 2016.

Richardson S. and Green P. On Bayesian analysis of mixtures with an unknown number of components (with discussion). *Journal of the Royal Statistical Society: Series B*, 59(4):731–792, 1997.

Ritov T., Bickel P., Gamst A., and Klein B. The Bayesian analysis of complex, high-dimensional models: can it be CODA? *Statistical Science*, 29:619–639, 2014.

Robbins H. The empirical bayes approach to statistical decision problems. *Annals of Mathemathical Statistics*, 35(1): 1–20, 1964.

Roberts J. A., Perry A., Lord A. R., Roberts G., Mitchell P. B., Smith R. E., Calamante F., and Breakspear M. The contribution of geometry to the human connectome. *NeuroImage*, 124:379–393, 2015.

Robinson E. C., Valstar M., Hammers A., Ericsson A., Edwards A. D., and Rueckert D. Multivariate statistical analysis of whole brain structural networks obtained using probabilistic tractography. *Medical image computing and computer-assisted intervention*, 11(Pt 1):486–493, 2008.

Robinson E. C., Hammers A., Ericsson A., Edwards A. D., and Rueckert D. Identifying population differences in whole-brain structural networks: a machine learning approach. *NeuroImage*, 50(3):910–919, 2010.

Robinson P., Sarkar S., Pandejee G., and Henderson J. Determination of effective brain connectivity from functional connectivity with application to resting state connectivities. *Physical Review E*, 90(1):012707, 2014.

Rosvall M. and Bergstrom C. T. Maps of random walks on complex networks reveal community structure. *Proceedings of the National Academy of Sciences of the USA*, 105(4):1118–1123, 2008.

Roverato A. Hyper inverse Wishart distribution for non-decomposable graphs and its application to Bayesian inference for Gaussian graphical models. *Scandinavian Journal of Statistics*, 29(3):391–411, 2002.

Rubinov M. and Sporns O. Complex network measures of brain connectivity: Uses and interpretations. *NeuroImage*, 52(3):1059–1069, 2010.

Rubinov M. and Sporns O. Weight-conserving characterization of complex functional brain networks. *NeuroImage*, 56(4):2068–2079, 2011.

Rushworth M. F. S., Boorman E., and Mars R. B. Comparing brain connections in different species using diffusion weighted imaging. In *Diffusion MRI: From quantitative measurement to in vivo neuroanatomy*, number 2009, pages 445–460. 2009.

Ryali S., Chen T., Supekar K., and Menon V. Estimation of functional connectivity in fMRI data using stability selection-based sparse partial correlation with elastic net penalty. *NeuroImage*, 59(4):3852–3861, 2012.

Rykhlevskaia E., Gratton G., and Fabiani M. Combining structural and functional neuroimaging data for studying brain connectivity: A review. *Psychophysiology*, 45(2):173–187, 2008.

Salimans T., Kingma D. P., and Welling M. Markov chain Monte Carlo and variational inference: Bridging the gap. volume 37 of *JMLR Proceedings*, pages 1218–1226. JMLR.org, 2015.

Salinas E. and Sejnowski T. J. Correlated neuronal activity and the flow of neural information. *Nature Reviews Neuroscience*, 2(1):539–550, 2001.

Salvador R., Suckling J., Coleman M. R., Pickard J. D., Menon D., and Bullmore E. T. Neurophysiological architecture of functional magnetic resonance images of human brain. *Cerebral Cortex*, 15(9):1332–1342, 2005.

Schleicher A., Amunts K., Geyer S., Morosan P., and Zilles K. Observer-independent method for microstructural parcellation of cerebral cortex: a quantitative approach to cytoarchitectonics. *NeuroImage*, 9(1):165–177, 1999.

Schmidt M. *Graphical model structure learning with L1-regularization*. PhD thesis, University of British Columbia, 2010.

Schmidt M., Fung G., and Rosales R. Fast optimization methods for L1 regularization: A comparative study and two new approaches. *Machine Learning: ECML 2007*, pages 286–297, 2007.

Schmidt M. N. and Mørup M. Nonparametric Bayesian modeling of complex networks: An introduction. *IEEE Signal Processing Magazine*, 30(3):110–128, 2013.

Schmittmann V. D., Jahfari S., Borsboom D., Savi A. O., and Waldorp L. J. Making Large-Scale Networks from fMRI Data. *PLoS ONE*, 10(9):e0129074, 2015.

Schultz T., Vilanova A., Brecheisen R., and Kindlmann G. Fuzzy fibers: Uncertainty in dMRI tractography. In *Scientific Visualization: Uncertainty, Multifield, Biomedical, and Scalable Visualization*, Mathematics + Visualization. Springer, 2014.

Schwarz A. J. and McGonigle J. Negative edges and soft thresholding in complex network analysis of resting state functional connectivity data. *NeuroImage*, 55(3):1132–1146, 2011.

Serra L., Cercignani M., Carlesimo G. A., Fadda L., Tini N., Giulietti G., Caltagirone C., and Bozzali M. Connectivity-based parcellation of the thalamus explains specific cognitive and behavioural symptoms in patients with bilateral thalamic infarct. *PLoS ONE*, 8(6):e64578, 2013.

Seung S. *Connectome: how the brain's wiring makes us who we are*. A Mariner Book. Houghton Mifflin Harcourt, 2012.

Shen K., Hutchison R. M., Bezgin G., Everling S., and McIntosh A. R. Network structure shapes spontaneous functional connectivity dynamics. *Journal of Neuroscience*, 35(14):5579–5588, 2015.

Shu N., Liu Y., Li K., Duan Y., Wang J., Yu C., Dong H., Ye J., and He Y. Diffusion tensor tractography reveals disrupted topological efficiency in white matter structural networks in multiple sclerosis. *Cerebral Cortex*, 21(11):2565–2577, 2011.

Simpson S. L., Hayasaka S., and Laurienti P. J. Exponential Random Graph Modeling for complex brain networks. *PLoS ONE*, 6(5):e20039, 2011.

Skudlarski P., Jagannathan K., Calhoun V., Hampson M., Skudlarska B. A., and Pearlson G. Measuring brain connectivity: diffusion tensor imaging validates resting state temporal correlations. *NeuroImage*, 43(3):554–561, 2008.

Skudlarski P., Jagannathan K., Anderson K., Stevens M. C., Calhoun V. D., Skudlarska B. A., and Pearlson G. Brain connectivity is not only lower but different in schizophrenia: a combined anatomical and functional approach. *Biological Psychiatry*, 68(1):61–69, 2010.

Smith A. D. and Bolam J. P. The neural network of the basal ganglia as revealed by the study of synaptic connections of identified neurones. *Trends in Neuroscience*, 13(7):259–265, July 1990.

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

Smith S. M. The future of FMRI connectivity. *NeuroImage*, 62(2):1257–1266, 2012.

Smith S. M., Miller K. L., Salimi-Khorshidi G., Webster M., Beckmann C. F., Nichols T. E., Ramsey J. D., and Woolrich M. W. Network modelling methods for fMRI. *NeuroImage*, 54(2):875–891, 2011.

Smith S. M., Vidaurre D., Beckmann C. F., Glasser M. F., Jenkinson M., and Miller e. a., K L. Functional connectomics from resting-state fMRI. *Trends in Cognitive Sciences*, 17(12):666–682, 2013.

Smith S. M., Nichols T. E., Vidaurre D., Winkler A. M., Behrens T. E. J., Glasser M. F., Ugurbil K., Barch D. M., van Essen D. C., and Miller K. L. A positive-negative mode of population covariation links brain connectivity, demographics and behavior. *Nature Neuroscience*, 18:1565–1567, 2015.

Sporns O. The human connectome: Origins and challenges. *NeuroImage*, 80:53–61, 2013.

Sporns O. Contributions and challenges for network models in cognitive neuroscience. *Nature Neuroscience*, 17(5): 652–660, 2014.

Sporns O., Tononi G., and Kötter R. The human connectome: A structural description of the human brain. *PLoS Computational Biology*, 1(4):e42, 2005.

Stephan K. E. and Roebroeck A. A short history of causal modeling of fMRI data. *NeuroImage*, 62(2):856–863, 2012.

Stephan K. E., Tittgemeyer M., Knösche T. R., Moran R. J., and Friston K. J. Tractography-based priors for dynamic causal models. *NeuroImage*, 47(4):1628–1638, 2009.

Stephens M. Bayesian analysis of mixture models with an unknown number of components — an alternative to reversible jump methods. *Annals of Statistics*, 28(1):40–74, 2000.

Stosiek C., Garaschuk O., Holthoff K., and Konnerth A. In vivo two-photon calcium imaging of neuronal networks. *Proceedings of the National Academy of Sciences of the USA*, 100(12):7319–7324, 2003.

Sui J., Adali T., Yu Q., Chen J., and Calhoun V. D. A review of multivariate methods for multimodal fusion of brain imaging data. *Journal of Neuroscience Methods*, 204(1):68–81, 2011.

Tang Y., Nyengaard J. R., de Groot D. M. G., and Gundersen H. J. G. Total regional and global number of synapses in the human brain neocortex. *Synapse*, 41(3):258–273, 2001.

Tank A., Foti N., and Fox E. B. Bayesian structure learning of stationary time series. In *Proceedings of the 31st Conference on Uncertainty in Artificial Intelligence*, number 31 in UAI, Amsterdam, The Netherlands, 2015.

Thiebaut De Schotten M., Urbanski M., Valabregue R., Bayle D. J., and Volle E. Subdivision of the occipital lobes: An anatomical and functional MRI connectivity study. *Cortex*, 56:121–137, 2014.

Thirion B., Varoquaux G., Dohmatob E., and Poline J.-B. Which fMRI clustering gives good brain parcellations? *Frontiers in Neuroscience*, 8(167), 2014.

Tournier J. D., Mori S., and Leemans A. Diffusion tensor imaging and beyond. *Magnetic Resonance in Medicine*, 65 (6):1532–1556, 2011.

Tzourio-Mazoyer N., Landeau B., Papathanassiou D., Crivello F., Etard O., Delcroix N., Mazoyer B., and Joliot M. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *NeuroImage*, 15(1):273–289, 2002.

Uhler C., Lenkoski A., and Richards D. Exact formulas for the normalizing constants of wishart distributions for graphical models. *arXiv preprint arXiv:1406.4901*, 2014.

Vaessen M. J., Hofman P. A. M., Tijssen H. N., Aldenkamp A. P., Jansen J. F. A., and Backes W. H. The effect and reproducibility of different clinical DTI gradient sets on small world brain connectivity measures. *NeuroImage*, 51 (3):1106–1116, 2010.

Valdés-Sosa P. A., Sánchez-Bornot J. M., Lage-Castellanos A., Vega-Hernández M., Bosch-Bayard J., Melie-García L., and Canales-Rodríguez E. Estimating brain functional connectivity with sparse multivariate autoregression. *Philosophical Transactions of the Royal Society: Series B*, 360(1457):969–981, 2005.

van de Geer S., Bühlman P., Ritov Y., and Dezeure R. On asymptotically optimal confidence regions for high-dimensional models. *Annals of Statistics*, 42:1166–1202, 2014.

van den Heuvel M. P. and Hulshoff Pol H. E. Exploring the brain network: A review on resting-state fMRI functional connectivity. *European neuropsychopharmacology*, 20(8):519–534, 2010.

van den Heuvel M. P. and Sporns O. Rich-club organization of the human connectome. *Journal of Neuroscience*, 31 (44):15775–15786, 2011.

van den Heuvel M. P. and Sporns O. Network hubs in the human brain. *Trends in Cognitive Sciences*, 17(12):683–696, 2013.

van den Heuvel M. P., Mandl R., and Hulshoff Pol H. Normalized cut group clustering of resting-state fMRI data. *PLoS ONE*, 3(4):e2001, 2008.

van den Heuvel M. P., Kersbergen K. J., de Reus M. A., Keunen K., Kahn R. S., Groenendaal F., de Vries L. S., and Benders M. J. N. L. The Neonatal Connectome During Preterm Brain Development. *Cerebral Cortex*, 25(9):3000–3013, 2015.

van Dijk K. R. A., Hedden T., Venkataraman A., Evans K. C., Lazar S. W., and Buckner R. Intrinsic functional connectivity as a tool for human connectomics: theory, properties, and optimization. *Journal of Neurophysiology*, 103 (1):297–321, 2010.

Van Essen D. C. A tension-based theory of morphogenesis and compact wiring in the central nervous system. *Nature*, 385(6614):313–318, 1997.

Van Essen D. C. Cartography and connectomes. *Neuron*, 80(3):775–790, 2013.

Van Essen D. C. and Ugurbil K. The future of the human connectome. *NeuroImage*, 62(2):1299–1310, 2012.

Van Essen D. C., Ugurbil K., Auerbach E., Barch D., Behrens T. E. J., Bucholz R., Chang A., Chen L., Corbetta M., Curtiss S. W., Della Penna S., Feinberg D., Glasser M. F., Harel N., Heath A. C., Larson-Prior L., Marcus D., Michalareas G., Moeller S., Oostenveld R., Petersen S. E., Prior F., Schlaggar B. L., Smith S. M., Snyder A. Z., Xu J., and Yacoub E. The human connectome project: A data acquisition perspective. *NeuroImage*, 62(4):2222–2231, 2012.

van Laarhoven T. and Marchiori E. Graph clustering with local search optimization: The resolution bias of the objective function matters most. *Physics Review E*, 87:012812, 2013.

van Oort E. S. B., van Cappellen van Walsum A. M., and Norris D. G. An investigation into the functional and structural connectivity of the default mode network. *NeuroImage*, 90:381–389, 2014.

Varoquaux G. and Craddock R. C. Learning and comparing functional connectomes across subjects. *NeuroImage*, 80 (0):405–415, 2013.

Varoquaux G., Gramfort A., Poline J. B., and Thirion B. Brain covariance selection: better individual functional connectivity models using population prior. In Lafferty J. D., Williams C. K. I., Shawe-Taylor J., Zemel R. S., and Culotta A., editors, *Advances in Neural Information Processing*, volume 23 of *NIPS*, pages 2334–2342, Vancouver, Canada, 2010. John Lafferty.

Venkataraman A., Rathi Y., Kubicki M., Westin C.-F., and Golland P. Joint generative model for fMRI/DWI and its application to population studies. *Medical image computing and computer-assisted intervention*, 13(Pt 1):191–199, 2010.

Vértes P. E., Alexander-Bloch A. F., Gogtay N., Giedd J. N., Rapoport J. L., and Bullmore E. T. Simple models of human brain functional networks. *Proceedings fo the National Acadmy of Sciences of the USA*, 109(15):5868–5873, 2012.

Vinh N. X., Epps J., and Bailey J. Information theoretic measures for clusterings comparison: Variants, properties, normalization and correction for chance. *Information Processing & Management*, 11:2837–2854, 2010.

Wang H. and Li S. Z. Efficient Gaussian graphical model determination under G-Wishart distributions. *Electronic Journal of Statistics*, 6:168–198, 2012.

Watts D. and Strogatz S. Collective dynamics of 'small-world' networks. *Nature*, 393:440–442, 1998.

Wedeen V., Rosene D. L., Wang R., Dai G., Mortazavi F., Hagmann P., Kaas J. H., and Tseng W.-Y. I. The Geometric Structure of the Brain Fiber Pathways. *Science*, 335(6076):1628–1634, 2012a.

Wedeen V. J., Rosene D. L., Wang R., Dai G., Mortazavi F., Hagmann P., Kaas J. H., and Tseng W.-Y. I. Response to Comment on "The Geometric Structure of the Brain Fiber Pathways". *Science*, 337(6102):1605–1605, 2012b.

Whittaker J. *Graphical Models in Applied Multivariate Statistics*. Wiley, March 2009.

Wolfe A. and Jensen D. Playing multiple roles: Discovery overlapping roles in social networks. In *CML-04 Workshop on Statistical Relational Learning and its Connections to Other Fields*, 2004.

Woolrich M. W. and Stephan K. E. Biophysical network models and the human connectome. *NeuroImage*, 80:330–338, 2013.

Xu Z. and Akella R. Improving probabilistic information retrieval by modeling burstiness of words. *Information Processing & Management*, 46(2):143–158, 2010.

Xu Z., Tresp V., Yu S., Yu K., and Kriegel H. Fast inference in infinite hidden relational models. In *Proceedings of the 5th International Workshop on Mining and Learning with Graphs (MLG 2007)*, Firenze, Italy, 2007.

Yao Z., Zhang Y., Lin L., Zhou Y., Xu C., and Jiang T. Abnormal cortical networks in mild cognitive impairment and Alzheimer's disease. *PLoS Computational Biology*, 6:e1001006, 2010.

Yatsenko D., Josić K., Ecker A. S., Froudarakis E., Cotton R. J., and Tolias A. S. Improved estimation and interpretation of correlations in neural circuits. *PLoS Computational Biology*, 11(3):e1004083, 2015.

Yeo B. T. T., Krienen F. M., Chee M. W. L., and Buckner R. L. Estimates of segregation and overlap of functional connectivity networks in the human cerebral cortex. *NeuroImage*, 88C:212–227, 2013.

Zalesky A., Fornito A., Harding I. H., Cocchi L., Yücel M., Pantelis C., and Bullmore E. T. Whole-brain anatomical networks: does the choice of nodes matter? *NeuroImage*, 50(3):970–983, 2010.

Zalesky A., Fornito A., and Bullmore E. On the use of correlation as a measure of network connectivity. *NeuroImage*, 60(4):2096–2106, 2012.

Zhang Y., Brady M., and Smith S. Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *IEEE Transactions on Medical Imaging*, 20(1):45–57, 2001.

Zhang Y., Fan L., Zhang Y., Wang J., Zhu M., Zhang Y., Yu C., and Jiang T. Connectivity-based parcellation of the human posteromedial cortex. *Cerebral Cortex*, 24(3):719–727, 2014.

Zhao P. and Yu B. On model selection consistency of lasso. *Journal of Machine Learning Research*, 7:2541–2563, 2006.

Zhu D., Zhang T., Jiang X., Hu X., Chen H., Yang N., Lv J., Han J., Guo L., and Liu T. Fusing DTI and fMRI data: A survey of methods and applications. *NeuroImage*, pages 184–191, 2013.

SAMENVATTING VOOR LEKEN

Toen je dit boekje in handen kreeg, had je vast niet meteen een beeld had van wat ik de afgelopen jaren heb onderzocht. Het staat vol met vaktermen, en als je al even gebladerd hebt, heb je waarschijnlijk de nodige formules voorbij zien komen. Niet iedereen houdt daarvan. Hier volgt een poging dat allemaal wat toegankelijker samen te vatten.

De laatste jaren heb ik me gericht op het menselijk brein. Nu zijn er talloze manieren waarop je dat zou kunnen doen, maar ik heb het paradigma van een 'hersennetwerk' aangehouden. Deze manier van kijken naar het brein is pakweg de laatste tien jaar erg populair binnen de neurowetenschap. Je kan het brein op verschillende niveaus als een netwerk beschouwen. Het meest voor de hand liggend is te kijken naar de neuronen binnen het brein, die onderling allerlei verbindingen vormen. Dit is een gigantisch netwerk van miljarden neuronen en nog veel meer verbindingen. Het wordt iets overzichtelijker wanneer we als het ware de landkaart uitzoomen, en vooral kijken naar grote groepen van neuronen en de (bundels van) verbindingen hiertussen. Een beetje zoals je niet naar elk zandpad wilt kijken wanneer je een reis van Maastricht naar Groningen voorbereidt. Ook deze grovere netwerken zijn informatief gebleken. Zo heeft de wetenschap geleerd dat de ziekte van Alzheimer te herkennen is aan een afname in clustering (een netwerkcluster is een groepje gebieden met veel interne verbindingen, maar weinig verbindingen naar de rest van het netwerk) en kunnen we breinnetwerken gebruiken om bij patiënten met epilepsie te localiseren in welk gebied epileptische activiteit begint. Ook iemands IQ is gecorreleerd met de structuur van diens hersennetwerk.

Deze medische en cognitieve voorbeelden vormen slechts een greep uit het grote aantal toepassingen van het netwerkperspectief op ons brein. Ze geven aan dat het nuttig is om naar hersennetwerken te kijken, maar ze laten nog in het midden hoe we eigenlijk aan de informatie over die netwerken zijn gekomen. Een wegennet in kaart brengen is een hele klus, maar met een beetje geduld kom je een heel eind. Bij het brein is dat makkelijker gezegd dan gedaan. Hoe lezen we af hoe iemands netwerk er uitziet, zonder iemands hoofd open te schroeven? Technieken zoals magnetic resonance imaging (MRI) bieden hiervoor uitkomst. Met behulp van MRI kunnen we zien hoe deeltjes in het brein zich verplaatsen onder invloed van een sterk magnetisch veld. Met de nodige wiskunde kunnen we vervolgens reconstrueren hoe iemands breinanatomie is (een *structurele* meting), en zelfs (indirect) hoe delen van dit brein over de tijd actief worden en weer tot rust komen (een *functionele* meting). Mooi, denk je, dan zijn we er. Een mens de scanner in aan de ene kant, een kaart van diens hersennetwerk aan de andere kant er weer uit. Helaas zijn er nog een aantal beren op de weg. Zo zijn onze meetmethodes vaak indirect en zeer gevoelig voor ruis. Het is alsof we verkeersdrukte willen meten door met onze ogen

dicht naast een snelweg te gaan staan en te luisteren, terwijl er af en toe een vliegtuig overvliegt. Hierdoor ontstaan fouten, en als dat er veel zijn kan onze uiteindelijke schatting van een netwerk een heel eind naast de waarheid liggen. Met andere woorden, zelfs met onze geavanceerde meetmethoden zijn we *onzeker* over de daadwerkelijke staat van iemands hersennetwerk.

Om goed met onzekerheid om te gaan, hebben we statistiek nodig en praten we in termen van *kansen*. Zo kunnen we het hebben over 'de kans op de aanwezigheid van een verbinding in een netwerk', of over 'de kans op het hele netwerk', of over 'de kans op een verbinding met een bepaalde sterkte', en ga zo maar door. In mijn onderzoek heb ik zogenaamde *Bayesiaanse* statistiek gebruikt. Hierbij combineren we expliciet de voorkennis over de (kansen op de) parameters waarin we zijn geïnteresseerd, met een model dat ons vertelt hoe (c.q. wat de kans is dat) deze parameters zouden leiden tot de waargenomen data. Een voorbeeldje maakt dit een stuk tastbaarder. Stel, ik maak een foto van één van mijn paranimfen die een aantal vingers opsteekt, maar de camera bewoog flink tijdens het maken van de foto. Het resultaat is een wazig beeld waaruit we niet direct kunnen afleiden hoeveel vingers er werden opgestoken, net zoals we op basis van MRI niet zomaar weten welke gebieden met elkaar zijn verbonden. We kunnen echter wel bedenken dat het aantal vingers invloed heeft op de hoeveelheid gekleurde pixels op het beeld. We kunnen dit vangen in een 'data genererend model', waarbij we voor alle mogelijke aantallen vingers beschrijven wat voor foto we daarbij verwachten, en wat voor kans die foto heeft. Bij vijf vingers is de kans op een grote blob pixels groter dan wanneer mijn paranimf maar één vinger opstak. Ook hebben we voorkennis over het aantal vingers: de kans dat iemand nul tot vijf vingers opstak is ongeveer even groot, maar deze kans neemt snel af wanneer het om meer vingers gaat. In die gevallen neemt mijn paranimf mij waarschijnlijk in het ootje. Via een eenvoudige formule (zie (1.1) voor de liefhebber) combineren we de kansen op basis van onze voorkennis met de kansen uit ons data genererend model. Het resultaat is een kansverdeling die voor ieder mogelijk aantal vingers beschrijft wat de kans is dat dat aantal daadwerkelijk was opgestoken, gegeven de foto. De belangrijkste zaken die je hiervan mag onthouden is dat we niet kijken naar slechts één uitkomst met de grootste kans, maar naar het hele landschap van mogelijke antwoorden, en dat we onze voorkennis meenemen om deze ruimte van mogelijke aantallen vingers in te dammen.

Mijn onderzoek ging niet over vingers. Maar dezelfde machinerie is toepasbaar op de data die we uit bijvoorbeeld MRI verkrijgen. Het is dan zaak om te bepalen welke hersennetwerken waarschijnlijk zijn. Daarvoor hebben we steeds twee ingrediënten nodig:

1. Een beschrijving van de kans op een gegeven hersennetwerk, vóórdat we ook maar iets hebben gemeten en

2. een beschrijving van wat de kansen op specifieke metingen zijn, wanneer we aannemen dat een hersennetwerk een gegeven structuur heeft.

De centrale vraag in mijn onderzoek was: Welke kansbeschrijvingen zijn geschikt om hersennetwerken en de bijbehorende data te beschrijven, en hoe beschrijven we hiermee de onzekerheid in onze schattingen van hersennetwerken?

In hoofdstuk 2 beantwoordde ik deze vraag voor *structurele connectiviteit*, waarbij elke verbinding in het netwerk correspondeert met een anatomische verbinding tussen twee gebieden.

Ik bekeek twee modellen om voorkennis te beschrijven. Bij het eerste model bleef ik agnostisch en was de (vooraf-)kans op een verbinding 50% — oftewel, het model had geen voorkeur voor de aan- of afwezigheid van een verbinding. Bij het tweede model koppelde ik de netwerken van verschillende proefpersonen (qua kansen) aan elkaar. Hierin ging ik ervan uit dat ieder persoon een hersennetwerk heeft dat in grote mate hetzelfde is als bij anderen, en slechts op enkele verbindingen verschilt. Door deze structurele connectiviteit te vergelijken met *functionele connectiviteit*, waarbij een sterke verbinding aangeeft dat twee gebieden op dezelfde momenten actief zijn, kon ik laten zien dat de koppeling van proefpersonen forse winst oplevert ten opzichte van het agnostische model, omdat het model hiermee een stuk minder gevoelig wordt voor ruis en imperfecties in de data.

Hoofdstuk 3 gaat in op nog een derde voorkennismodel. In dit model nemen we aan dat structurele connectiviteit *clustering* vertoont. Zoals eerder gezegd wil dat zeggen dat delen van het netwerk sterk onderling verbonden zijn, maar slechts zwak met de rest van het netwerk. In de analogie van het wegennetwerk, komt dit overeen met steden *waarbinnen* heel veel wegen zijn, maar *waartussen* slechts een paar wegen liggen. Dit model brengt een interessante twist ten opzichte van hoofdstuk 2, want de Bayesiaanse aanpak stelt ons in staat tegelijk te leren welke gebieden bij welk cluster horen, en welke verbindingen tussen gebieden aanwezig zijn. Een voorbeeld van wat ik bedoel zie je in figuur 3.2. De resultaten van deze modellen lieten zien dat de clustering tussen proefpersonen consistent is, wat aannemelijk maakt dat deze clusters daadwerkelijk iets zeggen over de structurele organisatie van het brein. Recent werk laat met behulp van netwerkanalyse zien dat het brein niet alleen clusters bevat, maar dat deze clusters onderling verbonden zijn door een aantal zeer centrale gebieden die een belangrijke rol in het netwerk vervullen. De clusterstructuur die volgt uit het voorkennismodel van dit hoofdstuk vindt deze centrale gebieden keurig terug, en laat zien dat deze anders onderling verbonden zijn dan de perifere clusters.

In hoofdstuk 4 richtte ik mij op modellen voor functionele connectiviteit. Hierbij geeft de sterkte van een verbinding aan in welke mate de activiteit van de bijbehorende gebieden correleert over de tijd. Wanneer twee gebieden duidelijk eenzelfde activatiepatroon laten zien, nemen we aan dat er tussen deze twee gebieden communicatie plaatsvindt om dit te synchroniseren. Maar het schatten van al deze verbindingen tegelijk is om een aantal technische redenen lastig. We kunnen het aantal vergelijkingen dat we moeten oplossen terugdringen door een eigenlijk heel voor de hand liggende aanname te doen: wanneer er functionele connectiviteit is (dat wil zeggen, er is communicatie tussen twee gebieden), moet er ook anatomische connectiviteit zijn, anders is er geen kanaal waarover de communicatie kan plaatsvinden. Het is wederom zoals bij een wegennetwerk: er kan alleen verkeer zijn, als er wegen liggen. Andersom kan er best een weg zijn die (even) niet wordt gebruikt. Als we alleen maar de functionele connectiviteit leren voor die verbindingen waarvan we weten dat ze anatomisch zijn verbonden, maken we het probleem makkelijker, en kunnen we betere schattingen maken, ook al hebben we slechts weinig metingen. In hoofdstuk 4 beschrijf ik een aantal simulaties waarin ik laat zien hoe goed die schatting van de anatomie dan moet zijn, om wat aan deze aanpak te hebben. Deze simulaties laten tevens zien dat de gebruikte modellen geen last hebben van een aantal problemen waar alternatieve methoden mee kampen.

Hoofdstuk 5 is vooral technisch (al vond je dat misschien van de andere hoofdstukken ook al), maar nam een belangrijke stap ten opzichte van hoofdstuk 4. Daar nam ik aan dat de anatomie op de één of andere manier bekend was, en gebruikte ik die om de functionele verbinden te leren. Maar zoals we in hoofdstuk 3 de kansverdelingen over zowel clustering als connectiviteit tegelijk leerden, kunnen we ook hier tegelijk structurele (anatomische) en functionele connectiviteit leren. Dit levert echter een aantal problemen op. Wanneer we een kansverdeling over netwerken willen leren, zouden we het liefst voor elk mogelijk netwerk weten wat diens kans is. Maar het aantal mogelijke netwerken is enorm groot, ook al is het aantal gebieden maar klein. Een netwerk bestaande uit tien gebieden heeft bijvoorbeeld al 35 184 372 088 832 mogelijke netwerken, en zie daar maar eens wat mee te gaan rekenen. Om hier mee om te gaan, gebruiken we technieken die een benadering geven: ze beschrijven de kansen van de netwerken die een redelijke kans hebben, en negeren het grootste deel van de netwerken waarvan de kans bijna nul is. Wanneer we structuur en anatomie tegelijk willen bepalen, is de zoekruimte nog vele malen groter, en zijn andere manieren van benadering nodig. In dit hoofdstuk heb ik een aantal van dergelijke technieken vergeleken en een nieuwe variant voorgesteld, die sneller een bruikbare benadering geeft dan een aantal alternatieven. Een punt om te benadrukken is dat deze methode zowel structurele als functionele connectiviteit leert, maar slechts gebruik maakt van functionele metingen. Dat we structuur tegelijk kunnen leren volgt uit de aanname dat als er communicatie (functie) is, er ook anatomie moet zijn.

De kers op de taart is hoofdstuk 6. Allereerst heb ik in hoofdstuk 6 een aantal veelgebruikte simulaties voor het schatten van functionele connectiviteit gebruikt om verder inzicht te krijgen in de methode die ik in het vorige hoofdstuk voorstelde. Ook op deze realistischere simulaties was te zien dat de aanpak goed werkt. Daarnaast heb ik in dit hoofdstuk de modellen uit hoofdstukken 2 en 5 gecombineerd, om zo op basis van zowel structurele als functionele metingen betere schattingen te krijgen van zowel structurele als functionele hersennetwerken. We noemen dit *datafusie*. Het voordeel van het combineren van meerdere bronnen van metingen is dat de schattingen preciezer worden. Als onze structurele data aangeven dat een verbinding aanwezig is, zullen we hier ook geen functionele verbinding aantreffen. Anderzijds, als de functionele data een erg overtuigend beeld geven, kan het ruis in de structurele data compenseren. We zien dit effect inderdaad terug wanneer we de verbindingen bepalen tussen een veertiental subcorticale breingebieden.

In het afsluitende hoofdstuk 7 beschouw ik een aantal van de beperkingen van de aanpak die ik heb gehanteerd. Hoewel de Bayesiaanse aanpak elegant is en in vrijwel elk probleem kan modelleren, blijkt het in de praktijk moeilijk om tevens computationeel efficiënt te zijn. Ondanks dat we werken met benaderingen in plaats van exacte kansverdelingen, is de toepasbaarheid van de aanpak in hoofdstuk 5 beperkt, omdat het uitrekenen van de kansen simpelweg te lang duurt. Nieuwe inzichten of alternatieve modellen bieden hierin hopelijk uitkomst. Ondanks deze praktische kanttekeningen, geven bovenstaande hoofdstukken aan dat we met behulp van Bayesiaanse statistiek goed in staat zijn hersennetwerken en de bijbehorende onzekerheid in kaart te brengen.

DANKWOORD / ACKNOWLEDGEMENTS

De eerste mensen die ik wil bedanken voor hun hulp zijn natuurlijk mijn (co-)promotoren Tom en Marcel, zonder wie dit onderzoek überhaupt niet begonnen was. Ik had mij geen beter begeleidingsduo kunnen wensen. Jullie hielden het hoofd koel wanneer ik in de stress dreigde te schieten, en zagen kansen waar ik juist obstakels zag. Bedankt voor het tussen-neus-en-lippen-door oplossen van problemen waar ik al een week mee zwoegde en voor jullie enthousiasme en ideeën. Ik hoop veel van jullie te hebben geleerd en nog regelmatig met jullie te kunnen samenwerken.

Maar de wetenschappelijke loopbaan startte eigenlijk al wat jaren eerder. Stijn begeleidde mij bij het schrijven van mijn bachelorscriptie, en bij onze ellenlange en diepgaande discussies werd mij voor het eerst duidelijk dat ik van die manier van denken ook mijn werk kon maken. Bedankt daarvoor Stijn, je hebt me op een bijzonder interessant carrièrepad gezet.

Waar Stijn het vuur aanwakkerde, gooide Theo daar nog grote blokken hout bovenop. Tijdens mijn master informatiekunde benaderde ik je met een exotisch onderwerp voor een afstudeerproject. Je nam dat heel serieus, maar wist mij gelukkig gaandeweg ongemerkt bij te sturen naar een gedegen scriptie (die toen al in de geest van dit boekje was). Dit proces hebben we tijdens het afstuderen bij informatica nog eens verfijnd, terwijl we ondertussen samenwerkten in de 'Friday Afternoon Tea'-club, waar ik ervaring kreeg met het schrijven van wetenschappelijke stukken. In je begeleiding stel je de student volledig voorop en manoeuvreer als een schaakspeler de stukken zo dat je student zelf inziet dat en hoe het anders moet. Ik houd dit, wanneer ik nu op mijn beurt weer anderen begeleid, steeds in het achterhoofd.

Apart from these key figures, there are many others who have somehow contributed to this thesis. Some of these did so explicitly, as the co-authors of the papers on which this book is based: Alex, Christian, Luca and Matthias (as well as of course Marcel and Tom). One co-author should be named in particular: Ronald, thanks for all your help with data preprocessing, as well as for all of our sparring sessions on interesting topics, work or otherwise.

Others didn't specifically work on the research that is presented here, but helped in creating the environment in which doing so was possible. They are, of course, my dear colleagues on the one side of the Heyedaalseweg: Alexandra, Arjen, Carst, Daniel (slacklining at the office, yay!), Eduard, Elena M., Elena S., Fabian, Gabriel, Gido, Hans, Helle, Henning, Herman, Jacopo, Josef, Kasper, Maaike, Maya, Mohsen, Nicole, Perry, Tjeerd, Robbert, Rhido, Saiden, Saskia, Simone, Suzan, Tameem, Theo, Tom C., Tom H., Twan, Wessel, Wout, as well as those on the other side: Alex, Ali, Andrew, Claudia, Elena, Eliana, Franc, Frank, Giulio, Haiteng, Ida, Irina, Iris, Jana, Jeroen, Jordy, Karen, Katja, Lieve, Linda, Linsey, Louis, Makiko, Marieke, Marjolein,

Mark, Nadine, Pasi, Paul, Peter, Pim, Ronald, Rutger, Sander, Sanne, Silvan, Umut and all those people that I might have forgotten to name.

Twee mensen uit bovenstaand rijtje steunen mij in het bijzonder wanneer ik dit proefschrift verdedig: Maya en Sanne. Bedankt voor alle gedeelde koppen thee, schalen sushi, heerlijk flauwe kantoorhumor, maaltijden, frustratiedelingsmomenten, filmsessies en Michelin-ster barbecues (jawel).

Er zijn nog een paar bofkonten die ik hier dubbel noem, omdat ze nou eenmaal formeel collega's zijn, maar vrienden op de eerste plaats: Frank, Galina, Jeroen, Maarten en Mark. Bedankt voor jullie wekelijkse gezelligheid, intervisie-gesprekken en support. Tot weldra weer op een exotisch fantasy-slagveld, of gewoon bij mij aan de eettafel.

De mensen uit de laatste groep die ik hier wil bedanken voor al hun steun zijn moeilijk met één woord te vangen. Maar bij de term 'vrijdagavond' (of ze daar nu fysiek bij aanwezig kunnen zijn of niet) weten ze wie ik bedoel. Het is mij altijd een groot genoegen met jullie de week met een hoogtepunt te mogen afsluiten en ik hoop dat we dit nog lang voortzetten.

Aan het einde van deze rit resteert mij nog een drietal grootse rotsen in de branding te bedanken: Wil en Esther, mijn ouders, en Tamar, mijn zusje (of moet ik tegenwoordig 'zus' zeggen?). Bedankt voor al jullie steun de afgelopen jaren, bij het schrijven van dit proefschrift, maar veel meer nog bij alles daarbuiten.

CURRICULUM VITÆ

ACADEMIC CAREER

2016 – PRESENT	Post-doctoral researcher & lecturer *(0.8 fte)* Radboud University, Nijmegen
2011 – 2016	PhD. Computer Science *(0.8 fte)* Radboud University, Nijmegen
2007 – 2011	MSc. Computer Science, *cum laude* Radboud University, Nijmegen
2005 – 2007	MSc. Information Science, *cum laude* Radboud University, Nijmegen
2002 – 2005	BSc. Information Science Radboud University, Nijmegen
1997 – 2002	Gymnasium, *cum laude* Kandinsky College, Nijmegen

PUBLICATIONS PART OF THIS DISSERTATION

Hinne M., Janssen R. J., Heskes T., and van Gerven M. A. J. Bayesian estimation of conditional independence graphs improves functional connectivity estimates. *PLoS Computational Biology*, 11(11):e1004534, 2015a.

Hinne M., Ekman M., Janssen R. J., Heskes T., and van Gerven M. A. J. Probabilistic clustering of the human connectome identifies communities and hubs. *PLoS ONE*, 10(1):e0117179, 2015b.

Hinne M., Lenkoski A., Heskes T., and van Gerven M. A. J. Efficient sampling of Gaussian graphical models using conditional Bayes factors. *Stat*, 3:326–336, 2014a.

Hinne M., Ambrogioni L., Janssen R. J., Heskes T., and van Gerven M. A. J. Structurally-informed Bayesian functional connectivity analysis. *Neuroimage*, 68:294–305, 2014b.

Hinne M., Heskes T., Beckmann C. F., and van Gerven M. A. J. Bayesian inference of structural brain networks. *Neuroimage*, 66:543–552, 2013.

OTHER PUBLICATIONS

Hinne M., Meijers A., Bakker R., Tiesinga P., Mørup M., and van Gerven M. A. J. The missing link: Predicting connectomes from noisy and partially observed tract tracing data. *PLoS Computational Biology*, 13(1):e1005374, 2017.

Janssen M. A. M., Hinne M., Janssen R. J., van Gerven M. A. J., Steens S. C., Góraj B., Koopmans P. P., and Kessels R. P. C. Resting-state subcortical functional connectivity in HIV-infected patients on long-term cART. *Brain Imaging and Behavior*, pages 1–6, 2016.

Janssen R. J., Hinne M., Heskes T., and van Gerven M. A. J. Quantifying uncertainty in brain network measures using Bayesian connectomics. *Frontiers in Computational Neuroscience*, 8(126), 2014.

Verberne S., van der Heijden M., Hinne M., Sappelli M., Hoenkamp E., Koldijk S., and Kraaij W. Reliability and validity of query intent assessments. *Journal of the Association for Information Science and Technology*, 64(11): 2224–2237, 2013.

Sappelli M., Verberne S., van der Heijden M., Hinne M., and Kraaij W. Collecting ground truth data for query intent. In *Proceedings of the Dutch-Belgium information retrieval workshop*, DIR'12, 2012.

Hinne M. and Marchiori E. Cutting graphs using competing ant colonies and an edge clustering heuristic. In *Proceedings of the 11th European Conference on Evolutionary Computation in Combinatorial Optimization*, EvoCOP'11, pages 60–71, 2011.

Hinne M., van der Heijden M., Verberne S., and Kraaij W. A multi-dimensional model for search intent. In *Proceedings of the Dutch-Belgium information retrieval workshop*, DIR'11, 2011.

Verberne S., Hinne M., van der Heijden M., Hoenkamp E., Kraaij W., and van der Weide T. P. How does the library searcher behave? a contrastive study of library search against ad-hoc search. In *Proceedings of the Conference on Multilingual and Multimodal Information Access Evaluation, logCLEF workshop*, CLEF'10, 2010.

van der Heijden M., Hinne M., Verberne S., Hoenkamp E., van der Weide T. P., and Kraaij W. When is a query a question? reconstructing wh-requests from ad hoc-queries. In *Proceedings of the Query Representation and Understanding workshop in conjunction with SIGIR 2010*, QRU'10, pages 17–20, 2010.

Verberne S., Hinne M., van der Heijden M., D'hondt E., Kraaij W., and van der Weide T. P. Annotating URLs with query terms: what factors predict reliable annotations? In *Proceedings of the Understanding the User workshop in conjunction with SIGIR 2009*, UIIR'09, 2009.

Hinne M., van der Heijden M., Kraaij W., Raaijmakers S., Verberne S., and van der Weide T. P. Annotation of URLs: more than the sum of parts. In *Proceedings of SIGIR 2009*, SIGIR'09, pages 632–633, 2009.

van der Heijden M., Hinne M., Kraaij W., Verberne S., and van der Weide T. P. Using query logs and click data to create improved document descriptions. In *Proceedings of the 2009 workshop on Web Search Click Data*, WSCD'09, pages 64–67, 2009.

Hinne M. Local identification of web graph communities. In *Proceedings of the 1st International Conference on Theory of Information Retrieval*, ICTIR'07, pages 261–278, 2007.