Bayesian inference of structural brain networks

Max Hinne [a,b,*], Tom Heskes [a], Christian F. Beckmann [b], Marcel A.J. van Gerven [b]

[a] Radboud University Nijmegen, Institute for Computing and Information Sciences, Nijmegen, The Netherlands
[b] Radboud University Nijmegen, Donders Institute for Brain, Cognition and Behaviour, Nijmegen, The Netherlands

ARTICLE INFO

ABSTRACT

Structural brain networks are used to model white-matter connectivity between spatially segregated brain regions. The presence, location and orientation of these white matter tracts can be derived using diffusion-weighted magnetic resonance imaging in combination with probabilistic tractography. Unfortunately, as of yet, none of the existing approaches provide an undisputed way of inferring brain networks from the streamline distributions which tractography produces. State-of-the-art methods rely on an arbitrary threshold or, alternatively, yield weighted results that are difficult to interpret. In this paper, we provide a generative model that explicitly describes how structural brain networks lead to observed streamline distributions. This allows us to draw principled conclusions about brain networks, which we validate using simultaneously acquired resting-state functional MRI data. Inference may be further informed by means of a prior which combines connectivity estimates from multiple subjects. Based on this prior, we obtain networks that significantly improve on the conventional approach.

© 2012 Elsevier Inc. All rights reserved.

Introduction

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 fiber tracts. In this paper, 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 fiber bundles, connecting separate regions of gray matter. This process is known as deterministic tractography (Chung et al., 2010; Conturo et al., 1999; Shu et al., 2011). Alternatively, fibers may be estimated using probabilistic tractography (Behrens et al., 2003, 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.

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 (Gong et al., 2009; Hagmann et al., 2007; Robinson et al., 2008). 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 (Chung et al., 2011; Hagmann et al., 2007, 2008; Vaessen et al., 2010; Zalesky et al., 2010). However, probabilistic streamlining

* Corresponding author at: Radboud University Nijmegen, Faculty of Science, Institute for Computing and Information Sciences, Postbus 9010, 6500 GL Nijmegen, The Netherlands. Fax: +31 24 365 27 28.
E-mail address: mhinne@cs.ru.nl (M. Hinne).

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; Robinson et al., 2010; Zalesky 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, 90% of the streamlines connect from voxel A to voxel B, and 10% 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 (Groves et al., 2011; Horwitz and Poeppel, 2002). 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 on network inference in systems biology (Mukherjee and Speed, 2008) and consists of two ingredients. First, a network prior is defined in terms of the classical Erdős–Rényi model (Erdős and Rényi, 1960). 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 (Damoiseaux and Greicius, 2009; Greicius et al., 2009; Honey et al., 2009; Koch et al., 2002; Lv et al., 2010; Park et al., 2008; Skudlarski et al., 2008). 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.

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, resting-state functional data and diffusion-weighted images were obtained using a Siemens Magnetom Trio 3 T 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 = 13,000 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 Neuroscience, 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 functional data 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 ≥ 0.5 were discarded. This resulted in an average region count of 115 ± 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 $\hat{\Sigma}$. Example covariance matrices are shown in Fig. 1a.

Preprocessing of diffusion imaging data

The preprocessing steps for the diffusion data were conducted using FSL FDT (Behrens et al., 2003) and consisted of correction for eddy currents and estimation of the diffusion parameters. Raw color-coded fractional anisotropy maps are shown in Fig. 1b. To obtain a measure of white-matter connectivity, we used FDT Probtrackx 2.0 (Behrens et al., 2003, 2007). 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 fractional anisotropy to prevent them from wandering around in gray matter. Streamlines in which a sharp angle ($>80°$) occurred or that had a length less than 2 mm were discarded. The output thus obtained is a matrix **N** with n_{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.
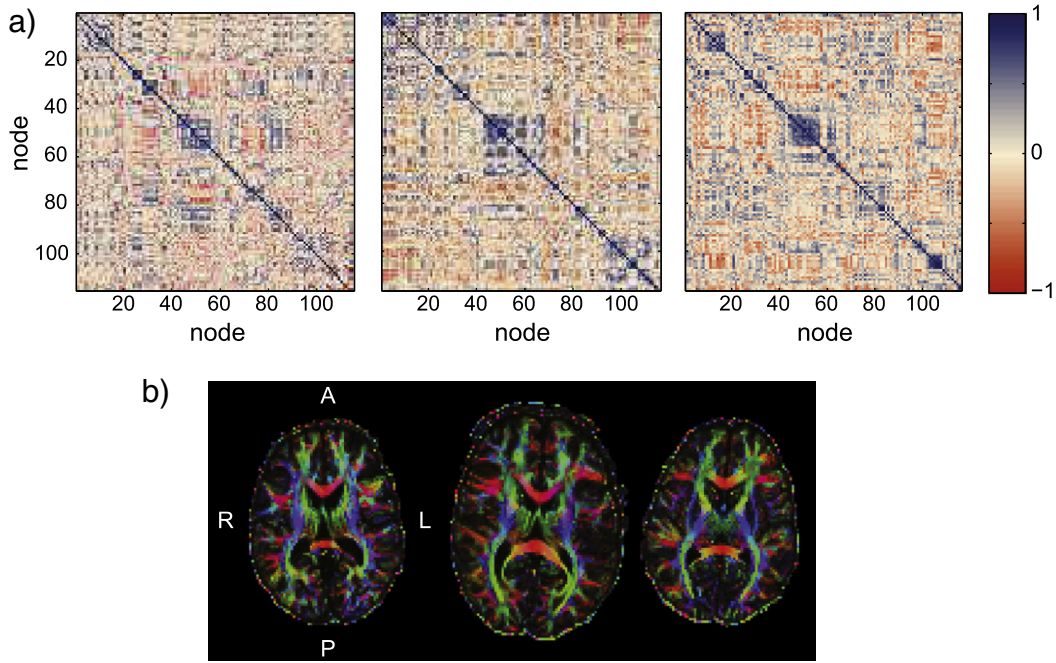
Fig. 1. a) Covariance matrices for the resting-state data for three randomly selected subjects. b) Axial view of RGB-FA maps for the diffusion weighted images, again for three randomly selected subjects. The nodes in the matrices are shown in the order they appear in the AAL atlas.

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 **A** given observed probabilistic streamlining data **N** and hyperparameters ξ. An element $a_{ij} \in \{0,1\}$ represents the absence or presence of an edge between brain region i and j. **A** is taken to be a simple graph, such that $a_{ij} = a_{ji}$ and $a_{ii} = 0$. 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:

$$P(\mathbf{A}|\mathbf{N}, \xi) \propto P\left(\mathbf{N}|\mathbf{A}, a^+, a^-\right) P(\mathbf{A}|p) \tag{1}$$

with hyperparameters $\xi = (a^+, a^-, p)$, 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 $P(\mathbf{A})$ and a forward model $P(\mathbf{N}|\mathbf{A})$ 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 the section 'Approximate inference'. We now proceed to formally define the components of the generative model as shown in Fig. 2.

Forward model

We begin with a specification of the forward model $P(\mathbf{N}|\mathbf{A})$. Here, we describe how the observed streamline distributions **N** depend on the underlying network **A** through latent streamline probabilities **X**.

Assume there are K 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 n_{ik} denote the number of streamlines which start in region i and terminate in region k. We assume that $n_{ii} = 0$. Probabilistic tractography produces a distribution over target vertices $\mathbf{n}_i = (n_{i1}, \ldots, n_{iK})^T$ by drawing S streamlines, $N_i = \sum_{k=1}^{K} n_{ik} \leq S$ of them ending up in a

target region.[1] A particular distribution $\mathbf{n}_i$ 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 $\mathbf{n}_i$ in terms of a multinomial distribution

$$P(\mathbf{n}_i|\mathbf{x}_i) \propto \prod_{j=1}^{K} x_{ij}^{n_{ij}},$$

in which $\mathbf{x}_i = (x_{i1}, \ldots, x_{iK})$ is a probability vector with $\sum_j x_{ij} = 1$. Each x_{ij} represents the probability of drawing a streamline from region i to region j. This streamlining probability itself depends on whether or not there actually exists a physical tract between region i and region j.

Let $\mathbf{a}_i$ denote the i-th row of **A** 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 **A**. This is captured by modeling the distribution of streamline probabilities using a Dirichlet distribution

$$P\left(\mathbf{x}_i|\mathbf{a}_i, a^+, a^-\right) \propto \prod_{j=1}^{K} x_{ij}^{b_{ij}-1},$$

where shorthand notation $b_{ij} \equiv a_{ij} a^+ + (1 - a_{ij}) a^-$ is used. The b_{ij} can be interpreted as the parameters that determine the probability of streamlining from region i to region j when an edge a_{ij} is either present (a^+) or absent (a^-).

To obtain a single expression for the likelihood of an adjacency matrix, let $\mathbf{N} = (\mathbf{n}_1; \ldots; \mathbf{n}_K)$ represent the combined probabilistic tractography data, i.e. for each of the K nodes a distribution of streamlines to all other nodes. Similarly, let $\mathbf{X} = (\mathbf{x}_1; \ldots; \mathbf{x}_K)$ denote the combined hidden

---

[1] It is possible that streamlines end up in voxels outside any region of the parcellation, hence the inequality.
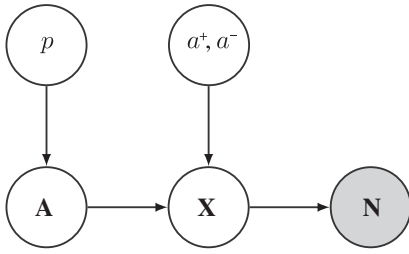
Fig. 2. The generative model that describes how the observed streamline distribution **N** depends on the (hidden) connectivity probabilities **X**. These in turn depend on the hyperparameters a^+ and a^- as well as the connectivity **A**, which is determined by hyperparameter p of the prior.

Fig. 3. 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 $\bar{\mathbf{A}}$.

connection probabilities and $\mathbf{A} = (\mathbf{a}_1; \ldots; \mathbf{a}_K)$ the adjacency matrix for all brain regions. The likelihood of the network **A** is expressed as

$$P(\mathbf{N}|\mathbf{A}, a^+, a^-) = \int P(\mathbf{N}|\mathbf{X})P\left(\mathbf{X}|\mathbf{A}, a^+, a^-\right)d\mathbf{X}. \tag{2}$$

By recognizing that the Dirichlet distribution is the conjugate prior for the multinomial distribution, it follows that Eq. (2) is a product of Dirichlet compound multinomial distributions (Madsen et al., 2005; Minka, 2000; Xu and Akella, 2010). 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 p (Erdős and Rényi, 1960). This allows the prior to be expressed in terms of a product of binomial distributions:

$$P(\mathbf{A}|p) = \prod_{i<j} p^{a_{ij}}(1-p)^{1-a_{ij}}.$$

Recall that $a_{ji} \equiv a_{ij}$ by definition, such that choosing $p = 0.5$ gives rise to a flat prior on simple graphs.

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. Fig. 3 depicts this hierarchical model.

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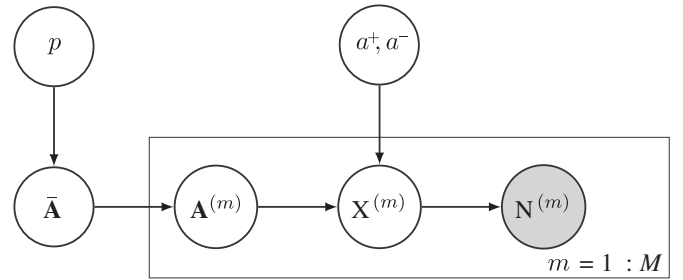
For a new subject $M+1$, the quantity of interest is the posterior marginal

$$P\left(\mathbf{A}^{(M+1)}\middle|\mathcal{N}, \mathbf{N}^{(M+1)}, \xi\right) \propto P\left(\mathbf{N}^{(M+1)}\middle|\mathbf{A}^{(M+1)}, a^+, a^-\right)P\left(\mathbf{A}^{(M+1)}\middle|\mathcal{N}, \xi\right).$$

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 $\bar{\mathbf{A}}$. 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 p to specify the connection probability for complete networks. This multi-subject prior is derived from the hierarchical model in A and is equal to:

$$P\left(\mathbf{A}^{(M+1)}|N, \xi\right) = \prod_{i<j} p_{ij}^{a_{ij}^{(M+1)}}\left(1-p_{ij}\right)^{\left(1-a_{ij}^{(M+1)}\right)}, \tag{3}$$

where $p_{ij} \equiv \left(\sum_{m=1}^{M} \hat{a}_{ij}^{(m)} + 1\right)/(M+2)$ with $\hat{a}_{ij}^{(m)}$ the maximum likelihood (ML) estimate for subject m. 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 subsequently be plugged into Eq. (1) to produce the posterior for subject $M+1$.

Approximate inference

Since the posterior (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 **A** by flipping an edge, resulting in a network $\mathbf{A}'$ (which, because of the symmetry of **A**, implies $a'_{ij} = 1-a_{ij}$ and $a'_{ij} = 1-a_{ji}$). The acceptance of the proposed sample is determined by the ratio

$$\gamma = \frac{P(\mathbf{A}'|\mathbf{N}, \xi)}{P(\mathbf{A}|\mathbf{N}, \xi)}.$$

A proposed network becomes a new sample 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 a_{kl}. A complete derivation of these terms is given in Appendix B.

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 $\{\mathbf{A}^{(1)}, \ldots, \mathbf{A}^{(T)}\}$ forms an approximation of the posterior $P(\mathbf{A}|\mathbf{N}, \xi)$. 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 $\mathbb{E}\left[a_{ij}|\mathbf{N}\right] = \frac{1}{T}\sum_{t=1}^{T}a_{ij}^{(t)}$. Other summary statistics for the distribution may be estimated in a similar manner.

Validation of structural connectivity estimates

Functional connectivity is constrained by structural connectivity (Cabral et al., 2012; Honey et al., 2010). In other words, when there is functional connectivity, there is often structural connectivity, although structural connectivity is not a necessary requirement for functional connectivity (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 (Deligianni et al., 2011; Marrelec et al., 2006; Smith et al., 2010; Varoquaux et al., 2010). Assume that a $K \times 1$ vector of BOLD responses $\mathbf{y}$ can be modeled by a zero-mean Gaussian density with inverse covariance matrix $\mathbf{Q}$. That is,

$$P(\mathbf{y}|\mathbf{Q}) = (2\pi)^{-K/2}|\mathbf{Q}|^{1/2}exp\left\{-\frac{1}{2}\mathbf{y}^{\top}\mathbf{Q}\mathbf{y}\right\}. \tag{4}$$

Then, given acquired resting-state data $\mathbf{D} = (\mathbf{y}^{1};...;\mathbf{y}^{T})$ for T time points, model estimation reduces to finding the maximum likelihood solution $\hat{\mathbf{Q}} = \arg\max_{\mathbf{Q}}\prod_{t}P(\mathbf{y}^{t}|\mathbf{Q})$. However, in general for fMRI data, $K > T$, 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., 2011). As a solution to this problem, regularization approaches have been suggested to find sparse approximations of the inverse covariance matrix (Friedman et al., 2008; Huang et al., 2010). In our setup, the sparsity structure is readily available in the form of structural connectivity $\mathbf{A}$. In order to use $\mathbf{A}$ as a constraint when estimating $\hat{\mathbf{Q}}$, we can make use of the fact that variables y_i and y_j are conditionally independent if and only if $q_{ij} = 0$ (Dempster, 1972). That is, we can interpret Eq. (4) as a Gaussian Markov random field with respect to network $\mathbf{A}$ such that $a_{ij} = 0 \Leftrightarrow q_{ij} = 0$ for all $i \neq j$ (Whittaker, 1990). We will use notation $\mathbf{Q} \sim \mathbf{A}$ to denote that the independence structure in $\mathbf{Q}$ is dictated by $\mathbf{A}$.

Let $\hat{\Sigma} = \frac{1}{T}\sum_{t=1}^{T}\mathbf{y}^{t}(\mathbf{y}^{t})^{\top}$ denote the empirical covariance matrix. As shown by Dahl et al. (2008), the ML estimate can be formulated as the following convex optimization problem:

$$\hat{\mathbf{Q}} = \arg\max_{\mathbf{Q} \sim \mathbf{A}}\ell(\mathbf{Q}) \quad \text{s.t.} \quad \left\{q_{ij} = 0 \Leftrightarrow a_{ij} = 0\right\},$$

where $\ell(\mathbf{Q}) = (T/2)\left(\log\det\mathbf{Q} - \text{trace}\left(\mathbf{Q}\hat{\Sigma}\right)\right)$ is, up to a constant, the log-likelihood function of $\mathbf{Q}$. We made use of a standard convex solver to find this constrained maximum likelihood estimate (Schmidt et al., 2007) and use it to define the score for a particular matrix $\mathbf{A}$: $S(\mathbf{A}) \equiv \ell\left(\hat{\mathbf{Q}}\right)$. 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 $\mathbf{A}_{ML} = \arg\max_{\mathbf{A}}P(\mathbf{N}|\mathbf{A}, a^{+}, a^{-})$ of that particular subject. Recall that this maximum likelihood solution is equivalent to the solution obtained when using a flat prior in our generative model. 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.

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

Comparing ML estimates with thresholded networks

The sparsity of the maximum likelihood estimates $\mathbf{A}_{ML}$, as obtained with the flat prior, was fairly constant (1019 ± 39.4 out of 6670 possible edges). As an example, Fig. 4 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 n_{ij} and n_{ji}, respectively. To compare our generative 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-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-T} over 20 subjects, given the different threshold methods, are shown in Table 1. When the threshold is applied to the maximum of n_{ij} and n_{ji}, the generative model outperforms thresholding. However, when either the mean or the minimum of n_{ij} and n_{ji} is used, samples obtained from the posterior with a flat prior score somewhat less than thresholded networks, on average. To understand the behavior of the generative model, it is instructive to consider Eq. (B.3) in Appendix B. Given that hyperparameters a^{+} and a^{-} are very small compared to elements of $\mathbf{N}$, the change in log-likelihood after flipping edge a_{ij} from absent to present boils down to

$$\Delta L_{ij} \approx \left(a^{+} - a^{-}\right)\left[\log\left(\frac{n_{ij}}{\sum_{k}n_{ik}}\right) + \log\left(\frac{n_{ji}}{\sum_{k}n_{jk}}\right)\right].$$

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 DCM by thresholding on either the mean or the minimum of n_{ij} and n_{ji} and by normalizing the streamline distribution by the number of outgoing streamlines.
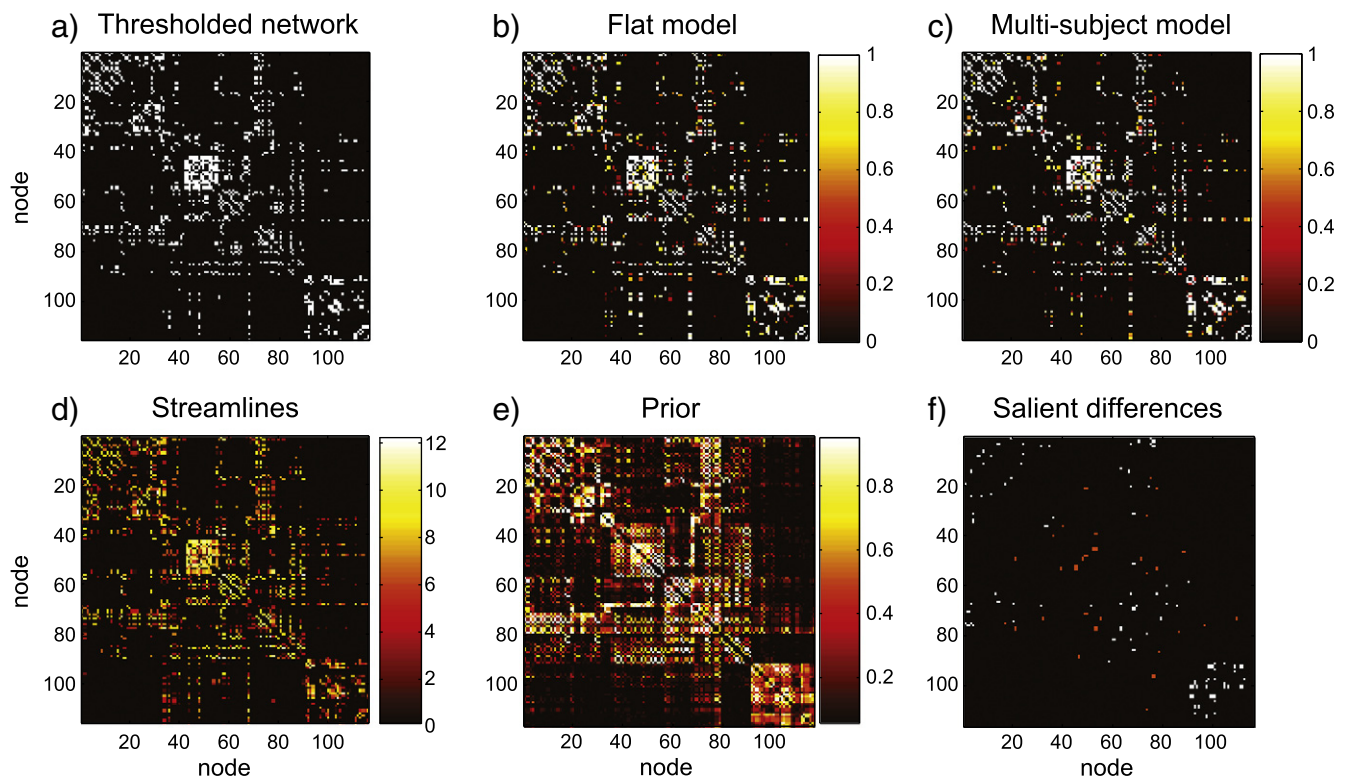
Fig. 4. (a–e) Connectivity results for that subject for which sampling in conjunction with the multi-subject prior showed the largest improvement. Shown are (a) the network that is obtained through the thresholding approach, (b) the posterior connection probabilities according to the flat model, (c) 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 6 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 of the AAL atlas.

Multi-subject prior

Thresholded networks perform roughly similar to the networks we infer through the posterior distribution using a flat prior. However, our model is capable of incorporating additional constraints, such as the multi-subject prior. Let f_{M-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-T} over 20 subjects, given the different threshold methods, are shown in Table 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-F}. We found that this distribution had a mean of 0.73 ± 0.02 ($p < 10^{-8}$).

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 Fig. 5. In addition, the figure shows the score for the thresholded network, with a threshold

applied to the mean of n_{ij} and n_{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 threshold approaches, but that it is capable of surpassing thresholded networks by using informative priors.

Lastly, Fig. 6 shows the connections for which our multi-subject approach differs most from those of thresholding, across all subjects. The edges correspond with those shown in Fig. 4(f). The figure shows edges that are present in the maximum a posteriori networks while being absent in the corresponding thresholded networks for at least 6 subjects and vice versa. The edges consistently and exclusively included by either of the approaches do not differ much in length. In fact, the mean edge lengths are very close: 17.6 ± 1.6 mm for threshold-favored edges and 17.0 ± 1.4 mm for DCM-favored edges. However, we do observe that when using the multi-subject prior, consistency for cerebellar and anterior cortical tracts is increased.

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

Table 1
The fraction of samples that have a higher score than thresholded networks. The fraction of samples from the distribution with a flat and a multi-subject prior are represented by f_{F-T} and f_{M-T}, respectively. The different threshold approaches are max, mean and min. The p-values were obtained using a one-sample t-test with $\mu_0 = 0.5$.

	f_{F-T}	p	f_{M-T}	p
Max	0.56 ± 0.06	0.15	0.76 ± 0.07	<0.001
Mean	0.45 ± 0.06	0.23	0.67 ± 0.07	0.02
Min	0.45 ± 0.07	0.27	0.66 ± 0.08	0.03

#	f_{F-T}	f_{M-T}	f_{M-F}
1	0.01	0.01	0.53
2	0.01	0.15	0.64
3	0.10	0.23	0.69
4	0.03	0.25	0.79
5	0.16	0.34	0.68
6	0.09	0.37	0.76
7	0.47	0.42	0.47
8	0.21	0.58	0.78
9	0.33	0.65	0.72
10	0.44	0.81	0.74
11	0.53	0.88	0.76
12	0.68	0.90	0.65
13	0.60	0.92	0.75
14	0.61	0.93	0.77
15	0.80	0.96	0.76
16	0.76	0.97	0.77
17	0.76	0.98	0.84
18	0.72	0.99	0.86
19	0.92	1.00	0.71
20	0.64	1.00	0.95

Multi-subject prior   Flat prior   —— Thresholded network

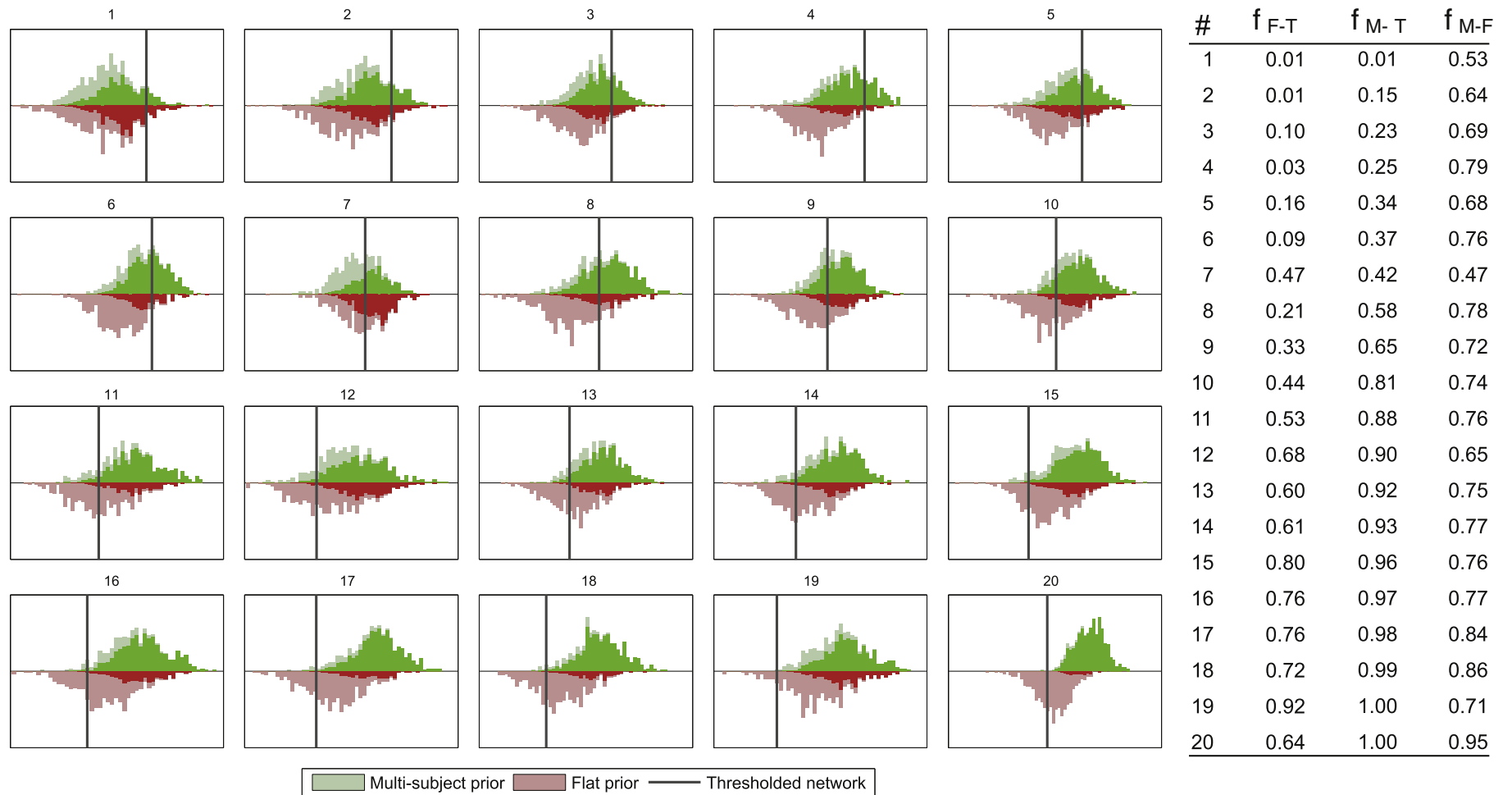**Fig. 5.** The scores $S(\mathbf{A})$ (horizontal axis) for the thresholded network $\mathbf{A}_T$, samples from the generative model given a flat prior (lower histogram, red) and given the multi-subject prior (upper histogram, green). The fraction of each bin that has a bright color corresponds with 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-T} is the fraction of samples with a flat prior that outperform thresholding, f_{M-T} is the fraction of samples with the multi-subject prior that outperform thresholding and f_{M-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.
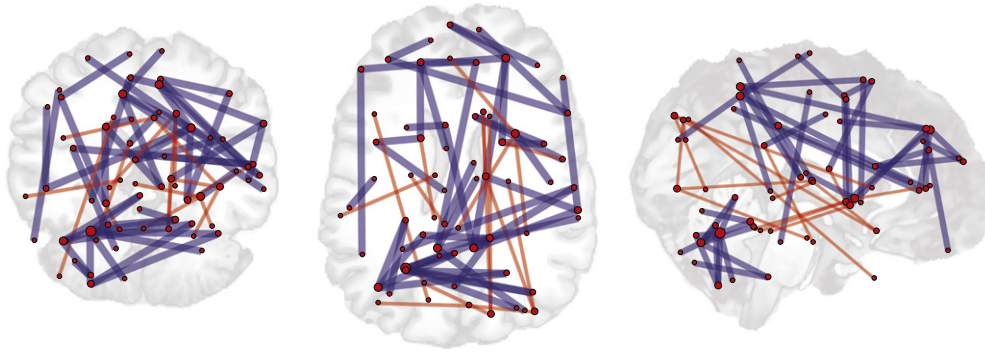
Fig. 6. The most salient differences in connectivity between the maximum a posteriori networks with the multi-subject prior and the thresholded networks, across all subjects. The edges are color-coded. Blue, thick edges indicate those connections that were present in at least 6 subjects whereas these edges would not be part of the thresholded network. Red, thin edges show converse findings. Nodes that are not adjacent to any of these edges are omitted.

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.

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 similar to the 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 a^+ and a^- were set by hand and the edge probability p 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 apart from the validation results, 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 Fig. 3 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 $\mathcal{A}$ as well as the group-level parent network $\bar{A}$. 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 paper 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 from both modalities simultaneously.

An additional benefit of our framework is that the network sparsity follows directly from optimizing Eq. (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.

Appendix A. Derivation of the multi-subject prior

We describe here the derivation of the multi-subject prior based on the maximum likelihood estimates for previously seen subjects, as described in the section 'Hierarchical model'. The prior on $\mathbf{A}' \equiv \mathbf{A}^{(M+1)}$ is given by

$$
\begin{aligned}
\Pr\left(\vec{\vec{\mathbf{A}}}' \mid \mathcal{N}, \vec{\xi}\right) &= \sum_{\vec{\vec{\mathbf{A}}}} \Pr\left(\vec{\vec{\mathbf{A}}}' \mid \vec{\vec{\mathbf{A}}}\right) \Pr\left(\vec{\vec{\mathbf{A}}} \mid \mathcal{N}, \vec{\xi}\right) \\
&\propto \sum_{\vec{\vec{\mathbf{A}}}} \Pr\left(\vec{\vec{\mathbf{A}}}' \mid \vec{\vec{\mathbf{A}}}\right) \Pr\left(\vec{\vec{\mathbf{A}}} \mid p\right) \\
&\quad \times \sum_{\mathcal{A}} \Pr\left(\mathcal{N} \mid \mathcal{A}, a^+, a^-\right) \Pr\left(\mathcal{A} \mid \vec{\vec{\mathbf{A}}}\right) \\
&\propto \sum_{\vec{\vec{\mathbf{A}}}} \Pr\left(\vec{\vec{\mathbf{A}}}' \mid \vec{\vec{\mathbf{A}}}\right) \Pr\left(\vec{\vec{\mathbf{A}}} \mid p\right) \\
&\quad \times \prod_{m=1}^{M} \sum_{\vec{\mathbf{A}}^{(m)}} \Pr\left(\vec{\mathbf{N}}^{(m)} \mid \vec{\mathbf{A}}^{(m)}, a^+, a^-\right) \\
&\quad \times \Pr\left(\vec{\mathbf{A}}^{(m)} \mid \vec{\vec{\mathbf{A}}}\right).
\end{aligned}
$$

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

$$\hat{\mathbf{A}}^{(m)} = \arg\max_{\mathbf{A}^{(m)}} P\left(\mathbf{N}^{(m)}\middle|\mathbf{A}^{(m)}, a^+, a^-\right).$$

Following the Erdős–Rényi model with $p = 0.5$ for $P(\bar{\mathbf{A}}|p)$ gives a flat prior on simple graphs. Up to irrelevant constants and keeping in mind that $\hat{\mathbf{A}}$ depends on $\mathbf{N}^{(m)}$, the prior is rewritten as

$$P\left(\mathbf{A}'|\mathcal{N}, \xi\right) \approx \sum_{\bar{\mathbf{A}}} P\left(\mathbf{A}'\middle|\bar{\mathbf{A}}\right) \prod_{m=1}^{M} P\left(\hat{\mathbf{A}}^{(m)}\middle|\bar{\mathbf{A}}\right),$$

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

$$P(\mathbf{A}^{(M+1)}|\mathcal{N}, \xi) = \prod_{i<j} P(a_{ij}^{(M+1)}|\mathcal{N}, \xi). \tag{A.1}$$

Next, we define the probability that $a_{ij}^{(m)}$, $m = (1, ..., M+1)$, inherits the connectivity from the parent network $\bar{a}_{ij}$ by

$$P\left(a_{ij}^{(m)} = 1\middle|\bar{a}_{ij} = 1\right) = P\left(a_{ij}^{(m)} = 0\middle|\bar{a}_{ij} = 0\right) \equiv q_{ij},$$

with q_{ij} close to 1. That is, each $a_{ij}^{(m)}$ is a copy of $\bar{a}_{ij}$ 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 $\sum_{m=1}^{M} \hat{a}_{ij}^{(m)}$ successes, then our best estimate of the probability that the next repetition $a_{ij}^{(M+1)}$ will be a success is:

$$P\left(a_{ij}^{(M+1)} = 1\middle|\hat{a}_{ij}^{(1)}, ..., \hat{a}_{ij}^{(M)}\right) = \frac{\sum_{m=1}^{M} \hat{a}_{ij}^{(m)} + 1}{M+2} \equiv p_{ij}.$$

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

$$P(\mathbf{A}^{(M+1)}|\mathcal{N}, \xi) = \prod_{i<j} p_{ij}^{a_{ij}^{(M+1)}} \left(1 - p_{ij}\right)^{\left(1 - a_{ij}^{(M+1)}\right)}.$$

Appendix B. MCMC sampling

We derive here the acceptance rate γ of a sample $\mathbf{A}'$ in the sampling chain as a function of one edge flip in $\mathbf{A}$ (see the 'Approximate inference' section). Note that each of the $2^{K(K-1)/2}$ possible networks $\mathbf{A}$ 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 a_{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

$$L \equiv \sum_i \left[\log \frac{N_i!}{\prod_j n_{ij}!} + \log \frac{\Gamma\left(\sum_j b_{ij}\right)}{\Gamma\left(\sum_j \left(b_{ij} + n_{ij}\right)\right)} + \sum_j \log \frac{\Gamma\left(b_{ij} + n_{ij}\right)}{\Gamma\left(b_{ij}\right)}\right] \tag{B.1}$$

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

$$\Delta L_{kl} = \log P\left(\mathbf{N}\middle|\mathbf{A}', a^+, a^-\right) - \log P\left(\mathbf{N}\middle|\mathbf{A}, a^+, a^-\right), \tag{B.2}$$

with the sole difference that $a'_{kl} = a'_{lk} = (1 - a_{kl})$. Plugging Eq. (B.1) into Eq. (B.2) yields

$$\begin{aligned}
\Delta L_{kl} = & \log\left[\frac{\Gamma\left(b'_{kl} + n_{kl}\right)}{\Gamma(b_{kl} + n_{kl})}\right] + \log\left[\frac{\Gamma\left(b'_{lk} + n_{lk}\right)}{\Gamma(b_{lk} + n_{lk})}\right] \\
& + \log\left[\frac{\Gamma\left(\sum_j b'_{kj}\right)}{\Gamma\left(\sum_j b_{kj}\right)}\right] + \log\left[\frac{\Gamma\left(\sum_j b'_{lj}\right)}{\Gamma\left(\sum_j b_{lj}\right)}\right] \\
& - \log\left[\frac{\Gamma\left(\sum_j \left(b'_{kj} + n_{kj}\right)\right)}{\Gamma\left(\sum_j \left(b_{kj} + n_{kj}\right)\right)}\right] - \log\left[\frac{\Gamma\left(\sum_j \left(b'_{lj} + n_{lj}\right)\right)}{\Gamma\left(\sum_j \left(b_{lj} + n_{lj}\right)\right)}\right] \\
& - 2\log\left[\frac{\Gamma\left(b'_{kl}\right)}{\Gamma(b_{kl})}\right].
\end{aligned} \tag{B.3}$$

The change in the log-prior as a consequence of flipping a_{kl} to $1 - a_{kl}$ for the prior follows from its definition in Eq. (3):

$$\begin{aligned}
\Delta P_{kl} = & \log \Pr\left(\vec{\mathbf{A}}'|\mathcal{N}, \vec{\xi}\right) - \log \Pr\left(\vec{\mathbf{A}}|\mathcal{N}, \vec{\xi}\right) \\
= & (1 - 2a_{kl}) \log\left(\frac{p_{kl}}{1 - p_{kl}}\right).
\end{aligned}$$

Here the edge probability p_{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.

References

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

Behrens, T.E.J., Woolrich, M.W., Jenkinson, M., Johansen-Berg, H., Nunes, R.G., Clare, S., Matthews, P.M., Brady, J.M., Smith, S.M., 2003. Characterization and propagation of uncertainty in diffusion-weighted MR imaging. Magn. Reson. Med. 50, 1077–1088.

Behrens, T.E.J., Johansen-Berg, H., Jbabdi, S., Rushworth, M.F.S., Woolrich, M.W., 2007. Probabilistic diffusion tractography with multiple fibre orientations: what can we gain? Neuroimage 34, 144–155.

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

Catani, M., 2007. From hodology to function. Brain 130, 602–605.

Chung, H.W., Chou, M.C., Chen, C.Y., 2010. Principles and limitations of computational algorithms in clinical diffusion tensor MR tractography. Am. J. Neuroradiol. 32, 3–13.

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

Conturo, T.E., Lori, N.F., Cull, T.S., Akbudak, E., Snyder, A.Z., Shimony, J.S., McKinstry, R.C., Burton, H., Raichle, M.E., 1999. Tracking neuronal fiber pathways in the living human brain. Proc. Natl. Acad. Sci. U. S. A. 96, 10422–10427.

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

Dahl, J., Vandenberghe, L., Roychowdhury, V., 2008. Covariance selection for non-chordal graphs via chordal embedding. Optim. Method. Softw. 23, 501–520.

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

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

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

Erdős, P., Rényi, A., 1960. On the evolution of random graphs. Publ. Math. Inst. Hung. Acad. Sci. 5, 17–61.

Friedman, J., Hastie, T., Tibshirani, R., 2008. Sparse inverse covariance estimation with the graphical lasso. Biostatistics 9, 432–441.

Friman, O., Farnebäck, G., Westin, C.F., 2006. A Bayesian approach for stochastic white matter tractography. IEEE Trans. Med. Imaging 25, 965–978.

Gong, G., Rosa-Neto, P., Carbonell, F., Chen, Z.J., He, Y., Evans, A., 2009. Age- and gender-related differences in the cortical anatomical network. J. Neurosci. 29.

Greicius, M.D., Supekar, K., Menon, V., Dougherty, R.F., 2009. Resting-state functional connectivity reflects structural connectivity in the default mode network. Cereb. Cortex 19, 72–78.

Groves, A.R., Beckmann, C.F., Smith, S.M., Woolrich, M.W., 2011. Linked independent component analysis for multimodal data fusion. Neuroimage 54, 2198–2217.

Hagmann, P., Kurant, M., Gigandet, X., Thiran, P., Wedeen, V., Meuli, R., Thiran, J.P., 2007. Mapping human whole-brain structural networks with diffusion MRI. PLoS One 2, e597.

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

Honey, C.J., Sporns, O., Cammoun, L., Gigandet, X., Thiran, J.P., Meuli, R., Hagmann, P., 2009. Predicting human resting-state functional connectivity from structural connectivity. Proc. Natl. Acad. Sci. U. S. A. 106, 2035–2040.

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

Horwitz, B., Poeppel, D., 2002. How can EEG/MEG and fMRI/PET data be combined? Hum. Brain Mapp. 17, 1–3.

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

Jbabdi, S., Woolrich, M.W., Andersson, J.L.R., Behrens, T.E.J., 2007. A Bayesian framework for global tractography. Neuroimage 37, 116–129.

Kaden, E., Knösche, T.R., Anwander, A., 2007. Parametric spherical deconvolution: inferring anatomical connectivity using diffusion MR imaging. Neuroimage 37, 474–488.

Koch, M., Norris, D.G., Hund-Georgiadis, M., 2002. An investigation of functional and anatomical connectivity using magnetic resonance imaging. Neuroimage 16, 241–250.

Lv, J., Guo, L., Hu, X., Zhang, T., Li, K., Zhang, D., Yang, J., Liu, T., 2010. Fiber-centered analysis of brain connectivities using DTI and resting state fMRI data. Med. Image Comput. Assist. Interv. 13, 143–150.

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

Marrelec, G., Krainik, A., Duffau, H., Pélégrini-Issac, M., Lehéricy, S., Doyon, J., Benali, H., 2006. Partial correlation for functional brain interactivity investigation in functional MRI. Neuroimage 32, 228–237.

Minka, T.P., 2000. Estimating a Dirichlet Distribution. Technical Report. MIT.

Mukherjee, S., Speed, T.P., 2008. Network inference using informative priors. Proc. Natl. Acad. Sci. U. S. A. 105, 14313–14318.

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

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

Poser, B.A., Versluis, M.J., Hoogduin, J.M., Norris, D.G., 2006. BOLD contrast sensitivity enhancement and artifact reduction with multiecho EPI: parallel-acquired inhomogeneity-desensitized fMRI. Magn. Res. Med. 55, 1227–1235.

Robinson, E.C., Valstar, M., Hammers, A., Ericsson, A., Edwards, A.D., Rueckert, D., 2008. Multivariate statistical analysis of whole brain structural networks obtained using probabilistic tractography. Med. Image Compute. Comput. Assist. Interv. 11, 486–493.

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

Ryali, S., Chen, T., Supekar, K., Menon, V., 2011. Estimation of functional connectivity in fMRI data using stability selection-based sparse partial correlation with elastic net penalty. Neuroimage 59, 3852–3861.

Rykhlevskaia, E., Gratton, G., Fabiani, M., 2008. Combining structural and functional neuroimaging data for studying brain connectivity: a review. Psychophysiology 45, 173–187.

Schmidt, M., Fung, G., Rosales, R., 2007. Fast optimization methods for L1 regularization: a comparative study and two new approaches. Mach. Learn. ECML 2007, 286–297.

Shu, N., Liu, Y., Li, K., Duan, Y., Wang, J., Yu, C., Dong, H., Ye, J., He, Y., 2011. Diffusion tensor tractography reveals disrupted topological efficiency in white matter structural networks in multiple sclerosis. Cereb. Cortex 21.

Simpson, S.L., Hayasaka, S., Laurienti, P.J., 2011. Exponential Random Graph Modeling for complex brain networks. PLoS One 6, e20039.

Skudlarski, P., Jagannathan, K., Calhoun, V., Hampson, M., Skudlarska, B.A., Pearlson, G., 2008. Measuring brain connectivity: diffusion tensor imaging validates resting state temporal correlations. Neuroimage 43, 554–561.

Smith, S.M., 2002. Fast robust automated brain extraction. Hum. Brain Mapp. 17, 143–155.

Smith, S.M., Miller, K.L., Salimi-Khorshidi, G., Webster, M., Beckmann, C.F., Nichols, T.E., Ramsey, J.D., Woolrich, M.W., 2010. Network modelling methods for fMRI. Neuroimage 54, 875–891.

Sui, J., Pearlson, G., Caprihan, A., Adali, T., Kiehl, K.A., Liu, J., Yamamoto, J., Calhoun, V.D., 2011. Discriminating schizophrenia and bipolar disorder by fusing fMRI and DTI in a multimodal CCA+ joint ICA model. Neuroimage 57, 839–855.

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., Joliot, M., 2002. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. Neuroimage 15, 273–289.

Vaessen, M.J., Hofman, P.A.M., Tijssen, H.N., Aldenkamp, A.P., Jansen, J.F.A., Backes, W.H., 2010. The effect and reproducibility of different clinical DTI gradient sets on small world brain connectivity measures. Neuroimage 51, 1106–1116.

Varoquaux, G., Gramfort, A., Poline, J.B., Thirion, B., 2010. Brain covariance selection: better individual functional connectivity models using population prior. In: Zemel, R., Shawe-Taylor, J. (Eds.), Advances in Neural Information Processing Systems, Vancouver, Canada, pp. 2334–2342.

Whittaker, J., 1990. Graphical Models in Applied Multivariate Statistics, 1st edition. Wiley Publishing.

Xu, Z., Akella, R., 2010. Improving probabilistic information retrieval by modeling burstiness of words. Inf. Process. Manag. 46, 143–158.

Zalesky, A., Fornito, A., Harding, I.H., Cocchi, L., Yücel, M., Pantelis, C., Bullmore, E.T., 2010. Whole-brain anatomical networks: does the choice of nodes matter? Neuroimage 50, 970–983.

Zhang, Y., Brady, M., Smith, S., 2001. Segmentation of brain MR images through a hidden Markov random field model and the expectation maximization algorithm. IEEE Trans. Med. Imaging 20, 45–57.