Bayesian Inference of Whole-Brain Networks

M. Hinne[1,2,*], T. Heskes[2], and M.A.J. van Gerven[1]

[1]Donders Institute for Brain, Cognition and Behaviour: Centre for Cognition.
Radboud University, Nijmegen, The Netherlands
[2]Institute for Computing and Information Sciences. Radboud University,
Nijmegen, The Netherlands
[*]mhinne@cs.ru.nl

February 9, 2012

Abstract

In structural brain networks the connections of interest consist of white-matter fibre bundles between spatially segregated brain regions. The presence, location and orientation of these white matter tracts can be derived using diffusion MRI in combination with probabilistic tractography. Unfortunately, as of yet no approaches have been suggested that provide an undisputed way of inferring brain networks from tractography. In this paper, we provide a computational framework which we refer to as *Bayesian connectomics*. Rather than applying an arbitrary threshold to obtain a single network, we consider the posterior distribution of networks that are supported by the data, combined with an exponential random graph (ERGM) prior that captures a priori knowledge concerning the graph-theoretical properties of whole-brain networks. We show that, on simulated probabilistic tractography data, our approach is able to reconstruct whole-brain networks. In addition, our approach directly supports multi-model data fusion and group-level network inference.

1 Introduction

Human behaviour ultimately arises through the interactions between multiple brain regions that together form a network that can be characterized in terms of structural, functional and effective connectivity (Penny et al., 2006). The science of characterizing neural connectivity, that is, to elucidate the wiring diagram of the human brain, has come to be known as *connectomics* and is seen as one of the great challenges in neuroscience (Sporns et al., 2005).

Structural connectivity presupposes the presence 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. Furthermore, structural connectivity has been related to several diseases, such as Alzheimer's disease, attention deficit hyperactivity disorder and multiple sclerosis (He et al., 2008, 2009; Konrad and Eickhoff, 2010), and is therefore also of major importance in clinical neuroscience (Catani, 2007).

Currently, the only way to estimate structural connectivity of whole-brain networks in vivo is through the use of diffusion imaging; a variant of magnetic resonance imaging 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 (Chung et al., 2010; Zhang et al., 2005). There are however several problems associated with this deterministic

approach as it cannot easily deal with various ambiguous but common fiber configurations such as kissing, crossing and splaying fibre bundles (Jbabdi and Johansen-Berg, 2011).

As a solution to this problem, probabilistic tractography has been introduced (Behrens et al., 2007, 2003; Friman et al., 2006; Jbabdi et al., 2007). By repeating the streamlining process multiple times, probabilistic tractography ultimately produces for each voxel a measure of uncertainty about in which other voxel a hypothesized connection will terminate. A benefit of the probabilistic approach is that it can deal with ambiguous fiber configurations by taking streamlining uncertainty into account and produces probability estimates concerning the presence or absence of particular white-matter tracts.

Often, one is not interested so much in particular tracts but rather in whole-brain structural connectivity. Several approaches have been suggested to derive whole-brain networks from probabilistic tractography results (Chung et al., 2011; Hagmann et al., 2007; Iturria-Medina et al., 2008; Skudlarski et al., 2008). The resulting networks are all reported to exhibit certain graph-theoretical properties, such as a characteristic short path length, clustering, the existence of hubs and modular structure (Bassett et al., 2011; Bassett and Bullmore, 2006; Bullmore and Sporns, 2009; Gong et al., 2009; Sporns, 2006). Unfortunately, the inference of whole-brain networks from probabilistic tractography estimates remains somewhat ad hoc. Networks are often constructed by counting the number of streamlines that connect regions of interest and interpreting this as an undirected, weighted graph (e.g. (Zalesky et al., 2010)). Alternatively, a binary graph is obtained by simply adding an edge for each pair of regions that have one or more streamlines between them (e.g. (Bassett et al., 2011; Hagmann et al., 2007; Vaessen et al., 2010)). It is hard, however, to maintain a direct correspondence between connection strength and the probability distribution over tracts that is produced by probabilistic tractography (Jbabdi and Johansen-Berg, 2011).

Another important observation is that the mentioned approaches do not easily support the integration of probabilistic streamlining data with other datasets. This is a particularly relevant topic in contemporary neuroscience since it is assumed by multi-modal data fusion, where data acquired via different recording modalities is integrated to provide a coherent picture of brain function (Horwitz and Poeppel, 2002), and required by group-level inference, where the interest is in estimating a network that characterizes a particular population, e.g. when comparing patients with controls in a clinical setting (Simpson et al., 2011b).

In the following, we provide for the first time a computational framework for the estimation of whole-brain networks from probabilistic tractography outcomes. In our approach, which we refer to as *Bayesian connectomics*, we consider the posterior distribution of graphs that are supported by our data, instead of a generating a single network based on an arbitrary threshold. Our approach relies on defining a generative model for whole-brain networks which is inspired by recent work on network inference in systems biology (Mukherjee and Speed, 2008) and consists of two ingredients. First, a likelihood model based on a multivariate distribution which views the streamline distributions produced by probabilistic tractography as noisy data. Second, an exponential random graph prior which models prior knowledge concerning the graph-theoretical properties of whole-brain networks. The extension to multi-modal data fusion or group-level inference follows by defining the likelihood model as a product of multivariate distributions. In order to validate our methodology we make use of simulated data which allows us to compare the estimates produced by Bayesian connectomics with ground truth. We show that our approach is able to reconstruct whole-brain networks from simulated data as produced by probabilistic tractography.

2 Bayesian inference of whole-brain networks

In this section we derive our Bayesian approach to the inference of whole-brain networks, referred to as Bayesian connectomics. We start by defining the likelihood term for our generative model. Subsequently, an exponential random graph prior will be defined which can be used to capture the essential graph-theoretical properties of whole-brain networks. Finally, we derive a Markov chain Monte-Carlo algorithm to sample from the posterior distribution of whole-brain networks.

2.1 Likelihood model

Assume we want to infer the connectivity between K brain regions. That is, we want to find a graph $G \in \mathcal{G}$ where $\mathcal{G}$ is the family of undirected graphs. The graph $G = (V, E)$ consists of vertices $V = \{V_1, \dots, V_K\}$ and edges $E \subseteq V \times V$ where loops (V_i, V_i) for $i \in \{1, \dots, K\}$ are excluded from the edge set. We use $e_{ij} = 1$ when $(V_i, V_j) \in E$ and $e_{ij} = 0$ when $(V_i, V_j) \notin E$.

We start by considering one region and the possible targets in which a postulated tract may terminate. Probabilistic tractography will produce a distribution over target regions $\mathbf{n}_i = (n_{i1}, \dots, n_{iK})^T$ by drawing S streamlines, $N_i = \sum_{k=1}^{K} n_{ik} \leq S$ of them ending up in a target region. Let $n_{ii} = 0$ and assume that all streamlines are drawn independently. The probability of finding a particular distribution $\mathbf{n}_i$ is expressed as a multinomial distribution

$$P\left(\mathbf{n}_i \mid \mathbf{x}_i\right) \propto \prod_{j=1}^{K} x_{ij}^{n_{ij}} \;, \tag{1}$$

in which x_{ij} with $\sum_j x_{ij} = 1$ 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 tract between region i and region j. We model the distribution of streamlining probabilities using a Dirichlet distribution

$$P\left(\mathbf{x}_i \mid \mathbf{a}_i\right) \propto \prod_{j=1}^{K} x_{ij}^{a_{ij}-1} \tag{2}$$

where the

$$a_{ij} = e_{ij}\overline{a} + (1 - e_{ij})\underline{a} \tag{3}$$

can be interpreted as pseudo-counts that determine the probability of streamlining from region i to region j when an edge e_{ij} in the underlying graph G is either present ($\overline{a}$) or absent ($\underline{a}$). Because the Dirichlet prior with parameters $\mathbf{a}_i = (a_{i1}, \dots, a_{iK})$ is conjugate to the multinomial distribution, the posterior distribution of tract probabilities is again Dirichlet such that

$$P\left(\mathbf{x}_i \mid \mathbf{n}_i, \mathbf{a}_i\right) \propto P(\mathbf{n}_i \mid \mathbf{x}_i)P(\mathbf{x}_i \mid \mathbf{a}_i) \propto \prod_{j=1}^{K} x_{ij}^{n_{ij}+a_{ij}-1} \;. \tag{4}$$

Let $\mathbf{N} = (\mathbf{n}_1; \dots; \mathbf{n}_K)$ represent the probabilistic tractography data, $\mathbf{X} = (\mathbf{x}_1; \dots; \mathbf{x}_K)$ the streamlining probabilities, and $\mathbf{A} = (\mathbf{a}_1; \dots; \mathbf{a}_K)$ the parameters of the Dirichlet distribution, for all brain regions $1 \leq i \leq K$ combined. The likelihood of the graph G is then expressed as

$$\begin{aligned}
P\left(\mathbf{N} \mid G\right) &= \int P\left(\mathbf{N} \mid \mathbf{X}\right) P\left(\mathbf{X} \mid \mathbf{A}\right) d\mathbf{X} \\
&= \prod_i \left[\frac{N_i!}{\prod_j n_{ij}!} \frac{\Gamma\left(\sum_j a_{ij}\right)}{\Gamma\left(\sum_j (a_{ij} + n_{ij})\right)} \prod_j \frac{\Gamma\left(a_{ij} + n_{ij}\right)}{\Gamma\left(a_{ij}\right)} \right] \;,
\end{aligned} \tag{5}$$

where we have replaced $\mathbf{A}$ with G on the left-hand side to emphasize that we are interested in a probability distribution over a set of graphs. Note that $\mathbf{A}$ and G are easily exchanged through Eq. (3). The second step in Eq. (5) follows by recognizing $P\left(\mathbf{N} \mid G\right)$ as a product of compound Dirichlet-Multinomial distributions, also known as multivariate Pólya distributions (Minka, 2003). This leads to the log-likelihood

$$L \equiv \log P\left(\mathbf{N} \mid G\right) = \sum_i \left[\log \frac{N_i!}{\prod_j n_{ij}!} + \log \frac{\Gamma\left(\sum_j a_{ij}\right)}{\Gamma\left(\sum_j (a_{ij} + n_{ij})\right)} + \sum_j \log \frac{\Gamma\left(a_{ij} + n_{ij}\right)}{\Gamma\left(a_{ij}\right)} \right] \;. \tag{6}$$

3

2.2 Exponential random graph prior

We would like to define a prior such that, when sampling from the prior, we generate reasonable whole-brain networks. Here, reasonable is taken to mean that the networks should respect graph-theoretical properties that have been inferred by previous studies in connectomics. One way to construct a prior that has this property is using the theory of exponential random graph models (ERGMs) (Frank and Strauss, 1986; Holland and Leinhardt, 1981; Wasserman, 1996). ERGMs, developed in the context of social network analysis (Snijders et al., 2006; Snijders, 2002) and systems biology (Mukherjee and Speed, 2008), have recently been used to fit models of whole-brain networks (Simpson et al., 2011a,b). However, they have not been used before as priors in a generative model of whole-brain networks.

The exponential random graph prior is defined as a log-linear distribution

$$P(G) \propto \exp\left(\lambda \sum_i w_i f_i(G)\right) \ , \tag{7}$$

in which λ is a strength parameter and the parameters w_i tune the relative strengths of the individual concordance functions $f_i : \mathcal{G} \to \mathbb{R}$ that capture several graph-theoretical properties of whole-brain networks.

For illustrative purposes, we will use the prior in order to capture the small-world phenomenon that occurs so frequently in brain networks (Sporns, 2010). That is, we wish to generate networks that have short average path length and a large clustering coefficient. This can be achieved by starting with a graph whose neighboring vertices are connected and rewiring edges at random with a certain probability (Watts and Strogatz, 1998). The choice of functions f_i is restricted however by the Markov chain Monte Carlo sampling algorithm since, as we will note in Section 2.3, the change after each sampling step must be calculated efficiently. Unfortunately, this prevents us from using average path length as as a concordance function since the change in path length due to an edge mutation cannot be calculated locally. It requires an all-pairs shortest paths algorithm, for which the fastest – Dijkstra's algorithm – is known to have time complexity $O(|E|\cdot|V|+|V|^2 \log|V|)$ on sparse graphs (Dijkstra, 1959). Instead, as a surrogate model, we use a prior that prefers graphs which are characterized by a small number of short edges. Let d_{ij} be the Euclidian distance between vertices i and j, then the prior is defined as

$$P(G) \propto \exp\left(\beta f(G)\right) \tag{8}$$

with $\beta = \lambda \cdot w$ and concordance function $f(G) = -\sum_{i,j} e_{ij} d_{ij}$ which assumes that the edge length follows an exponential distribution. In Section 3, we will show that this prior prefers graphs which exhibit the small-world property.

2.3 Sampling from the posterior distribution

By using the likelihood and the prior in conjunction with Bayes' rule, we obtain the following posterior distribution over whole-brain networks:

$$P(G \mid \mathbf{N}) \propto P(\mathbf{N} \mid G) P(G) \ . \tag{9}$$

Since the posterior $P(G \mid \mathbf{N})$ cannot be calculated analytically, we use Metropolis Markov chain Monte Carlo (MCMC) sampling to approximate this distribution (Mukherjee and Speed, 2008). The acceptance of each sample G' in the sampling chain is determined by the ratio

$$\alpha = \frac{P(G' \mid \mathbf{N})}{P(G \mid \mathbf{N})} \ . \tag{10}$$

A proposed graph becomes a new sample with probability $\min(1, \alpha)$. The acceptance probability of a suggested sample can be calculated as

$$\log \alpha = \Delta L_{kl} + \lambda \sum_i w_i \Delta f_i(G) \ , \tag{11}$$

with ΔL_{kl} the change in log-likelihood after flipping edge e_{kl} and $\Delta f_i(G) = f_i(G') - f_i(G)$. This requires that we can efficiently update both the likelihood and the prior for new samples in the Markov chain.

The change in log-likelihood as a consequence of the flipping of an edge e_{kl} to $1 - e_{kl}$ is defined as

$$\Delta L_{kl} = \log P\left(\mathbf{N} \mid G'\right) - \log P\left(\mathbf{N} \mid G\right) \;, \tag{12}$$

with the sole difference that the parameter a_{kl} associated with G' and denoted by $\tilde{a}_{kl}$ is given by $\tilde{a}_{kl} = \tilde{a}_{lk} = (1 - e_{kl})\overline{a} + e_{kl}\underline{a}$ with e_{kl} the edge as defined in G. Plugging (6) into (12) yields

$$\begin{aligned} \Delta L_{kl} = & \log\left[\frac{\Gamma\left(\tilde{a}_{kl} + n_{kl}\right)}{\Gamma\left(a_{kl} + n_{kl}\right)}\right] + \log\left[\frac{\Gamma\left(\tilde{a}_{lk} + n_{lk}\right)}{\Gamma\left(a_{lk} + n_{lk}\right)}\right] + \log\left[\frac{\Gamma\left(\sum_j \tilde{a}_{kj}\right)}{\Gamma\left(\sum_j a_{kj}\right)}\right] + \log\left[\frac{\Gamma\left(\sum_j \tilde{a}_{lj}\right)}{\Gamma\left(\sum_j a_{lj}\right)}\right] \\ & - \log\left[\frac{\Gamma\left(\sum_j(\tilde{a}_{kj} + n_{kj})\right)}{\Gamma\left(\sum_j(a_{kj} + n_{kj})\right)}\right] - \log\left[\frac{\Gamma\left(\sum_j(\tilde{a}_{lj} + n_{lj})\right)}{\Gamma\left(\sum_j(a_{lj} + n_{lj})\right)}\right] - 2\log\left[\frac{\Gamma\left(\tilde{a}_{lk}\right)}{\Gamma\left(a_{lk}\right)}\right] \;. \end{aligned} \tag{13}$$

The change in the log prior as a consequence of the flipping of an edge e_{kl} to $1 - e_{kl}$ depends on the employed concordance function(s). The concordance function used in (8) factorizes over individual edges:

$$P(G) \propto \prod_{i,j} \exp\left[-\beta e_{ij} d_{ij}\right] \tag{14}$$

such that $\Delta f(G) = (2e_{kl} - 1)d_{kl}$.

All configurations of the $2^{N(N-1)/2}$ possible graphs have a probability greater than zero to be constructed thus guaranteeing that the Markov chain is irreducible. The collection of accepted samples $\{G^{(1)}, \ldots, G^{(T)}\}$ forms an approximation of the posterior $P\left(G \mid \mathbf{N}\right)$. When we desire to generate a whole-brain network, the maximum a posterior graph

$$G_{\mathrm{MAP}} = \operatorname*{argmax}_{G}\left\{\log P\left(\mathbf{N} \mid G\right) + \log P\left(G\right)\right\} \tag{15}$$

can be used. Alternatively, the samples can be used to estimate posterior probabilities of network features, such as the existence of a tract. Assuming the Markov chain has converged, the posterior probability of a single tract is given by

$$\mathbb{E}\left[e_{ij} | \mathbf{N}\right] = \frac{1}{T}\sum_{t=1}^{T} e_{ij}^{(t)} \;. \tag{16}$$

Other quantities may be estimated in a similar fashion.

2.4 Dealing with multiple datasets

As pointed out previously, probabilistic streamlining data need not be collected in isolation. It may be collected in conjunction with other data such as resting-state fMRI or MEG data or it may be collected for multiple subjects. We show that our approach quite naturally supports multi-modal data fusion and/or group-level inference.

Suppose we have acquired M datasets $\mathbf{Y}_1, \ldots, \mathbf{Y}_M$. The goal is to derive the graph G which best explains these data. The posterior for G will be of the form

$$P\left(G \mid \mathbf{Y}_1, \ldots, \mathbf{Y}_M\right) \propto P\left(\mathbf{Y}_1, \ldots, \mathbf{Y}_M \mid G\right) P\left(G\right) \;. \tag{17}$$

Conditional on G, we assume that the datasets are independent, such that we have the following factorized representation of the posterior:

$$P\left(G \mid \mathbf{Y}_1, \ldots, \mathbf{Y}_M\right) \propto \left[\prod_{i=1}^{M} P\left(\mathbf{Y}_i \mid G\right)\right] P\left(G\right) \;. \tag{18}$$

In case of multi-modal data fusion this requires that we are able to write down a likelihood model for each of the datasets $\mathbf{Y}_i$. In case of group-level inference, $\mathbf{Y}_i$ represents probabilistic streamlining data for subject i. Note that this is not the same as simply summing the streamlines, by virtue of Eq. (5). It follows from the factorization implied by (18) that, in order to run the MCMC algorithm at the group level, all that is required is to use the following term for the change in log-likelihood when computing the acceptance probability:

$$\Delta L_{kl} = \sum_{i=1}^{M} \left[\log P \left(\mathbf{N}_i \mid G' \right) - \log P \left(\mathbf{N}_i \mid G \right) \right]. \tag{19}$$

3 Simulation study

In our simulation study, we used the Watts & Strogatz model (Watts and Strogatz, 1998) in order to generate a small-world graph G consisting of 20 edges and 20 vertices, laid out in a circular arrangement that defines the physical distance between the nodes. For this ground-truth network, we generated 'probabilistic streamlining' data as follows. For each node i a streamlining probability vector $\mathbf{x}$ was constructed by drawing from a Dirichlet distribution with parameters $\mathbf{a}_i = (a_{i1}, \ldots, a_{iN})$. Parameters were set to zero except for those a_{ij} for which $(i,j) \in E$ (the true tracts) as well as for a randomly selected a_{ik} for each $1 \leq i \leq 20$ (representing a false positive due to kissing, crossing or splaying fiber bundles). These parameters were all set equal to 20. Subsequently, for each node i the streamlining probability vectors $\mathbf{x}_i$ were used to generate 1000 streamlines drawn from a multinomial distribution. Figure 1 shows G together with the generated probabilistic streamlining data S.
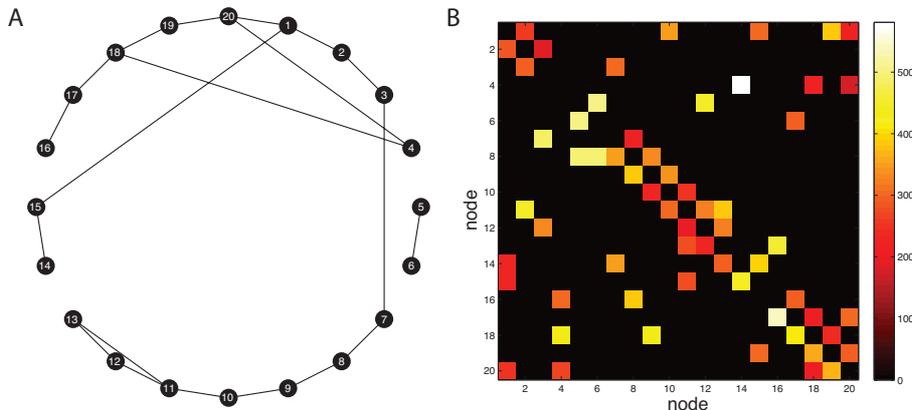

Figure 1: Simulated data. A. Small-world graph G. B. Probabilistic streamlining data generated for graph G indicating the number of streamlines drawn between nodes i and j.

Next, we wanted to determine whether the ground-truth network could be retrieved using Bayesian connectomics. That is, can we reconstruct G from probabilistic streamlining data S using our MCMC algorithm, does the exploitation of prior information help in retrieving the original network, and does it outperform a conventional thresholding approach? We ran the MCMC algorithm using five parallel chains, a burn-in of 2000 steps and an effective sample size of 8000. For the hyperparameters, we used $\bar{a} = 20$ and $\underline{a} = 15$ to indicate that absent edges may still have a substantial probability of being streamlined. For our prior, we used $\beta = 2$ as this generated networks whose number of edges agreed with what one expects if the underlying network is generated by the Watts & Strogatz model.

Figure 2 shows the true connectivity as well as the reconstruction results for a conventional thresholding approach (containing those edges for which the number of streamlines exceeds zero), the maximum likelihood solution, which is obtained by running the MCMC algorithm with a flat
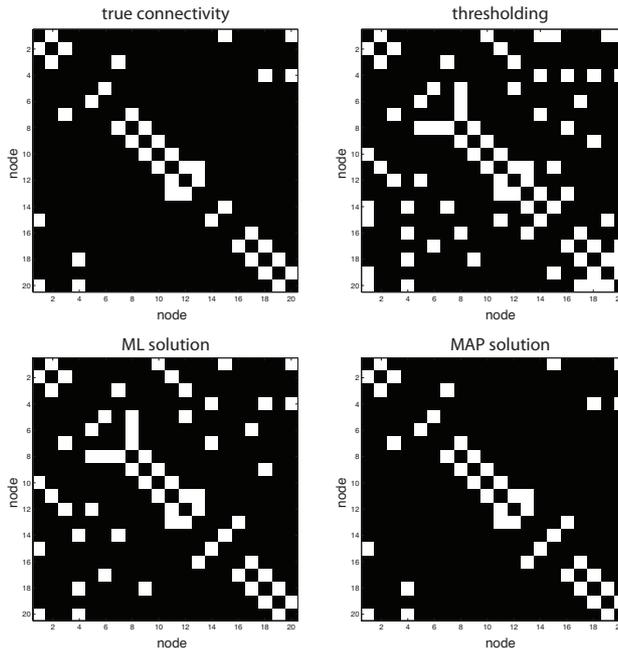
Figure 2: Comparison between the true connectivity and the connectivity obtained with a thresholding approach, a sample from the prior, the maximum likelihood (ML) solution and the maximum a posteriori (MAP) solution afforded by Bayesian connectomics.

prior, and the maximum a posteriori solution, which is obtained by running the MCMC algorithm with an informed prior. The latter is the procedure we refer to as Bayesian connectomics. Clearly, both the thresholding approach and the maximum likelihood solution suffer from the false positives introduced during the streamlining process. The MAP solution, in contrast, is able to retrieve the true connectivity almost perfectly well. Note that this cannot be explained by the use of the prior alone since the prior assumes that the empty graph is the most probable graph and samples from the prior are not likely to recover the long-range connections.

4 Discussion

Diffusion imaging promises to reveal the the wiring diagram of the human brain in vivo as achieved by probabilistic tractography. So far, the generation of whole-brain networks from probabilistic tractography data has relied on heuristic thresholding methods. In this paper we propose Bayesian connectomics as a principled framework that may replace these thresholding methods. The framework allows for the exploitation of prior knowledge when deriving the true connectivity based on inherently noisy probabilistic streamlining data. An additional advantage of Bayesian connectomics is that it easily allows dealing with multiple datasets as is the case with multi-modal data fusion or group-level inference. We have shown that our approach outperforms thresholding in retrieving the true connectivity underlying simulated probabilistic streamlining data.

In the simulation, we made a number of assumptions. First, we assumed that errors during probabilistic streamlining arise through the existence of a restricted number of false positives which arise due to the existence of a few kissing, crossing and splaying fibers. It would be interesting to examine how our approach behaves under other, possibly more realistic, assumptions about how probabilistic streamlining data is generated. Furthermore, we assumed that the hyperparameters $\overline{a}$, $\underline{a}$ and β were known. The hyperparameters of the Dirichlet distribution are unknown in practice, but these could be chosen based on the distribution of the number of streamlines that are generated during the streamlining process. The parameter β, which determined the influence of the prior

was chosen such that samples from the prior corresponded to the small-world networks we were interested in. In order to construct priors which work well in practice, various other concordance functions can be employed, whose parameters can be estimated on existing data using maximum likelihood approaches (Hunter et al., 2008; Simpson et al., 2011a).

In this paper, we have laid the foundations for a Bayesian approach to connectivity analysis. What remains is to validate the approach not only on simulated data but also on empirical data. This not only requires the reconstruction of whole-brain networks from diffusion imaging data, but also an assessment of their validity on independent data. One may think here of determining how well reconstructed networks correspond to networks estimated from resting-state fMRI or from tracer data (Jbabdi and Johansen-Berg, 2011). Eventually, we would like to use our approach to determine how differences between (groups of) people can be understood in terms of variations in their underlying wiring diagrams.

References

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

Bassett, D. and Bullmore, E. (2006). Small-world brain networks. *Neuroscientist*, 12(6):512–523.

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

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

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

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

Chung, H.-W., Chou, M.-C., and Chen, C.-Y. (2010). Principles and Limitations of Computational Algorithms in Clinical Diffusion Tensor MR Tractography. *A. J. Neuroradiol.*, 32:3–13.

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

Dijkstra, E. (1959). A note on two problems in connexion with graphs. *Numer. Math.*, 1:269–271.

Frank, O. and Strauss, D. (1986). Markov graphs. *J. Amer. Statistical Assoc.*, 81(395):832–842.

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

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

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

He, Y., Chen, Z., and Evans, A. (2008). Structural insights into aberrant topological patterns of large-scale cortical networks in Alzheimer's disease. *J. Neurosci*, 28(18):4756–4766.

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

Holland, P. and Leinhardt, S. (1981). An exponential family of probability distributions for directed graphs. *J. Amer. Stat. Assoc.*, 76(373):33–50.

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

Hunter, D., Goodreau, S., and Handcock, M. (2008). Goodness of fit of social network models. *J Am. Stat. Assoc*, 103(481):248–258.

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

Jbabdi, S. and Johansen-Berg, H. (2011). Tractography: Where Do We Go from Here? *Brain Connect.*, 1(3):169–183.

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

Konrad, K. and Eickhoff, S. (2010). Is the ADHD brain wired differently? A review on structural and functional connectivity in attention deficit hyperactivity disorder. *Hum. Brain Mapp.*, 31(6):904–916.

Minka, T. (2003). Estimating a Dirichlet distribution. Technical report, M.I.T.

Mukherjee, S. and Speed, T. (2008). Network inference using informative priors. *Proc. Natl. Acad. Sci.*, 105(38):14313–8.

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

Simpson, S., Hayasaka, S., and Laurienti, P. (2011a). Exponential Random Graph Modeling for Complex Brain Networks. *PLoS ONE*, 6(5):–.

Simpson, S., Moussa, M., and Laurienti, P. (2011b). An exponential random graph modeling approach to creating group-based representative whole-brain connectivity networks. *arXiv*, stat.AP.

Skudlarski, P., Jagannathan, K., Calhoun, V., Hampson, M., Skudlarska, B., and Pearlson, G. (2008). Measuring brain connectivity: diffusion tensor imaging validates resting state temporal correlations. *NeuroImage*, 43(3):554–561.

Snijders, T., Pattison, P., Robins, G., and Handcock, M. (2006). New Specifications for Exponential Random Graph Models. *Sociol. Methodol.*, 36(1):99–153.

Snijders, T. A. B. (2002). Markov chain Monte Carlo estimation of exponential random graph models. *J. of Soc. Struct.*, 3.

Sporns, O. (2006). Small-world connectivity, motif composition, and complexity of fractal neuronal connections. *Biosystems*, 85(1):55–64.

Sporns, O. (2010). *Networks of the Brain*. M.I.T. Press.

Sporns, O., Tononi, G., and Kötter, R. (2005). The human connectome: A structural description of the human brain. *PLoS Comput. Biol.*, 1(4):e42.

Vaessen, M., Hofman, P., Tijssen, H., Aldenkamp, A., Jansen, J., and Backes, W. (2010). The effect and reproducibility of different clinical DTI gradient sets on small world brain connectivity measures. *NeuroImage*, 51(3):1106–1116.

Wasserman, S. (1996). Logit models and logistic regressions for social networks: I. An introduction to Markov graphs and p*. *Psychometrika*, 61(3):401–425.

Watts, D. J. and Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. *Nature*, 393(6684):440–442.

Zalesky, A., Fornito, A., Harding, I., Cocchi, L., Yücel, M., Pantelis, C., and Bullmore, E. (2010). Whole-brain anatomical networks: does the choice of nodes matter? *NeuroImage*, 50(3):970–983.

Zhang, J., Ji, H., Kang, N., and Cao, N. (2005). Fiber Tractography in Diffusion Tensor Magnetic Resonance Imaging: A Survey and Beyond. Technical report, Dept. of Comp. Sci., Uni. of Kentucky.