

RESEARCH ARTICLE

# Probabilistic Clustering of the Human Connectome Identifies Communities and Hubs

Max Hinne<sup>1,2\*</sup>, Matthias Ekman<sup>1</sup>, Ronald J. Janssen<sup>1</sup>, Tom Heskes<sup>2</sup>, Marcel A. J. van Gerven<sup>1</sup>

**1** Radboud University Nijmegen, Donders Institute for Brain, Cognition and Behaviour, Nijmegen, The Netherlands, **2** Radboud University Nijmegen, Institute for Computing and Information Sciences, Nijmegen, The Netherlands

\* [mhinne@cs.ru.nl](mailto:mhinne@cs.ru.nl)



**OPEN ACCESS**

**Citation:** Hinne M, Ekman M, Janssen RJ, Heskes T, van Gerven MAJ (2015) Probabilistic Clustering of the Human Connectome Identifies Communities and Hubs. PLoS ONE 10(1): e0117179. doi:10.1371/journal.pone.0117179

**Academic Editor:** Gaolang Gong, Beijing Normal University, CHINA

**Received:** September 16, 2014

**Accepted:** December 14, 2014

**Published:** January 30, 2015

**Copyright:** © 2015 Hinne et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All data are available from the DANS database under persistent identifier urn:nbn:nl:ui:13-j9q9-o2.

**Funding:** The research was funded by the BrainGain Smart Mix Programme of the Netherlands Ministry of Economic Affairs and the Netherlands Ministry of Education, Culture and Science (<http://www.senternovem.nl/smartmix>). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

## Abstract

A fundamental assumption in neuroscience is that brain function is constrained by its structural properties. This motivates the idea that the brain can be parcellated into functionally coherent regions based on anatomical connectivity patterns that capture how different areas are interconnected. Several studies have successfully implemented this idea in humans using diffusion weighted MRI, allowing parcellation to be conducted in vivo. Two distinct approaches to connectivity-based parcellation can be identified. The first uses the connection profiles of brain regions as a feature vector, and groups brain regions with similar connection profiles together. Alternatively, one may adopt a network perspective that aims to identify clusters of brain regions that show dense within-cluster and sparse between-cluster connectivity. In this paper, we introduce a probabilistic model for connectivity-based parcellation that unifies both approaches. Using the model we are able to obtain a parcellation of the human brain whose clusters may adhere to either interpretation. We find that parts of the connectome consistently cluster as densely connected components, while other parts consistently result in clusters with similar connections. Interestingly, the densely connected components consist predominantly of major cortical areas, while the clusters with similar connection profiles consist of regions that have previously been identified as the ‘rich club’; regions known for their integrative role in connectivity. Furthermore, the probabilistic model allows quantification of the uncertainty in cluster assignments. We show that, while most clusters are clearly delineated, some regions are more difficult to assign. These results indicate that care should be taken when interpreting connectivity-based parcellations obtained using alternative deterministic procedures.

## Introduction

The brain can be described as a vast network of interconnected neurons. The acquisition and subsequent analysis of this network has spawned a discipline known as *connectomics* [1, 2].

Part of connectomics focusses on *structural connectivity*, which concerns the layout of physical tracts consisting of axonal fibers. At a macroscopic scale, structural connectivity is defined in terms of brain regions that consist of neuronal populations which are inter-connected via white matter fiber bundles. As neuronal activity is constrained by neuroanatomy, correctly identifying these structural connections aids in understanding how spatially remote regions of the brain cooperate [3, 4]. Both functional and structural connectivity have been shown to be relevant in clinical applications, for instance by characterizing the connectomes that correspond to neurological and/or psychological disorders [5–7]. Similarly, connectivity studies have become a useful aid in understanding cognition by elucidating which networks are related to particular functions [8–11].

An important application of structural connectivity is the delineation of functionally specialized clusters of brain regions based on their structural connectivity patterns. Importantly, connectivity-based parcellation can be obtained *in vivo*, contrasting it with other methods such as histological analysis of cytoarchitecture [12, 13]. Other non-invasive methods that delineate structural cortical boundaries exist, based on identification of major anatomical landmarks, but these approaches have been shown to be susceptible to large inter-subject variability [14]. Intuitively, a cluster consists of a set of regions that are more similar to other regions in the cluster than to regions outside of it, but this idea can be operationalized in different ways.

Under the first interpretation, which we refer to as *profile-based clustering*, regions may be clustered based on similarity of their connectivity profiles (e.g. [15–19], or see [20] for a literature review). In other words, it assumes that regions in the same cluster connect with the same areas. Several studies have used this approach to parcellate regions of interest such as the frontal pole [21], posteromedial cortex [22], occipital lobes [23], cingulate cortex [24] and thalamus [25–27]. Note that in this interpretation of connectivity-based clusters, regions within a cluster are not necessarily mutually connected. Implicitly, this approach aims to find regions that similarly integrate information from other parts of the brain.

Under the second interpretation, which we refer to as *community-based clustering*, parcellations are taken to consist of densely connected clusters that are only sparsely connected to regions outside the cluster. This approach is typically used for whole-brain parcellation [28–32]. Here, one implicitly assumes that clusters are structurally (and thus, indirectly, functionally) specialized and mostly interact with the other regions in their respective cluster.

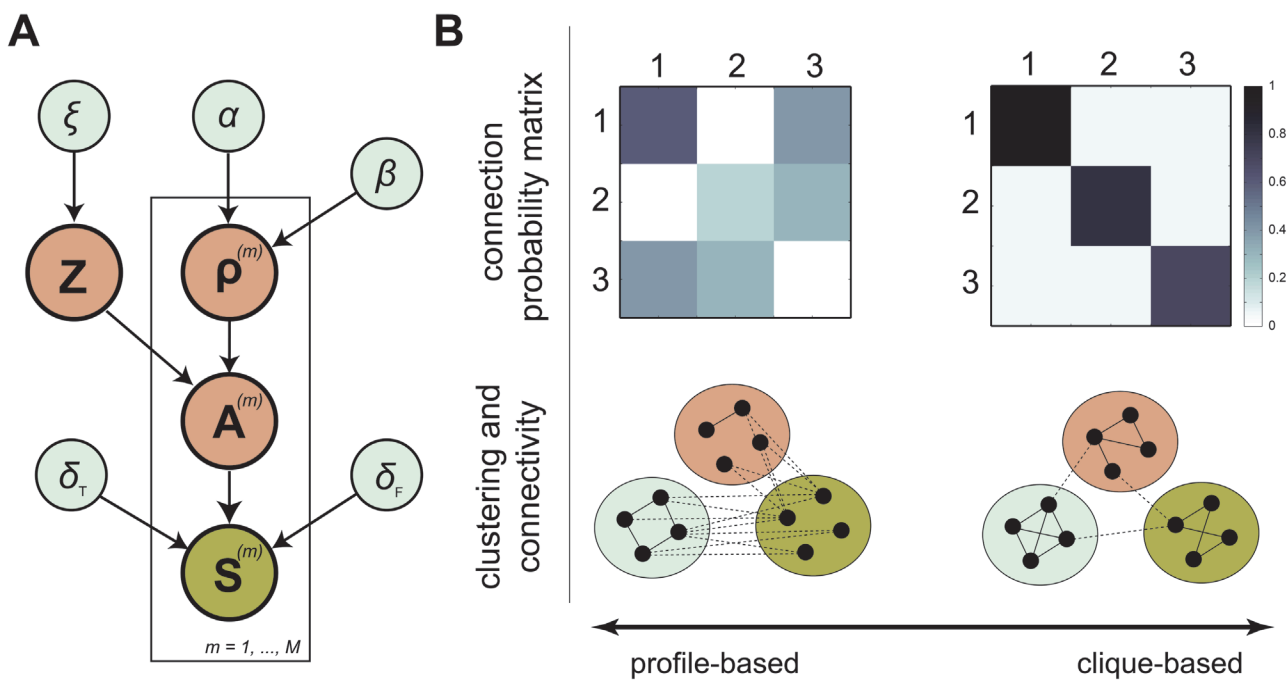
Either operationalization of connectivity-based clustering is applicable to the human brain. As a consequence, choosing either perspective may hinder finding meaningful clusters that only adhere to the other definition of a cluster. Furthermore, clustering behavior may not be uniform across the brain, and instead may be a mixture of profile-based and community-based clusters. In this study we introduce a probabilistic model that is able to parcellate structural connectivity and is sufficiently flexible to incorporate the different cluster interpretations. The model reveals community-based clusters in the human connectome, but also a set of nodes that are assigned to small profile-based clusters that function as hubs. These small clusters all contain regions that have been labeled the ‘rich club’ [33]. In addition, we find that the whole-brain parcellations which the model provides, are up to par with algorithms that have previously been used to parcellate connectivity. Moreover, in contrast to parcellations obtained with the other approaches, parcellations estimated by the probabilistic model can be obtained without the need to choose the number of clusters in advance. Finally, because the model is probabilistic, it is able to explicitly represent the uncertainty in the obtained parcellations. These visualizations show that the cluster assignment for particular regions is uncertain, indicating that care should be taken when interpreting connectivity-based parcellations obtained using alternative deterministic procedures.

### Probabilistic model for connectivity-based clustering

First, we describe a probabilistic model for clustering based on structural connectivity. Next, we extend this model to allow for direct estimation of cluster structure from probabilistic tractography data. The complete model is shown in Fig. 1A.

#### Stochastic block model

Let us assume that each cluster has a distinct connectivity profile. In other words, each cluster has a set of parameters that govern its connectivity behavior. This is known as a stochastic block model (SBM) [34]. Within a cluster, all nodes (i.e. brain regions) use the same connectivity parameters. As a consequence, regions within a cluster are stochastically exchangeable. Formally, each node  $1, \dots, N$  is assigned a cluster label, using the unobserved (latent) cluster assignment variable  $Z_{nk} = 1$  if node  $n$  is in cluster  $k$ , and 0 otherwise. Each node is assigned to exactly one cluster. Structural connectivity is described by a symmetric and binary adjacency matrix  $A$ . Because of the stochastic exchangeability assumption, the probability of a connection  $a_{ij}$  depends solely on the clusters to which nodes  $i$  and  $j$  are assigned. The cluster connection probabilities between clusters  $a$  and  $b$  are given by  $\rho_{ab}$ . The set of cluster connection probabilities is collected in the connection probability matrix  $\rho$ . In practice,  $\rho$  is unknown so we use a prior on  $\rho$  to reflect our assumptions about it. Here, this is a Beta prior that depends on two hyperparameters  $\alpha$  and  $\beta$ , which model the probability of a connection or non-connection between different clusters, respectively. We assume an uninformative prior with  $\alpha = \beta = 1$ . After observing the data, the posterior expectation for  $\rho$  will reflect cluster connection probabilities and therefore informs about the contributions of profile-based and community-based clusters



**Fig 1. Probabilistic model for connectivity-based parcellation.** **A.** The streamline infinite relational model combines a forward model for streamline data  $S$  with an infinite relational model that allows estimation of the cluster assignment matrix  $Z$  as well as the connection probability matrix  $\rho$ . Hyperparameters  $\{\xi, \alpha, \beta, \delta_T, \delta_F\}$  complete the model. **B.** The probabilistic model supports both profile-based clustering as well as community-based clustering. The top row shows simulated connection probability matrices that correspond to profile-based clustering (left) and community-based clustering (right). Example networks that correspond to these probabilities are shown at the bottom.

doi:10.1371/journal.pone.0117179.g001

[35]. As an example of how block models may represent different connectivity patterns, Fig. 1B shows the posterior expectations for the connection probability matrix using a toy network of twelve nodes distributed evenly over three clusters. Here, the network on the left shows how profile-based clusters may be captured by the off-diagonal weights of  $\rho$ , while the network on the right shows how strong weights on the diagonal of  $\rho$  generate a traditional community-based network.

SBM have seen widespread application in literature, ranging from discovery of roles in social networks [34] to identification of protein-protein interactions [36, 37]. In addition, several model variants have been introduced, such as approaches that deal with overlapping clusters [38] or SBM tailored to weighted networks [39]. However, these approaches assume that the number of clusters  $K$  is known a priori, which is frequently not the case with empirical data. A nonparametric extension was introduced to learn the number of clusters from data as well [40]. This is achieved by placing a prior distribution on the cluster assignment matrix  $\mathbf{Z}$ . Doing so allows the model to accommodate a (potentially) infinitely large number of clusters, rather than needing to specify the number of clusters  $K$  beforehand. Specifically, we draw  $\mathbf{Z}$  from a Chinese restaurant process (CRP) [41]. This distribution over partitions can be used to generate samples from in the following way. Consider nodes that are assigned to clusters one by one, as customers entering a restaurant and choosing a table to sit at. Each customer is assigned to a non-empty table  $k$  with probability  $\frac{m_k}{N-1+\xi}$ , with  $m_k$  the number of customers already assigned to table  $k$  and with probability  $\frac{\xi}{N-1+\xi}$  to an empty (new) table. Its concentration parameter  $\xi$  determines how likely it is for a customer to sit at an empty table, which affects the total number of tables with customers, i.e. the number of clusters. Using the CRP, the generative model is then given by

$$\begin{aligned} \mathbf{Z} \mid \xi &\sim \text{CRP}(\xi) \\ \rho_{ab} \mid \alpha, \beta &\sim \text{Beta}(\alpha, \beta) \\ a_{ij} \mid \rho_{ab}, \mathbf{Z} &\sim \text{Bernoulli}(\mathbf{z}_i \rho \mathbf{z}_j^T), \end{aligned} \tag{1}$$

where we introduce the notation  $\mathbf{m}_i$  to indicate the  $i$ th row of a matrix  $\mathbf{M}$ . The model is known as the infinite relational model (IRM) [35, 40, 42, 43]. The infinite relational model is easily generalized to encompass multiple (conditionally independent) subjects that share a parcellation [42] by changing the model into

$$\begin{aligned} \rho_{ab}^{(m)} \mid \alpha, \beta &\sim \text{Beta}(\alpha, \beta) \\ a_{ij}^{(m)} \mid \rho_{ab}^{(m)}, \mathbf{Z} &\sim \text{Bernoulli}(\mathbf{z}_i \rho^{(m)} \mathbf{z}_j^T) \end{aligned} \tag{2}$$

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

### Forward model for streamlines

To infer cluster assignments, structural connectivity data must be provided in the form of a binary adjacency matrix  $\mathbf{A}$ . This matrix can be obtained using probabilistic tractography [44]. Specifically, probabilistic tractography proceeds by drawing streamlines between brain regions based on local estimates of anisotropic diffusion. These streamlines are collected in a streamline count matrix  $\mathbf{S}$ , reflecting the number of streamlines between pairs of brain regions. Thresholding of the streamline count matrix produces a binary matrix that reflects structural connectivity. However, as a threshold results in a point estimate, we make use of a probabilistic model

that describes how a structural network generates the distributions of probabilistic streamlines that are obtained through tractography [45]. Using this forward model, a streamline threshold is no longer required.

Ideally, probabilistic streamlines show a distribution that perfectly reflects the underlying structural connectivity. However, tractography is prone to noise and errors, in particular in the presence of kissing, splaying and crossing fibres [46]. Hence, we distinguish between true connections along which we expect to observe streamlines, and false connections that may occasionally display streamlines, but do not correspond to actual anatomical pathways. The probability of a streamline between a pair of regions is represented by the matrix  $\mathbf{X}$ . Formally, the streamline probability vector  $\mathbf{x}_i$  for a particular region  $i$  is determined by a Dirichlet distribution with parameter  $\delta_T$  for true connections and  $\delta_F$  for false connections. According to these probabilities, streamlines are distributed amongst the target regions using a multinomial distribution. By integrating out the streamline probability vectors, we obtain the following forward model:

$$\mathbf{s}_i \mid \mathbf{a}_i, \delta_T, \delta_F \sim \text{DirMul}(\delta_T \mathbf{a}_i + \delta_F (\mathbf{1}_n - \mathbf{a}_i)) \tag{3}$$

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

To estimate connectivity from streamline data, (3) is supplied with a prior on  $\mathbf{A}$ . In the most straightforward case, this prior is uniform, i.e.  $P(\mathbf{A}) \propto 1$  [45]. Alternatively, this uninformative prior may be replaced by the IRM (see Fig. 1A), expressing our assumption of clustering in the network. The interpretation of the integrated model is that most observed streamlines indicate a structural connection (although there is some noise in the tractography process), and regions in the same cluster share their connectivity preferences (although some exceptions are allowed). We refer to the combined model that operates on streamline data as sIRM. Details of the approximate inference algorithm used to compute posterior estimates of interest are provided in S11.

### Visualizing clustering uncertainty

In addition to the display of individual parcellations, such as the maximum a posteriori (MAP) estimate, the proposed method may be used to visualize uncertainty about a parcellation. However, as the number of clusters may vary within a clustering distribution, and cluster labels are arbitrary for each sample, such a visualization is not trivial. A representation that allows different samples to be compared is the cluster co-assignment matrix  $\mathbf{M} = \mathbf{Z}^T \mathbf{Z}$ . The expectation of  $\mathbf{M}$ , i.e.  $\mathbb{E}[\mathbf{M}] = \frac{1}{S} \sum_{s=1}^S \mathbf{M}_s$ , with  $S$  the number of obtained samples (see S11), describes the posterior probability that two regions are assigned to the same cluster. The probability of co-assignment is used to color a region  $i$  with a weighted color coding given by

$$c_i = \sum_j m_{ij} \hat{c}_j / \sum_j m_{ij}, \text{ with } \hat{c}_j \text{ the color representation in the MAP estimate for region } j. \text{ For ex-}$$

ample, suppose that the MAP estimate consists of two clusters, one colored red and one colored yellow. If a region  $i$  is co-assigned to a region in the red cluster in half the samples, and to a region in the yellow cluster in the other half of the samples, it will be colored orange.

### Evaluating parcellation quality

Since a clustering ground truth is unavailable, we used the reproducibility of the parcellations as an indicator of parcellation quality [14, 47, 48]. We quantify the reproducibility as the

similarity between the parcellations of different participants (or groups of participants), expressed using adjusted mutual information (AMI) [49]. The AMI measure differs from the more traditional normalized mutual information measure in that it compensates for possible bias as a result of different numbers of clusters per parcellation. The measure is defined as

$$S(\mathbf{Z}_1, \mathbf{Z}_2) = \frac{\text{MI}(\mathbf{Z}_1, \mathbf{Z}_2) - \mathbb{E}[\text{MI}(\mathbf{Z}_1, \mathbf{Z}_2)]}{\max(H(\mathbf{Z}_1), H(\mathbf{Z}_2)) - \mathbb{E}[\text{MI}(\mathbf{Z}_1, \mathbf{Z}_2)]}, \quad (4)$$

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

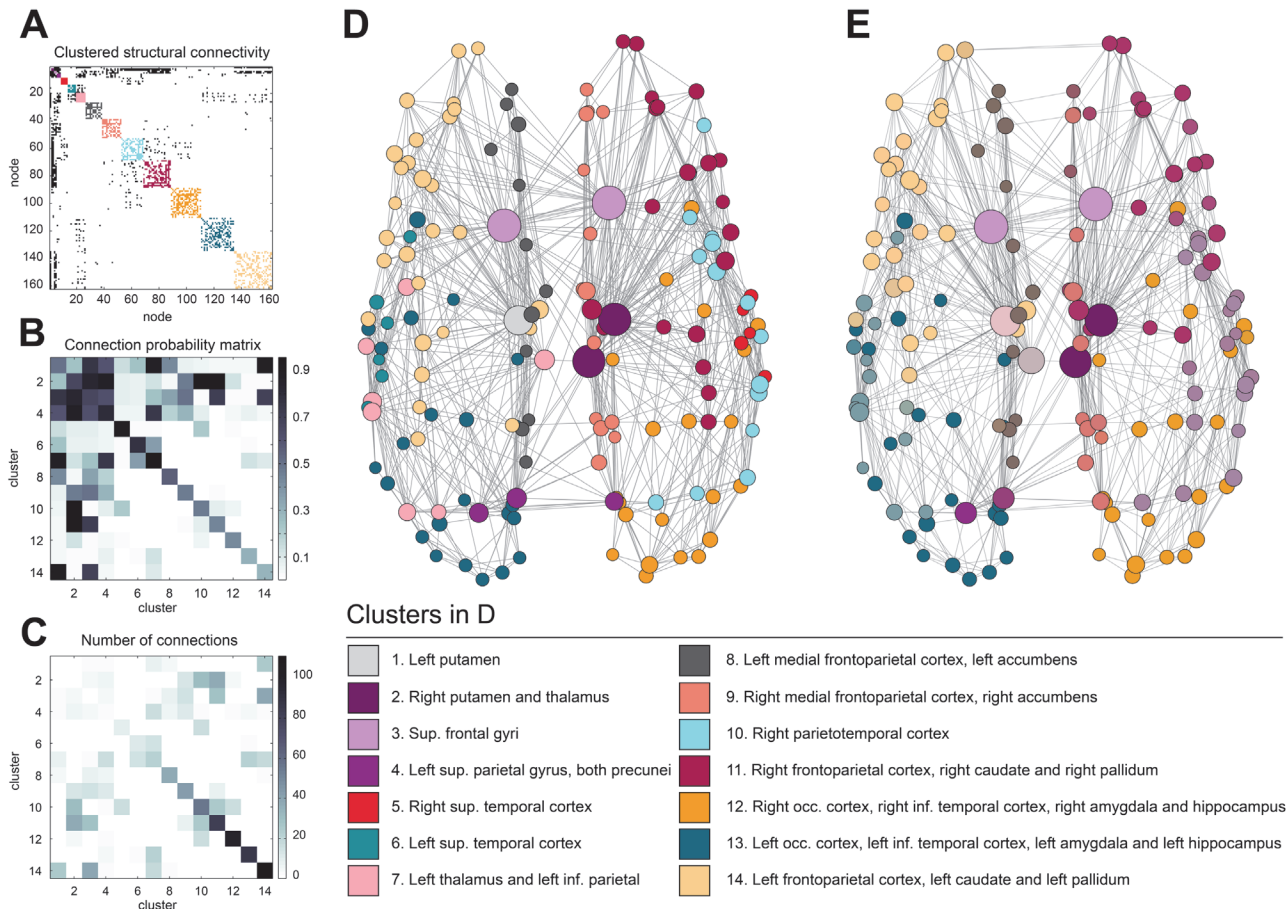
For comparison, we also obtain parcellations using  $K$ -means, the canonical algorithm for profile-based clustering, and the Infomap algorithm [50], which relies on community-based clustering (see S13). Both methods have been used before in the context of brain parcellation [18, 32]. Each of the comparison algorithms was applied to the empirical streamline distributions. For  $K$ -means and Infomap the streamline matrices were made symmetric, i.e.  $\mathbf{S}' = \mathbf{S} + \mathbf{S}^T$ . The number of clusters for  $K$ -means and Infomap is fixed to be the same as for the MAP estimate of the sIRM approach. To apply the bIRM, we first obtained the MAP estimate of connectivity with a flat prior, i.e.  $P(\mathbf{A}) \propto 1$ , which was subsequently clustered using the IRM approach. For the group-level analysis, a group-level streamline matrix was created by summing the streamline counts of all ten participants per split. This matrix was provided as input for  $K$ -means and Infomap. The bIRM method uses the ten individual MAP estimates of the forward model as input, while the sIRM method uses the ten streamline matrices.

## Results

The sIRM was used to obtain posterior distributions as well as MAP estimates for twenty participants, using the parameter settings described in S12. On average, the MAP clusterings of these participants consisted of 12.10 (SD 1.29) clusters. The posterior distributions of the numbers of clusters are centered tightly around the MAP estimates, as evidenced by a mean range of the 95% credible intervals of only 0.85 (SD 0.81) (i.e. most of the samples of the approximated distribution have the exact same number of clusters). To analyze the behavior of group-level parcellations, we created 20 random splits of the set of subjects and obtained the MAP parcellation for all of the 40 halves. These 40 parcellations consisted on average of 15.03 (SD 0.83) clusters. At both the individual and the group-level we find that all identified clusters are spatially contiguous, with the exception that the superior frontal gyrus and the precuneus are sometimes assigned to the same cluster as their functional homologue in the contralateral hemisphere.

As an example, the MAP estimate for one participant is shown in Fig. 2A–D and Fig. 3. The other MAP estimates are shown in S2 Fig. and S3 Fig. for the individual participants and the 40 halves, respectively. Fig. 2A shows the connectivity matrix  $\mathbf{A}$ . Within-cluster connections are colored with the color of their respective clusters and between-cluster connections are colored black. Fig. 2B shows the probability of a connection between pairs of clusters, as represented by the connection probability matrix  $\rho$ . The number of connections between pairs of clusters,  $\mathbf{ZAZ}^T$ , is shown in Fig. 2C. Note that the number of possible connections grows with  $m_k^2$ , so while the amount of connection increases, the cluster connection probability may decrease. A visualization of the layout of the connectivity  $\mathbf{A}$  and clustering  $\mathbf{Z}$  of this MAP estimate is shown in Fig. 2D.

Fig. 2E shows the interpolated clustering colors based on the approximated posterior distribution of  $\mathbf{Z}$  for this participant. The figure reveals that although the expectation is not very different from the MAP estimate, there remains some room for uncertainty. For instance, the right parietotemporal cortex and the right superior temporal cortex are seen as separate



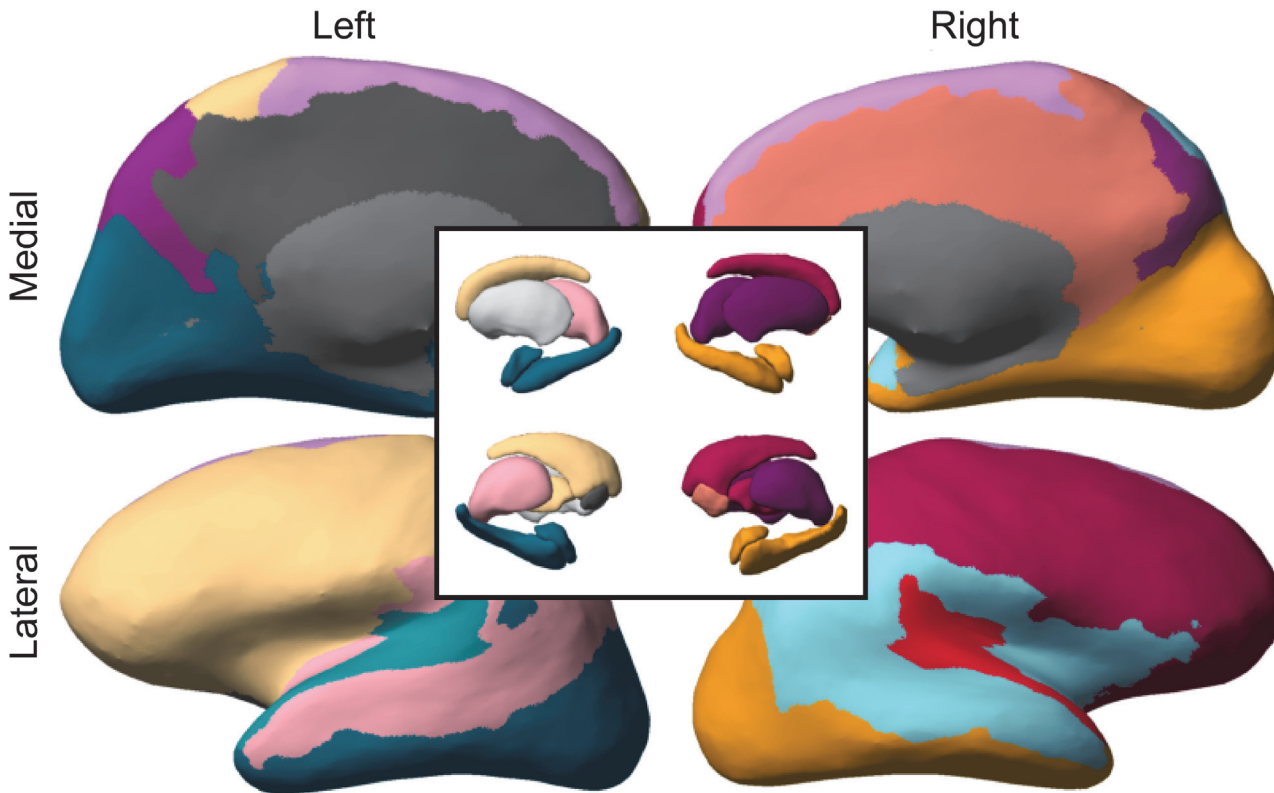
**Fig 2. (Color online) Maximum a posteriori parcellations for one participant.** **A.** The adjacency matrix **A**. **B.** The connection probability matrix  $\rho$ . **C.** The number of connections between clusters  $\mathbf{ZAZ}^T$ . **D.** Visualization of the maximum a posteriori network structure and parcellation. The clusters are color coded to be able to compare the network with the adjacency matrix in **A**. Node sizes are scaled by their degree. **E.** Visualization of the expectation of network structure and parcellations. Colors are interpolated with the MAP estimate as point of reference (see text). To keep the visualization uncluttered, only the  $m$  most probable edges are shown, where  $m$  is the number of edges in the MAP estimate.

doi:10.1371/journal.pone.0117179.g002

clusters in the MAP estimate, but appear to be merged into a single cluster in a substantial part of the distribution. Further uncertainty is shown in the assignment of left thalamus, right precuneus and left inferior parietal cortex. Zooming in on these regions results in what we will refer to as a ‘cluster probability map’; a map that, for any given region of interest  $i$ , shows the probability that it is assigned to the same cluster as another region. This corresponds to a row  $\mathbf{m}_i$  of  $\mathbf{M}$ . Fig. 4 visualizes the cluster probability maps for the right inferior frontal gyrus and the left postcentral gyrus, for the same participant as shown in Fig. 2. Maps like these serve as further illustration that there may be substantial uncertainty in cluster assignments, and that point estimates should be used with care.

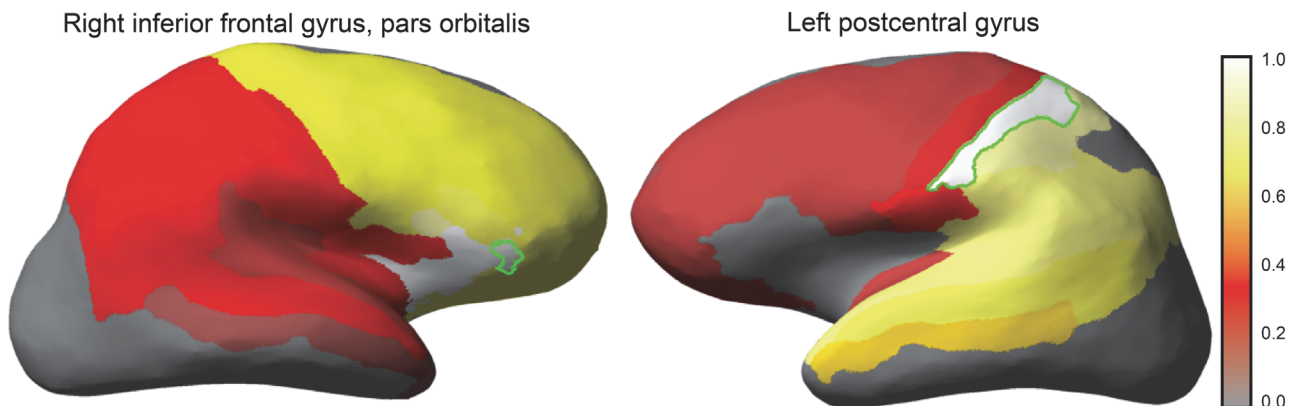
### Comparison with other methods

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



**Fig 3. (Color online)** The maximum a posteriori parcellation for one participant (see Fig. 2), projected onto the inflated cortical surface and subcortical areas. The coloring corresponds to the colors used in Fig. 2.

doi:10.1371/journal.pone.0117179.g003

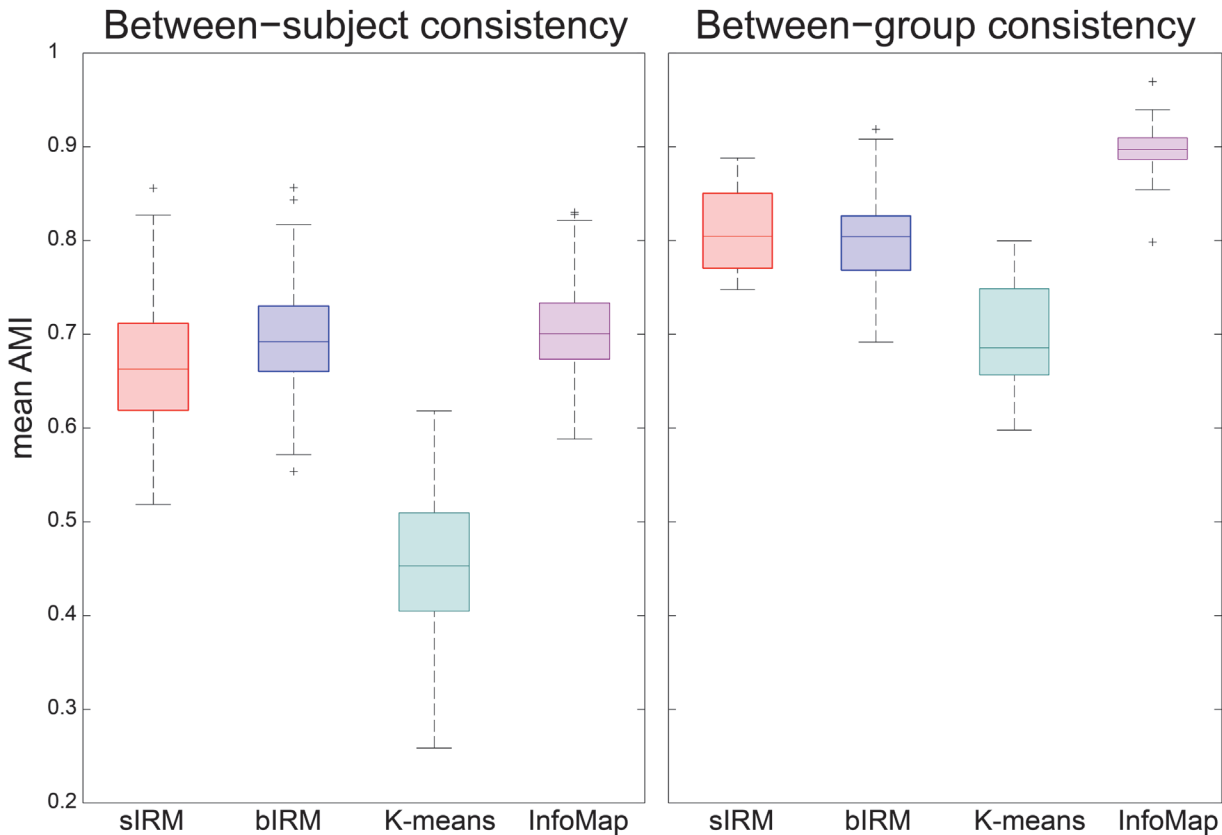


**Fig 4. Cluster probability map to visualize the uncertainty of the resulting cluster assignments.** (left) Cluster probability for the inferior frontal gyrus. The map shows that the highlighted region is likely to be assigned to regions in the frontal cortex, and to a much lesser degree to regions in the parietal cortex. The opposite pattern is shown for the postcentral gyrus (right).

doi:10.1371/journal.pone.0117179.g004

The results of the reproducibility comparisons are shown in Fig. 5. At the level of individual parcellations, only *K*-means performs notably less consistent than the other methods. The SIRM results in sparse connectivity (the MAP estimates of the twenty subjects have a density of  $\pm 7\%$ , on average) and its parcellations are up to par with the BIRM and Infomap. At the group level, Infomap appears to be the most consistent method and *K*-means is again the method





**Fig 5. The reproducibility of parcellations.** **A.** Mean adjusted mutual information (AMI, see main text) for pairwise comparisons between parcellations of all participants. **B.** Mean AMI for each of the pairs of parcellations that were created by randomly splitting the participant group into halves and obtaining a parcellation for each half.

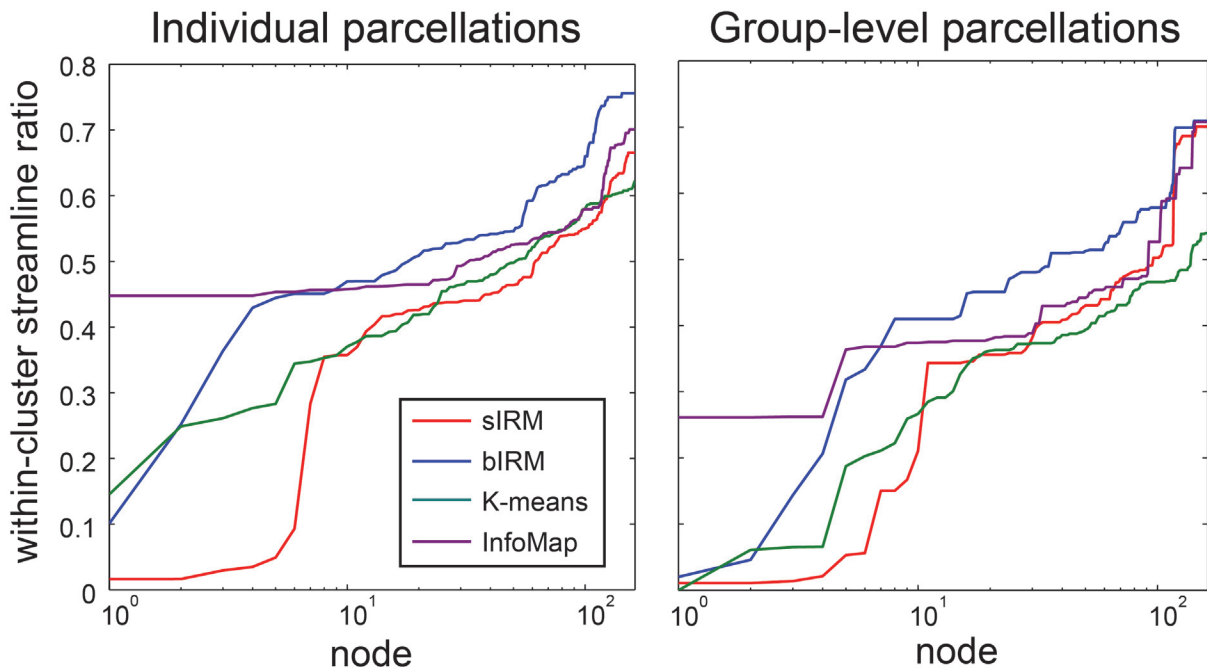
doi:10.1371/journal.pone.0117179.g005

that is least consistent. The group-level parcellations obtained by the sIRM are similar in consistency to those obtained by the bIRM.

### Community-based versus profile-based clusters

The MAP estimates of  $\rho$  as well as the number of connections between clusters in 2 show that clustering behavior depends strongly on cluster identity and that some clusters show community-like tendencies while others do not. To quantify the extent to which a cluster forms a community, and to be able to compare this with the other methods, we computed for each cluster  $c$  the ratio  $r_c$  of within-cluster streamlines versus the total number of streamlines connected to this cluster. When this ratio approaches 1, the corresponding cluster is a community. If instead the ratio approaches 0, it can be regarded a profile-based cluster. As the number of clusters can be different between the sIRM and the bIRM methods, we then assigned to each node the score  $r_c$  its corresponding cluster had.

Fig. 6 shows for both the single-participant as well as for the group-level parcellations the ratio  $r_c$  for all nodes, as averaged over all participants and all group-level parcellations, respectively. The results for the sIRM method clearly reveal that parcellation can be divided into two regimes. The first consists of clusters with relatively high  $r_c$  values (up to 0.76). These are large clusters that correspond predominantly to major cortical areas or lobes that are highly intra-connected. They are connected via the second regime with low  $r_c$  values (as low as 0.02), that



**Fig 6. The ratio describing how community-like the clusters are that each region is assigned to.** The plots show, for the cluster each region is assigned to, the ratio of within-cluster streamlines versus the total number of streamlines connected to that cluster, for parcellations of individual participants (left) and group-parcellations (right). This reveals that in particular for the sIRM method, this ratio is small for a number of regions, indicating that these regions are not part of community-based clusters. Regions that consistently have a low ratio, are described in the text. Note that the nodes corresponding to each line are ordered differently.

doi:10.1371/journal.pone.0117179.g006

contains small clusters (containing less than five nodes) and have few within-cluster streamlines. In at least 15 out of 20 participants, the small clusters contain the bilateral putamina, bilateral superior frontal gyri and right thalamus. For the group-level parcellations we observe similar behavior. Most clusters tend to be communities, connected via small clusters consisting of one or two regions. Small-cluster regions that occur in at least 26 out of 40 group-level parcellations are the bilateral precune, superior frontal gyri, thalami, putamina and left superior parietal gyrus.

The other approaches show different patterns. For the single-participant parcellations, bIRM, K-means or Infomap show no regions that consistently appear in clusters with low  $r_c$ . For group-level parcellations, bIRM finds bilateral superior frontal gyri and right putamen (a subset of the regions found by the sIRM), K-means finds right occipital gyrus, right orbital sulcus and right medial olfactory sulcus and Infomap finds bilateral pallida and left putamen. Remarkably, the regions that stand apart according to the sIRM approach (for both the single-subject and the group-level parcellations) are all known to be part of the ‘rich club’ regions that integrate more remote cortical regions [33]. Note that these regions do not mutually form a single large cluster, as each of them has substantially different connections to the rest of the brain.

## Discussion

Human brain connectivity is shown to exhibit clustering according to two different principles. Multiple brain regions show community-based clustering where clusters are both spatially contiguous and densely intraconnected. These clusters are tied together by brain regions that reveal

connectivity-based clustering. The latter contain only one or two nodes and do not 'fit in' with other clusters due to substantially different connection profiles. These clusters consist predominantly of the superior frontal gyri, the superior parietal gyri, the precuneus, the thalamus and the putamina. All these regions have previously been identified as members of the 'rich club' [33]. In addition, all regions except for the thalamus have been pointed out as being the most vulnerable and central regions in structural brain networks [51]. Since the thalamus can be considered to act as a relay station, it stands to reason that it connects different cortical clusters. The rich-club regions stand apart from the rest of the connectome, as they are not part of the community-like clusters. Rather, they are each assigned their own cluster or group together with a similar region. Intuitively, this is an appealing observation. The community-like clusters contain regions dedicated to specialized processing. Here, an abundance of local connectivity, required for extensive within-cluster communication, results in dense intra-cluster connectivity. Simultaneously, the signal from these clusters must be integrated and disseminated, which is presumably achieved via the rich club regions [52]. Note that although small clusters consisting of one or a few regions will not be community-based clusters by definition, because there are simply not enough possible internal edges, it is the finding that these regions are not assigned to bigger clusters containing other nodes that is interesting. This means that a substantial fraction of the connections of these regions is to several different clusters, which emphasizes their integrative role.

Validating the empirical results from connectivity-based parcellation remains a difficult task [20]. Since ground truth is not available, reproducibility across subjects is often used as a proxy for parcellation quality [14, 47, 48]. Based on this criterion, we have shown that our approach performs well. Yet, by visual inspection of the parcellations (see S2 Fig, S4–S6) some pairs of regions appear grouped together that are not immediately obvious. For instance, in many parcellations we find that thalamus and putamen are assigned to the same cluster, as well as amygdala and hippocampus and bilateral superior frontal gyri. One may argue that in particular for the subcortical areas, these regions should appear in separate clusters as they have specialized functionality. However, since these parcellations are based on connectivity and these regions project similarly to the cortex, they are put together in a cluster. This is inherent to connectivity-based clustering and occurs for  $K$ -means and Infomap as well. In fact, of the algorithms we considered, sIRM appears to be the only method that consistently assigns many of the rich club regions to small clusters instead of agglomerating them into large clusters. Finally, we note that anatomical constraints such as enforcing that subcortical areas constitute singleton clusters may easily be added into the prior distribution.

Previously used methods are limited to one particular clustering behavior. A notable exception is a recent study that reveals that the connectome of the *C. elegans* roundworm also consists of a number of densely intraconnected clusters that are integrated via a core cluster that strongly connects to each community [53]. As we described above, the human connectome reveals both densely connected clusters as well as disconnected clusters with very similar connectivity. Consequently, picking one perspective inevitably neglects part of the available structure, which is exemplified by the parcellations found using  $K$ -means and Infomap that assign the rich club regions to larger clusters. Although, in the case of Infomap, this results in more consistent parcellations, the more detailed picture of communities and integrating regions is lost.

Estimating parcellation with the sIRM approach provides a number of additional interesting quantities apart from the parcellation itself. The cluster connection probabilities  $\rho$  act as a cluster-level estimation of connectivity, expressed in terms of the rich-club regions and major cortices. Region-to-region connectivity  $\mathbf{A}$  is estimated alongside clustering, and is potentially less prone to noise due to the prior that encourages regions in the same cluster to connect similarly, unless the data provides strong evidence to the contrary. Furthermore, other connectivity-based

parcellation strategies, such as  $K$ -means, which groups regions with similar connectivity, and Infomap, which optimizes for densely connected components, provide point estimates of a parcellation. Although these approaches have provided valuable insight in the organization of both structural and functional connectivity [32, 54, 55], they do not quantify the uncertainty in their results (this issue is also discussed in [48]). Instead, our proposed method characterizes the full posterior distribution of all variables involved and thus provides a richer representation of parcellation. We find that a number of regions show substantial uncertainty in their assignments to a cluster. This illustrates that point estimates should be used with care, as a number of regions could easily be assigned erroneously.

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

The IRM is a nonparametric method, which is helpful since the number of clusters is not known a priori. Still, parameters remain that affect the resulting parcellation. In particular the parameters  $\delta_T$  and  $\delta_F$  that govern the estimated connectivity and concentration parameter  $\xi$  that affects the number of identified clusters (although we observe that its influence is drowned out by the contribution of the likelihood) may be estimated from the data using empirical Bayes [56]. However, this will incur substantial additional computation costs that must be overcome to arrive at an efficient model.

There are a number of directions one could take the proposed approach. To start with, alternative generative models may be used instead. For instance, the forward model that now consists of a Dirichlet compound multinomial distribution, modeling the probabilities of streamlines, may be replaced by a Poisson model [57], modeling streamline counts instead. Additional analyses will be needed to identify which model best captures the underlying connectivity. Furthermore, alternative stochastic block models may be used, such as described in a recent study [35] that applies the IRM and two variants to resting-state functional MRI.

Lastly, an interesting avenue to pursue is identifying what causes the uncertainty in the posterior distribution of parcellation. Presumably, this can be attributed to a large extent to noise in data acquisition and tractography. But aside from methodological reasons, there may be other causes that could provide insight in the functional organization of the brain. For instance, uncertainty may in some cases be a result of overlapping clusters [58, 59]. It is likely that regions that are part of multiple, overlapping clusters show higher uncertainty in their cluster assignments. This may be clarified by embedding the infinite relational model within a larger framework to infer clusters at different levels of a nested hierarchy [30, 60]. One would expect that those clusters that show overlap and uncertainty become merged at a higher level of the hierarchy.

## Conclusion

In this paper, we have described an approach for connectivity-based parcellation that encompasses both community-like clusters as well as clusters that have similar connections, without being densely intraconnected. We find that both kinds of clusters are represented in the human connectome, and that the division into these two types corresponds to previous knowledge of structural connectivity. The model is able to quantify which regions are difficult to assign to a cluster, and it learns the number of clusters from the data. Finally, it does not depend on

thresholded connectivity, but derives connectivity simultaneously with the parcellation. We hope that connectivity-based parcellation based on probabilistic models such as the one presented here will help to better understand the structural organization principles of human brain networks.

## Materials and Methods

### Ethics statement

Twenty healthy volunteers were scanned after giving informed written consent in accordance with the guidelines of the local reviewing committee CMO Arnhem-Nijmegen. This study was approved by CMO Arnhem-Nijmegen (CMO 2001/095 and amendment “Imaging Human Cognition”).

### Data acquisition

The data used here was previously described in [61]. A T1 structural scan and diffusion-weighted images were obtained using a Siemens Magnetom Trio 3T system at the Donders Centre for Cognitive Neuroimaging, Radboud University Nijmegen, The Netherlands. An optimized acquisition order described by [62] was used in the DWI protocol (voxel size 2.0 mm isotropic, matrix size  $110 \times 110$ , TR = 13000 ms, TE = 101 ms, 70 slices, 256 directions at  $b = 1500 \text{ s/mm}^2$  and 24 images at  $b = 0$ ).

### Preprocessing

DWI data were preprocessed using FSL FDT [63], which consisted of motion correction, correction for eddy currents and estimation of the diffusion parameters. To obtain a measure of white-matter connectivity, we used FDT Probtrackx 2.0 [63, 64] using seed voxel to target voxel tracking. Structural scans were segmented using FAST [65] and FIRST [66] to generate seed and target voxels. Seed voxels were those voxels in the cortical gray matter mask with a non-zero white-matter partial volume estimate and the outermost voxels of the subcortical masks. The remainder of the cortical and subcortical voxels served as target voxels. All FSL preprocessing was performed using version 5.0 and default settings unless otherwise specified. In addition, streamlines were terminated once they hit the target mask. This prevents transitive connections being erroneously interpreted as direct connections.

The cortical surface was reconstructed using Freesurfer version 5.0 [67] and network nodes were defined according to the atlas provided in [68] (see S1 Fig.) in combination with subcortical segmentation using FIRST [66]. This resulted in a total of 162 nodes. Seed and target voxels were assigned to the nearest vertex in the reconstructed surface and by extension to the corresponding region. All streamline counts between two regions were summed to construct the final streamline matrix.

## Supporting Information

**S1 Fig. The Freesurfer template that defines the network nodes.**  
(EPS)

**S2 Fig. The participant-level parcellations for the streamline IRM.**  
(TIF)

**S3 Fig. The group-level parcellations for the streamline IRM.**  
(TIF)

**S4 Fig. The participant-level parcellations for the *K*-means algorithm.**  
(TIF)

**S5 Fig. The participant-level parcellations for the Infomap algorithm.**  
(TIF)

**S6 Fig. The participant-level parcellations for the IRM using the MAP estimates of the connectivity forward model as input.**  
(TIF)

**S1 Text. Approximate inference of the model.** Description of the block Gibbs sampler used to approximate the posterior distributions and maximum a posteriori estimates.  
(PDF)

**S2 Text. Parameter selection.**  
(PDF)

**S3 Text. Details regarding *K*-means and Infomap.**  
(PDF)

## Acknowledgments

The authors gratefully acknowledge the support of the BrainGain Smart Mix Programme of the Netherlands Ministry of Economic Affairs and the Netherlands Ministry of Education, Culture and Science. The authors thank Erik van Oort and David Norris for the acquisition of the DWI data. Finally, the authors express their gratitude towards Morten Mørup for providing the IRM sampling software and his valuable comments on the manuscript.

## Author Contributions

Conceived and designed the experiments: MvG TH ME MH. Performed the experiments: MH. Analyzed the data: MH RJ. Contributed reagents/materials/analysis tools: MvG TH ME MH RJ. Wrote the paper: MH MvG.

## References

1. Sporns O, Tononi G, Kötter R (2005) The human connectome: A structural description of the human brain. *PLoS Comput Biol* 1: e42. doi: [10.1371/journal.pcbi.0010042](https://doi.org/10.1371/journal.pcbi.0010042) PMID: [16201007](https://pubmed.ncbi.nlm.nih.gov/16201007/)
2. Hagmann P (2005) From diffusion MRI to brain connectomics. Ph.D. thesis, École Polytechnique Fédérale de Lausanne.
3. Greicius MD, Supekar K, Menon V, Dougherty RF (2009) Resting-state functional connectivity reflects structural connectivity in the default mode network. *Cereb Cortex* 19: 72–78. doi: [10.1093/cercor/bhn059](https://doi.org/10.1093/cercor/bhn059) PMID: [18403396](https://pubmed.ncbi.nlm.nih.gov/18403396/)
4. Honey CJ, Sporns O, Cammoun L, Gigandet X, Thiran JP, et al. (2009) Predicting human resting-state functional connectivity from structural connectivity. *Proc Natl Acad Sci USA* 106: 2035–2040. doi: [10.1073/pnas.0811168106](https://doi.org/10.1073/pnas.0811168106) PMID: [19188601](https://pubmed.ncbi.nlm.nih.gov/19188601/)
5. Catani M (2007) From hodology to function. *Brain* 130: 602–605. doi: [10.1093/brain/awm008](https://doi.org/10.1093/brain/awm008) PMID: [17322561](https://pubmed.ncbi.nlm.nih.gov/17322561/)
6. Craddock R, Jbabdi S, Yan CG, Vogelstein J, Castellanos F, et al. (2013) Imaging human connectomes at the macroscale. *Nature methods* 10: 524–539. doi: [10.1038/nmeth.2482](https://doi.org/10.1038/nmeth.2482) PMID: [23722212](https://pubmed.ncbi.nlm.nih.gov/23722212/)
7. Fornito A, Bullmore E (2012) Connectomic intermediate phenotypes for psychiatric disorders. *Frontiers in psychiatry* 3: 1–15. doi: [10.3389/fpsy.2012.00032](https://doi.org/10.3389/fpsy.2012.00032)
8. Bressler S, Menon V (2010) Large-scale brain networks in cognition: emerging methods and principles. *Trends in Cognitive Sciences* 14: 277–290. doi: [10.1016/j.tics.2010.04.004](https://doi.org/10.1016/j.tics.2010.04.004) PMID: [20493761](https://pubmed.ncbi.nlm.nih.gov/20493761/)

9. Baronchelli A, Ferrer-i Cancho R, Pastor-Satorras R, Chater N, Christiansen M (2013) Networks in cognitive science. *Trends in cognitive sciences* 17: 348–360. doi: [10.1016/j.tics.2013.04.010](https://doi.org/10.1016/j.tics.2013.04.010) PMID: [23726319](https://pubmed.ncbi.nlm.nih.gov/23726319/)
10. Sporns O (2014) Contributions and challenges for network models in cognitive neuroscience. *Nature neuroscience* 17: 652–660. doi: [10.1038/nn.3690](https://doi.org/10.1038/nn.3690) PMID: [24686784](https://pubmed.ncbi.nlm.nih.gov/24686784/)
11. Ekman M, Derrfuss J, Tittgemeyer M, Fiebach C (2012) Predicting errors from reconfiguration patterns in human brain networks. *Proceedings of the National Academy of Sciences* 109: 16714–16719. doi: [10.1073/pnas.1207523109](https://doi.org/10.1073/pnas.1207523109) PMID: [23012417](https://pubmed.ncbi.nlm.nih.gov/23012417/)
12. Schleicher A, Amunts K, Geyer S, Morosan P, Zilles K (1999) Observer-independent method for microstructural parcellation of cerebral cortex: a quantitative approach to cytoarchitectonics. *NeuroImage* 9: 165–177. doi: [10.1006/nimg.1998.0385](https://doi.org/10.1006/nimg.1998.0385) PMID: [9918738](https://pubmed.ncbi.nlm.nih.gov/9918738/)
13. Caspers S, Geyer S, Schleicher A, Mohlberg H, Amunts K, et al. (2006) The human inferior parietal cortex: cytoarchitectonic parcellation and interindividual variability. *NeuroImage* 33: 430–448. doi: [10.1016/j.neuroimage.2006.06.054](https://doi.org/10.1016/j.neuroimage.2006.06.054) PMID: [16949304](https://pubmed.ncbi.nlm.nih.gov/16949304/)
14. Klein JC, Behrens TEJ, Robson MD, Mackay CE, Higham DJ, et al. (2007) Connectivity-based parcellation of human cortex using diffusion MRI: Establishing reproducibility, validity and observer independence in BA 44/45 and SMA/pre-SMA. *NeuroImage* 34: 204–211. doi: [10.1016/j.neuroimage.2006.08.022](https://doi.org/10.1016/j.neuroimage.2006.08.022) PMID: [17023184](https://pubmed.ncbi.nlm.nih.gov/17023184/)
15. Knösche TR, Tittgemeyer M (2011) The role of long-range connectivity for the characterization of the functional-anatomical organization of the cortex. *Front Syst Neurosci* 5: 1–13. doi: [10.3389/fnsys.2011.00058](https://doi.org/10.3389/fnsys.2011.00058) PMID: [21779237](https://pubmed.ncbi.nlm.nih.gov/21779237/)
16. Johansen-Berg H, Behrens TEJ, Robson MD, Drobniak I, Rushworth MFS, et al. (2004) Changes in connectivity profiles define functionally distinct regions in human medial frontal cortex. *Proc Natl Acad Sci USA* 101: 13335–13340. doi: [10.1073/pnas.0403743101](https://doi.org/10.1073/pnas.0403743101) PMID: [15340158](https://pubmed.ncbi.nlm.nih.gov/15340158/)
17. Johansen-Berg H, Behrens T, Sillery E, Ciccarelli O, Thompson A, et al. (2005) Functional-anatomical validation and individual variation of diffusion tractography-based segmentation of the human thalamus. *Cereb Cortex* 15: 31–39. doi: [10.1093/cercor/bhh105](https://doi.org/10.1093/cercor/bhh105) PMID: [15238447](https://pubmed.ncbi.nlm.nih.gov/15238447/)
18. Anwender A, Tittgemeyer M, von Cramon D, Friederici A, Knösche T (2006) Connectivity-based parcellation of Broca's area. *Cereb Cortex* 17: 816–825. doi: [10.1093/cercor/bhk034](https://doi.org/10.1093/cercor/bhk034) PMID: [16707738](https://pubmed.ncbi.nlm.nih.gov/16707738/)
19. Mars RB, Jbabdi S, Sallet J, O'Reilly JX, Croxson PL, et al. (2011) Diffusion-weighted imaging tractography-based parcellation of the human parietal cortex and comparison with human and macaque resting-state functional connectivity. *J Neurosci* 31: 4087–4100. doi: [10.1523/JNEUROSCI.5102-10.2011](https://doi.org/10.1523/JNEUROSCI.5102-10.2011) PMID: [21411650](https://pubmed.ncbi.nlm.nih.gov/21411650/)
20. Cloutman LL, Lambon Ralph MA (2012) Connectivity-based structural and functional parcellation of the human cortex using diffusion imaging and tractography. *Front Neuroanat* 6: 1–18. doi: [10.3389/fnana.2012.00034](https://doi.org/10.3389/fnana.2012.00034) PMID: [22952459](https://pubmed.ncbi.nlm.nih.gov/22952459/)
21. Liu H, Qin W, Li W, Fan L, Wang J, et al. (2013) Connectivity-based parcellation of the human frontal pole with diffusion tensor imaging. *J Neurosci* 33: 6782–6790. doi: [10.1523/JNEUROSCI.4882-12.2013](https://doi.org/10.1523/JNEUROSCI.4882-12.2013) PMID: [23595737](https://pubmed.ncbi.nlm.nih.gov/23595737/)
22. Zhang Y, Fan L, Zhang Y, Wang J, Zhu M, et al. (2012) Connectivity-based parcellation of the human posteromedial cortex. *Cereb Cortex*: epub ahead of print.
23. Thiebaut De Schotten M, Urbanski M, Valabregue R, Bayle DJ, Volle E (2012) Subdivision of the occipital lobes: An anatomical and functional MRI connectivity study. *Cortex*: epub ahead of print.
24. Beckmann M, Johansen-Berg H, Rushworth MFS (2009) Connectivity-based parcellation of human cingulate cortex and its relation to functional specialization. *J Neurosci* 29: 1175–1190. doi: [10.1523/JNEUROSCI.3328-08.2009](https://doi.org/10.1523/JNEUROSCI.3328-08.2009) PMID: [19176826](https://pubmed.ncbi.nlm.nih.gov/19176826/)
25. Serra L, Cercignani M, Carlesimo GA, Fadda L, Tini N, et al. (2013) Connectivity-based parcellation of the thalamus explains specific cognitive and behavioural symptoms in patients with bilateral thalamic infarct. *PLoS ONE* 8: e64578. doi: [10.1371/journal.pone.0064578](https://doi.org/10.1371/journal.pone.0064578) PMID: [23755128](https://pubmed.ncbi.nlm.nih.gov/23755128/)
26. Behrens TEJ, Woolrich MW, Smith SM, Boulby PA, Barker GJ, et al. (2003) Non-invasive mapping of connections between human thalamus and cortex using diffusion imaging. *Nat Neurosci* 6: 750–757. doi: [10.1038/nn1075](https://doi.org/10.1038/nn1075) PMID: [12808459](https://pubmed.ncbi.nlm.nih.gov/12808459/)
27. O'Muircheartaigh J, Vollmar C, Traynor C, Barker GJ, Kumari V, et al. (2011) Clustering probabilistic tractograms using independent component analysis applied to the thalamus. *NeuroImage* 54: 2020–2032. doi: [10.1016/j.neuroimage.2010.09.054](https://doi.org/10.1016/j.neuroimage.2010.09.054) PMID: [20884353](https://pubmed.ncbi.nlm.nih.gov/20884353/)
28. Hilgetag C, Kaiser M (2004) Clustered organization of cortical connectivity. *Neuroinformatics* 2: 353–360. doi: [10.1385/NI:2:3:353](https://doi.org/10.1385/NI:2:3:353) PMID: [15365196](https://pubmed.ncbi.nlm.nih.gov/15365196/)
29. van den Heuvel M, Mandl R, Hulshoff Pol H (2008) Normalized cut group clustering of resting-state fMRI data. *PLoS ONE* 3: e2001. doi: [10.1371/journal.pone.0002001](https://doi.org/10.1371/journal.pone.0002001) PMID: [18431486](https://pubmed.ncbi.nlm.nih.gov/18431486/)

30. Meunier D, Lambiotte R, Fornito A, Ersche KD, Bullmore E (2009) Hierarchical modularity in human brain functional networks. *Front Neuroinform* 3: 1–12. doi: [10.3389/neuro.11.037.2009](https://doi.org/10.3389/neuro.11.037.2009) PMID: [19949480](https://pubmed.ncbi.nlm.nih.gov/19949480/)
31. Craddock RC, James GA, Holtzheimer PE, Hu XP, Mayberg HS (2012) A whole brain fMRI atlas generated via spatially constrained spectral clustering. *Hum Brain Mapp* 33: 1914–1928. doi: [10.1002/hbm.21333](https://doi.org/10.1002/hbm.21333) PMID: [21769991](https://pubmed.ncbi.nlm.nih.gov/21769991/)
32. Power JD, Cohen AL, Nelson SM, Wig GS, Barnes KA, et al. (2011) Functional network organization of the human brain. *Neuron* 72: 665–678. doi: [10.1016/j.neuron.2011.09.006](https://doi.org/10.1016/j.neuron.2011.09.006) PMID: [22099467](https://pubmed.ncbi.nlm.nih.gov/22099467/)
33. van den Heuvel MP, Sporns O (2011) Rich-club organization of the human connectome. *Journal of Neuroscience* 31: 15775–15786. doi: [10.1523/JNEUROSCI.3539-11.2011](https://doi.org/10.1523/JNEUROSCI.3539-11.2011) PMID: [22049421](https://pubmed.ncbi.nlm.nih.gov/22049421/)
34. Nowicki K, Snijders TAB (2001) Estimation and prediction for stochastic blockstructures. *Journal of the American Statistical Association* 96: 1077–1087. doi: [10.1198/016214501753208735](https://doi.org/10.1198/016214501753208735)
35. Andersen K, Madsen K, Siebner H, Schmidt M, Mrup M, et al. (2014) Non-parametric bayesian graph models reveal community structure in resting state fmri. *NeuroImage* 100: 301–315. doi: [10.1016/j.neuroimage.2014.05.083](https://doi.org/10.1016/j.neuroimage.2014.05.083) PMID: [24914522](https://pubmed.ncbi.nlm.nih.gov/24914522/)
36. Airoidi E, Blei D, Fienberg S, Xing E (2009) Mixed membership stochastic blockmodels. In: Koller D, Schuurmans D, Bengio Y, Bottou L, editors, *Advances in Neural Information Processing Systems 21*, Curran Associates, Inc. pp. 33–40.
37. Guimerà E, Sales-Pardo M (2009) Missing and spurious interactions and the reconstruction of complex networks. *Proceedings of the National Academy of Sciences* 106: 22073–22078. doi: [10.1073/pnas.0908366106](https://doi.org/10.1073/pnas.0908366106)
38. Wolfe A, Jensen D (2004) Playing multiple roles: Discovery overlapping roles in social networks. In: *CML-04 Workshop on Statistical Relational Learning and its Connections to Other Fields*.
39. Aicher C, Jacobs A, Clauset A (in press) Learning latent block structure in weighted networks. *Journal of Complex Networks*.
40. Kemp C, Tenenbaum J, Griffiths T, Yamada T, Ueda N (2006) Learning systems of concepts with an infinite relational model. In: *Proceedings of the 21st national conference on Artificial intelligence*. AAAI Press, volume 1 of *AAAI'06*, pp. 381–388.
41. Aldous D (1985) Exchangeability and related topics. In: Hennequin P, editor, *École d'Été de Probabilités de Saint-Flour XIII 1983*, Springer Berlin Heidelberg, volume 1117 of *Lecture Notes in Mathematics*. pp. 1–198.
42. Mørup M, Madsen KH, Dogonowski AM, Siebner H, Hansen LK (2010) Infinite relational modeling of functional connectivity in resting state fMRI. In: *Neural Information Processing Systems*. Curran Associates, Inc., volume 23 of *NIPS'10*, pp. 1750–1758.
43. Xu Z, Tresp V, Yu S, Yu K, Kriegel H (2007) Fast inference in infinite hidden relational models. In: *Proceedings of the 5th International Workshop on Mining and Learning with Graphs (MLG 2007)*. Firenze, Italy.
44. Behrens ea TEJ (2003) Characterization and propagation of uncertainty in diffusion-weighted MR imaging. *Magn Reson Med* 50: 1077–1088. doi: [10.1002/mrm.10609](https://doi.org/10.1002/mrm.10609) PMID: [14587019](https://pubmed.ncbi.nlm.nih.gov/14587019/)
45. Hinne M, Heskes T, Beckman CF, van Gerven M (2013) Bayesian inference of structural brain networks. *NeuroImage* 66: 543–552. doi: [10.1016/j.neuroimage.2012.09.068](https://doi.org/10.1016/j.neuroimage.2012.09.068) PMID: [23041334](https://pubmed.ncbi.nlm.nih.gov/23041334/)
46. Bassett DS, Brown JA, Deshpande V, Carlson JM, Grafton ST (2011) Conserved and variable architecture of human white matter connectivity. *NeuroImage* 54: 1262–1279. doi: [10.1016/j.neuroimage.2010.09.006](https://doi.org/10.1016/j.neuroimage.2010.09.006) PMID: [20850551](https://pubmed.ncbi.nlm.nih.gov/20850551/)
47. Blumensath T, Jbabdi S, Glasser MF, Van Essen DC, Ugurbil K, et al. (2013) Spatially constrained hierarchical parcellation of the brain with resting-state fMRI. *NeuroImage* 76: 313–324. doi: [10.1016/j.neuroimage.2013.03.024](https://doi.org/10.1016/j.neuroimage.2013.03.024) PMID: [23523803](https://pubmed.ncbi.nlm.nih.gov/23523803/)
48. Thirion B, Varoquaux G, Dohmatob E, Poline JB (2014) Which fMRI clustering gives good brain parcellations? *Frontiers in Neuroscience* 8. doi: [10.3389/fnins.2014.00167](https://doi.org/10.3389/fnins.2014.00167) PMID: [25071425](https://pubmed.ncbi.nlm.nih.gov/25071425/)
49. Vinh NX, Epps J, Bailey J (2010) Information theoretic measures for clusterings comparison: Variants, properties, normalization and correction for chance. *Inf Process Manage* 11: 2837–2854.
50. Rosvall M, Bergstrom CT (2008) Maps of random walks on complex networks reveal community structure. *Proc Natl Acad Sci USA* 105: 1118–1123. doi: [10.1073/pnas.0706851105](https://doi.org/10.1073/pnas.0706851105) PMID: [18216267](https://pubmed.ncbi.nlm.nih.gov/18216267/)
51. Iturria-Medina Y, Sotero RC, Canales-Rodríguez E, Alemán-Gómez Y, Melie-García L (2008) Studying the human brain anatomical network via diffusion-weighted MRI and graph theory. *NeuroImage* 40: 1064–76. doi: [10.1016/j.neuroimage.2007.10.060](https://doi.org/10.1016/j.neuroimage.2007.10.060) PMID: [18272400](https://pubmed.ncbi.nlm.nih.gov/18272400/)



52. van den Heuvel MP, Sporns O (2013) An anatomical substrate for integration among functional networks in human cortex. *J Neurosci* 33: 14489–500. doi: [10.1523/JNEUROSCI.2128-13.2013](https://doi.org/10.1523/JNEUROSCI.2128-13.2013) PMID: [24005300](https://pubmed.ncbi.nlm.nih.gov/24005300/)
53. Pavlovic D, Vértes P, Bullmore E, Schafer W, Nichols T (2014) Stochastic blockmodeling of the modules and core of the *Caenorhabditis elegans* connectome. *PLoS ONE* 9: e97584. doi: [10.1371/journal.pone.0097584](https://doi.org/10.1371/journal.pone.0097584) PMID: [24988196](https://pubmed.ncbi.nlm.nih.gov/24988196/)
54. Skudlarski P, Jagannathan K, Anderson K, Stevens MC, Calhoun VD, et al. (2010) Brain connectivity is not only lower but different in schizophrenia: a combined anatomical and functional approach. *Biological psychiatry* 68: 61–69. doi: [10.1016/j.biopsych.2010.03.035](https://doi.org/10.1016/j.biopsych.2010.03.035) PMID: [20497901](https://pubmed.ncbi.nlm.nih.gov/20497901/)
55. Nanetti L, Cerliani L, Gazzola V, Renken R, Keysers C (2009) Group analyses of connectivity-based cortical parcellation using repeated *k*-means clustering. *NeuroImage* 47: 1666–1677. doi: [10.1016/j.neuroimage.2009.06.014](https://doi.org/10.1016/j.neuroimage.2009.06.014) PMID: [19524682](https://pubmed.ncbi.nlm.nih.gov/19524682/)
56. Robbins H (1964) The empirical bayes approach to statistical decision problems. *Ann Math Statist* 35: 1–20. doi: [10.1214/aoms/1177703729](https://doi.org/10.1214/aoms/1177703729)
57. Schmidt MN, Mørup M (2013) Nonparametric Bayesian Modeling of Complex Networks: An Introduction. *IEEE Signal Processing Magazine* 30: 110–128. doi: [10.1109/MSP.2012.2235191](https://doi.org/10.1109/MSP.2012.2235191)
58. Rubinov M, Sporns O (2011) Weight-conserving characterization of complex functional brain networks. *NeuroImage* 56: 2068–2079. doi: [10.1016/j.neuroimage.2011.03.069](https://doi.org/10.1016/j.neuroimage.2011.03.069) PMID: [21459148](https://pubmed.ncbi.nlm.nih.gov/21459148/)
59. Yeo BTT, Krienen FM, Chee MWL, Buckner RL (2013) Estimates of segregation and overlap of functional connectivity networks in the human cerebral cortex. *NeuroImage* 88C: 212–227. doi: [10.1016/j.neuroimage.2013.10.046](https://doi.org/10.1016/j.neuroimage.2013.10.046) PMID: [24185018](https://pubmed.ncbi.nlm.nih.gov/24185018/)
60. Blundell C, Teh YW (2013) Bayesian hierarchical community discovery. In: Burges C, Bottou L, Welling M, Ghahramani Z, Weinberger K, editors, *Advances in Neural Information Processing Systems 26*, Curran Associates, Inc. pp. 1601–1609.
61. van Oort ESB, van Cappellen van Walsum AM, Norris DG (2013) An investigation into the functional and structural connectivity of the Default Mode Network. *NeuroImage*: epub ahead of print.
62. Cook PA, Symms M, Boulby PA, Alexander DC (2007) Optimal acquisition orders of diffusion-weighted MRI measurements. *J Magn Reson Imaging* 25: 1051–1058. doi: [10.1002/jmri.20905](https://doi.org/10.1002/jmri.20905) PMID: [17457801](https://pubmed.ncbi.nlm.nih.gov/17457801/)
63. Behrens TEJ, Woolrich MW, Jenkinson M, Johansen-Berg H, Nunes RG, et al. (2003) Characterization and propagation of uncertainty in diffusion-weighted MR imaging. *Magn Reson Med* 50: 1077–88. doi: [10.1002/mrm.10609](https://doi.org/10.1002/mrm.10609) PMID: [14587019](https://pubmed.ncbi.nlm.nih.gov/14587019/)
64. Behrens TEJ, Berg HJ, Jbabdi S, Rushworth MFS, Woolrich MW (2007) Probabilistic diffusion tractography with multiple fibre orientations: What can we gain? *NeuroImage* 34: 144–155. doi: [10.1016/j.neuroimage.2006.09.018](https://doi.org/10.1016/j.neuroimage.2006.09.018) PMID: [17070705](https://pubmed.ncbi.nlm.nih.gov/17070705/)
65. 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. doi: [10.1109/42.906424](https://doi.org/10.1109/42.906424) PMID: [11293691](https://pubmed.ncbi.nlm.nih.gov/11293691/)
66. Patenaude B, Smith SM, Kennedy DN, Jenkinson M (2011) A Bayesian model of shape and appearance for subcortical brain segmentation. *NeuroImage* 56: 907–922. doi: [10.1016/j.neuroimage.2011.02.046](https://doi.org/10.1016/j.neuroimage.2011.02.046) PMID: [21352927](https://pubmed.ncbi.nlm.nih.gov/21352927/)
67. Dale A, Fischl B, Sereno MI (1999) Cortical surface-based analysis: I. segmentation and surface reconstruction. *NeuroImage* 9: 179–194. doi: [10.1006/nimg.1998.0395](https://doi.org/10.1006/nimg.1998.0395) PMID: [9931268](https://pubmed.ncbi.nlm.nih.gov/9931268/)
68. Destrieux C, Fischl B, Dale A, Halgren E (2010) Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. *NeuroImage* 53: 1–15. doi: [10.1016/j.neuroimage.2010.06.010](https://doi.org/10.1016/j.neuroimage.2010.06.010) PMID: [20547229](https://pubmed.ncbi.nlm.nih.gov/20547229/)