A Hierarchical Method for Whole-Brain Connectivity-Based Parcellation

David Moreno-Dominguez,[1]* **Alfred Anwander,**[2] **and Thomas R. Knösche**[1]

[1]*Research Group "Cortical Networks and Cognitive Functions," Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany*
[2]*Department of Neuropsychology, Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany*

◆ ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ ◆

Abstract: In modern neuroscience there is general agreement that brain function relies on networks and that connectivity is therefore of paramount importance for brain function. Accordingly, the delineation of functional brain areas on the basis of diffusion magnetic resonance imaging (dMRI) and tractography may lead to highly relevant brain maps. Existing methods typically aim to find a predefined number of areas and/or are limited to small regions of grey matter. However, it is in general not likely that a single parcellation dividing the brain into a finite number of areas is an adequate representation of the function-anatomical organization of the brain. In this work, we propose hierarchical clustering as a solution to overcome these limitations and achieve whole-brain parcellation. We demonstrate that this method encodes the information of the underlying structure at all granularity levels in a hierarchical tree or dendrogram. We develop an optimal tree building and processing pipeline that reduces the complexity of the tree with minimal information loss. We show how these trees can be used to compare the similarity structure of different subjects or recordings and how to extract parcellations from them. Our novel approach yields a more exhaustive representation of the real underlying structure and successfully tackles the challenge of whole-brain parcellation. *Hum Brain Mapp 35:5000–5025, 2014.* © 2014 Wiley Periodicals, Inc.

Key words: diffusion MRI; connectome; whole-brain parcellation; hierarchical clustering; tractography; dendrogram

◆ ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ ◆

INTRODUCTION

It is commonly accepted among neuroscientists that the cerebral cortex can be subdivided into areas according to various structural criteria, including the distribution of different cell types (cytoarchitecture), the distribution of myelinated fibers (myeloarchitecture), and the distribution of different neurotransmitter receptors (receptorarchitecture) [Amunts et al., 2007, 2010; Brodmann, 1909; Vogt, 1910, 1911; Zilles, 2004; Zilles and Amunts, 2009, 2010; Zilles et al., 2004]. The most widely known such parcellation is still the cytoarchitectonic map of Brodmann, based on the specific variation in size and packing density of cell bodies over the layers of the cortical sheet in one single subject. It is also generally agreed that brain structure is closely related to brain function and, therefore, structurally defined cortical

areas tend to carry functional meaning. Consequently, many studies have aimed to find the boundaries between these areas, using a variety of techniques based on local structural tissue properties. However, the brain is not only a collection of isolated functional units; the different parts communicate and interact in a complex network ultimately resulting in higher cognitive capabilities. The connectivity pattern of a specific point in the cortex is, therefore, a major source of information about its function and an important parameter for the description and distinction of cortical areas [Barbas and Rempel-Clower, 1997; Knösche and Tittgemeyer, 2011; Passingham et al., 2002]. The subdivision of the brain into function-anatomically defined areas is also a necessary step for the *connectome*, characterized by elements (the regions being connected) and the connections between them [Sporns, 2011].

However, it is unlikely that a single parcellation dividing the brain into a finite number of functional areas would be an adequate representation of the functional organization of the brain, in the same way that a political map subdividing the earth's land surface is not a perfect representation of the cultural differences and kinships amongst its people. The measurable changes of properties on the cortical surface are often gradual rather than abrupt. In these cases, we might find different partitions depending on how we define the minimum structural difference that just merits distinction, that is, on the required level of granularity of the partition. Also, even in cases where these changes are sharp and a partition remains constant for a wide range of granularities, there can still exist nested divisions within the regions of this partition. This is exemplified by the cytoarchitecture work of Caspers et al. [2008] and the tractography work of Ruschel et al. [in press], where Brodmann's areas 39 and 40 were further subdivided. A partition should, therefore, be seen as an approximation of the similarity structure (e.g., expressed by a correlation matrix) of some structural properties at a particular level of granularity.

Brain connectivity is among the most relevant structural cues in terms of brain function [Knösche and Tittgemeyer, 2011]. The arrival of dMRI, and particularly the ability to describe the anatomical connectivity pattern of a point in the cortex by means of tractography, has enabled researchers to perform *in vivo* cortical parcellation based on brain connectivity [Anwander et al., 2007; Johansen-Berg et al., 2004]. Classical approaches usually focus on one particular subdivision of the cortical surface and apply rather strong constraints and assumptions. For example, *target-based* clustering [Behrens et al., 2003] involves the strong assumption that each parcel should be mainly connected to one out of a set of predefined target areas. On the other hand, so-called *free clustering* algorithms do not have this assumption, but the number of expected parcels, average size of clusters, or a similar parameter must be known in advance [Anwander et al., 2007], posing a classical model selection problem. The implicit assumption here is that there is a parcellation that can be considered a reasonably unique and complete representation of the connectivity similarity structure, which is rarely likely to be the case. There have been attempts to deal with non-uniqueness through having a series of parcellations [Kahnt et al., 2012], attempting to find an, in some sense, optimal parcellation (Jbabdi et al., 2009), or searching for a space of optimal parcellations [Gorbach et al., 2011; Roca et al., 2009]. However, when faced with a whole-brain approach, the challenge of not only having a high and unknown expected number of areas, but also that number being subject to the desired granularity of the partitioning, arises.

In this work, we propose *hierarchical clustering* as an approach to overcome these limitations. Hierarchical methods have been applied before to partition functional magnetic resonance imaging (fMRI) connectivity data [Cordes et al., 2002; Liu et al., 2012; Stanberry et al., 2003] and recently also to obtain full cortical fMRI parcellations at multiple granularities [Blumensath et al., 2013]. In dMRI, agglomerative hierarchical clustering has been used to perform parcellation of white matter pathways by Wassermann et al. [2010] and Guevara et al. [2011].

We aim to demonstrate that hierarchical clustering is also a promising means by which to characterize the similarity structure of anatomical connectivity patterns in the human brain, where the information of the underlying structure at all granularity levels is encoded in a hierarchical tree or dendrogram. For this purpose, we implemented and compared several hierarchical methods and the best performing algorithm, both by data-fit and computational cost criteria, was selected. It combines hierarchical centroid linkage clustering with a physical neighborhood restriction. Once trees are obtained, interpreting the large amount of data encoded and extracting the most relevant information is not an easy task. To aid this process, a dendrogram pre-processing pipeline was designed that reduces the complexity of the resulting trees, while keeping most of its information, to facilitate further analysis. Finally, we pursue the idea that these trees can then be sampled to obtain relevant partitions at different granularity levels.

Another important issue is the comparison of parcellations between subjects. This is a non-trivial issue, since, as argued above, in each subject many different parcellations

---

Abbreviations

CPCC	cophenetic correlation coefficient
dMRI	diffusion magnetic resonance imaging
FA	fractional anisotropy
fMRI	functional magnetic resonance imaging
GRAPPA	generalized autocalibrating partially parallel acquisitions
SNR	signal to noise ratio
SS	spread vs. separation
tCPCC	tree cophenetic correlation coefficient
TE	echo time
TR	repetition time
wTriples	weighted triples similarity.

are possible, depending on local and global granularity constraints. Also, even if one manages to identify matching parcellations in different subjects, the comparison based thereupon only applies for the respective granularity level, while for finer or coarser subdivisions the result could be completely different. On the other hand, the hierarchical tree is just a compact representation of all possible parcellations. Hence, comparing the entire tree, instead of individual parcellations, should circumvent the above mentioned issues. We show how the trees can be used to compare the similarity structure of different subjects or time points: globally, using the full connectivity structure information through dendrogram comparison, and at selected granularity levels through the use of partition finding algorithms. Importantly, this is performed while remaining in the subject space without the need to transform the data to a common space prior to partitioning [Wang et al., 2013].

METHODS

Data Acquisition and Preprocessing

High resolution dMRI images as well as T1- and T2-weighted images were acquired for four young and healthy participants (three males and a female) on a Siemens TimTrio scanner with a 32-channel array head coil and maximum gradient strength of 40 mT/m. For one of the participants, a second set of images was acquired after a one-week interval. Written informed consent was obtained from the subjects in accordance with the ethical approval from the University of Leipzig.

The dMRI data was acquired using spin-echo echo-planar imaging, with repetition time (TR) = 11 s, echo time (TE) = 90 ms, 85 axial slices, resolution 1.5 mm isotropic, GRAPPA/3, and three acquisitions. We used 60 diffusion gradient directions, which were evenly distributed over the half-sphere (b-value = 1,000 s/mm^2). The diffusion-weighted volumes were interspersed by acquisitions with no diffusion weighting (b0 images) at the beginning and after each block of 10 volumes (7 volumes). The total scan time for the dMRI protocol was approximately 45 min.

As a first preprocessing step, the three-dimensional T1-weighted (magnetization prepared-rapid gradient echo, TR = 1,300 ms, time to inversion = 650 ms, TE = 3.93 ms, resolution 1.0 × 1.0 × 1.5 mm, two acquisitions, reconstructed to 1 mm isotropic resolution) images were reoriented to the mid-sagittal plane through the anterior and posterior commissures and the brain volume was segmented using the *Lipsia* software package [Lohman et al. 2001]. The 21 images without diffusion weighting were used to estimate motion correction parameters using rigid-body transformations [Jenkinson et al., 2002], implemented in FSL (FMRIB Software Library, Oxford, UK). Motion correction parameters were interpolated for all 201 volumes and combined with a global registration to the T1 anatomy using a mutual information registration algorithm. The dif-

fusion gradient direction for each volume was corrected using the rotation parameters. The registered images were linearly interpolated to the new reference frame with an isotropic voxel resolution of 1 mm and the three corresponding acquisitions and gradient directions were averaged. Next, the diffusion tensor was calculated for each voxel after logarithmic transformation of the signal intensities [Basser et al., 1994]. Finally, the fractional anisotropy (FA) of the tensor in each voxel was subsequently determined, and a multi-slice FA image [Basser and Pierpaoli, 1996] was created. The combined motion correction and registration to the individual T1 anatomy provided some advantages. A simple motion correction to the first image in the diffusion weighted sequence would have introduced a variable amount of smoothing caused by the interpolation of the images to the reference image. For example, the first images in the sequence would have needed less interpolation and the reduced smoothing would have caused a directional bias. Using the independent orientation of the T1 image as reference removed this potential bias. Additionally, the sampling of the data with a higher spatial resolution (1mm instead of 1.5mm) allowed keeping more details of the data compared with a resampling with the original resolution. In this way, interpolation of the raw data provided some methodological advantages in the following tractography step.

White Matter Tractography

The brain volume was segmented into white and gray matter compartments by means of FA thresholding (white matter: FA ≥ 0.15) and interactive corrections for deep white matter imperfections. Using an FA based mask allows to define seed voxels at a clearly defined white matter boundary. This precession would not have been possible using the white matter mask from the segmented T1 image, since the diffusion image shows small non-linear distortions. Each white matter voxel that neighbored a cortical gray matter voxel was used as a seed voxel for the probabilistic dMRI tractography (that is, each single grey matter/white matter boundary voxel at 1 mm resolution, between 130,000 and 200,000 seed voxels per brain depending on size), as proposed by Anwander et al. [2007]. The tractography algorithm computed a transition probability of a simulated particle jumping from one voxel to the next from the diffusion data. Next, the probabilistic tractography started 100,000 particles in each seed voxel. The particles propagated in the white matter as guided by the local transition probabilities, defined by the probability density function from the diffusion tensor model. The target space was the whole white matter volume with a resolution of 1 mm^3. The diffusion data was not interpolated in this step and used the interpolation of the raw diffusion data as computed in the preprocessing steps. Finally, a visitation map was computed from the number of particles which cross each voxel. The tractography algorithm was

parallelized and implemented on a consumer PC graphic board (GPU) and took only a few seconds per seed point.

The three-dimensional distribution of the connectivity values (visitation map) of a particular seed voxel with all voxels in the brain is called a tractogram. In these tractograms, which we use as connectivity fingerprints, the value associated with a particular white matter voxel represents the visitation fraction, that is, what proportion of all particles started at the seed voxel went through that particular voxel. The visitation values ranging between 0 and 100,000 were *log* transformed to reduce the dynamic range (in order to palliate the intrinsic bias that visitation-based connectivity values have towards favoring short connections against longer distance ones, which are especially problematic for the computation of similarities between tractograms) and scaled between 0 and 1 (1 means all, 0 means none of the started streamlines touched the voxel). These values are taken as a correlate for the anatomical connectivity between that voxel and the seed voxel of the tractogram. Although based on a simple local model (diffusion tensor), this probabilistic tractography can, to a certain extent, account for fanning fibers and fiber crossings. This provides tractograms with enough overlap area to detect connectivity pattern differences between voxels at the discrimination level required for successful parcellation.

To analyze the effects of a reduced signal-to-noise ratio (SNR) onto the developed analysis methods, a second set of tractograms was obtained for the first three subjects using just a single acquisition of the diffusion data (in contrast to averaging the three available acquisitions).

Hierarchical Clustering

In order to characterize the similarity of structural connectivity in a granularity range as wide as possible, agglomerative hierarchical clustering was applied over the tractogram fingerprints. This type of clustering starts by considering every object in the dataset as a separate cluster, then it merges the closest (i.e., most similar) pair of clusters, according to some similarity criterion, and iterates until all of the data points belong to one single cluster. The result is essentially a binary tree, where each position in the *x*-axis corresponds to a connectivity fingerprint (also called leaves) and the values in *y*-axis where any two leaves join for the first time refer to the dissimilarity or distance between the two fingerprints as encoded by the tree. An outline of the clustering process applied to anatomical connectivity can be seen in Figure 1.

A noncentered variant of Pearson's correlation coefficient was used as a similarity metric [Eq. (A1)], as it is better suited for structural tractograms, where all values are positive and different degrees of negative linear dependency values do not hold relevant biological information (details on this choice can be found in the Appendix).

Different agglomerative methods use different measures to calculate the new distances when elements are merged (linked). The most widely used linkage methods in the literature are the four types of graph methods (single, com-
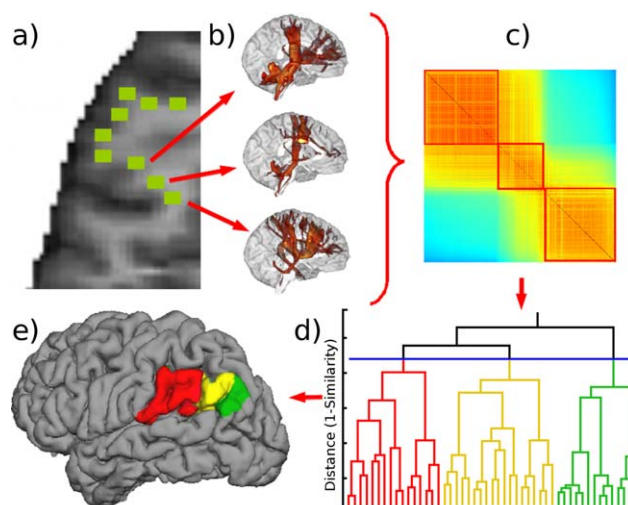

Figure 1.

Schema of the hierarchical clustering process. **a)** Select gray-matter/white-matter interface voxels; **b)** generate probabilistic tractograms of seed voxels; **c)** compute similarities between tractograms; **d)** build-up connectivity tree; **e)** select partitions within the tree and map back to the cortex. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

plete, weighted, and average linkage methods), called so as they stem from graph theory [Murtagh, 1983]. In all of these methods it is necessary to calculate the pairwise distances between all elements. This can prove costly when there are a large number of elements and the points are in a very high dimensional space, as is the case in the scenario of connectivity-based whole-brain parcellation using 1 mm resolution.

In order to reduce the computation and memory requirements, we elected to explore a fifth approach based on the *centroid linkage* method [Jain and Dubes, 1988]. In this method, each cluster is defined by its centroid: a data point that represents all the points included in the cluster. In the study presented here, the centroid was computed as the average of the tractograms in natural space.

If the assumption is made that a connectivity-defined region in the brain must always be a connected patch of gray matter, then only mergers between neighboring clusters are allowed, and only those distances have to be computed, drastically reducing the cost of the algorithm (a neighborhood restriction may also be used in the graph methods in order to force morphologically continuous clusters, but the whole distance matrix must still be calculated and thus it yields no computational advantage). The concept of spatially constrained hierarchical clustering has also been exploited for fMRI data [Blumensath et al. 2013], a modality where this particular restriction has proved of advantage for parcellation in the past [Craddock et al. 2011].

As an output of these algorithms (both graph and centroid) a binary tree (also called bifurcating rooted tree or

fully resolved dendrogram) is obtained. This tree encodes the connectivity similarity structure of the dataset at all granularity levels, transforming into a much-reduced dimensionality ($2n$, n being the number of seed elements) the information of the distance matrix (dimension n^2) obtained from the tractogram space (dimension $n \cdot m$, m being the number of white matter voxels).

One of the advantages of this method is the possibility of comparing the full connectivity structure across datasets through tree comparison, which we will further develop in a later subsection. In order to do this, the leaves of the trees to be compared must first be matched. With this in mind, an extra restriction was applied to the centroid method during the initial iterations of the tree building process. The objective is to ensure that at the lower levels of the tree (that is, the ones with highest granularity) the clusters are joined in a homogenous way, with roughly equal sizes, until a certain number of clusters has been reached. As will be explained later, this allows for easier leaf matching. There is, however, no restriction upon the shape of these clusters, and their merging is still guided by connectivity pattern similarity. The concept of a 2-stage clustering approach (where first a maximum granularity partition is obtained from which to build the hierarchical tree) has also been successfully used by Gorbach et al. [2011] and Blumensath et al. [2013] to partition dMRI and fMRI data, respectively (although the particular implementations are substantially different).

For thorough description of the methods implemented and their mathematical formulations please refer to the Appendix.

Dendrogram Preprocessing

Even after the optimal linking method has been chosen, the task of extracting relevant information from the resulting dendrogram is not simple: the high number of seed voxels involved could translate into many possible granularity levels and partitions. The nature of the clustering process also forces the dendrogram to always have binary bifurcations, whereas in reality the dataset is likely to have structures nested in a non-binary way. This means that some of the nodes in the tree do not contribute any real information about the similarity structure and are merely a byproduct of the pair-wise agglomerative method. Also, as in most real datasets, outliers could be present. Finally, in the case of the centroid linkage method, non-monotonic steps can occur, which, although not constituting an error in themselves, can complicate partition finding algorithms and make visual interpretation of the tree difficult.

In order to address these problems and ease the information extraction, several dendrogram preprocessing steps were developed and applied: elimination of outliers; monotonicity correction; limiting the maximum-granularity captured in the tree and detection of non-binary structures followed by removal of the corresponding intermediate nodes (see Appendix for details on this section). These preprocessing methods effectively reduce the number of branchings, which in turn reduces the tree complexity and possible confounds in the dendrogram, while still maintaining maximum usable information. This also facilitates the task of the information extraction algorithms, which are introduced below.

Tree Comparison Across Measurements

Once the connectivity similarity structure of a brain is encoded in a dendrogram, there is the possibility of using the information from the whole tree to assess the structural differences in brain connectivity between different subjects or measurements.

Dendrogram comparison techniques are already in use in other fields, with most efforts being dedicated to the field of phylogenetics [Critchlow et al., 1996; Restrepo et al., 2007]. However, these techniques are used to compare different trees built over the same dataset, relying on a perfect match between the leaf elements of both trees. In the scenario of brain connectivity trees from different measurements, this would only be the case if the dendrograms being compared originate from the same brain, and only if there have not been significant changes in morphology nor the data acquisition method.

Leaf-matching across trees

In order to be able to apply these comparison methods when assessing connectivity structure variability across subjects, the problem of leaf identification had to be tackled. Potentially, there are different possible criteria for the identification of associated pairs of leafs in two dendrograms, for example spatial proximity after a more or less sophisticated co-registration of the images or cortical surfaces derived from these images. However, as the dendrograms to be compared are based on the similarity of tractograms, it seems appropriate to use the same criterion for finding matched pairs of leafs. The solution provided involves several steps:

- First, the trees are preprocessed with the techniques previously introduced, in order to reduce the number of leaves and provide a maximum granularity partition. These maximum granularity partitions are fine-tuned so that all the trees to be compared have the same number of meta-leaves. This number is chosen to obtain an acceptable complexity reduction while incurring minimal information loss.
- Mean tractograms corresponding to each of the meta-leaves are obtained for all subjects. The mean tractogram of any given node is calculated as the log-transformed average of the raw (not log-transformed) seed tractograms contained in the respective node.
- The subjects' FA images are non-linearly registered to a common space, and this transformation is applied to the mean tractograms. The registration is performed through the *ANTS* package [SyN registration
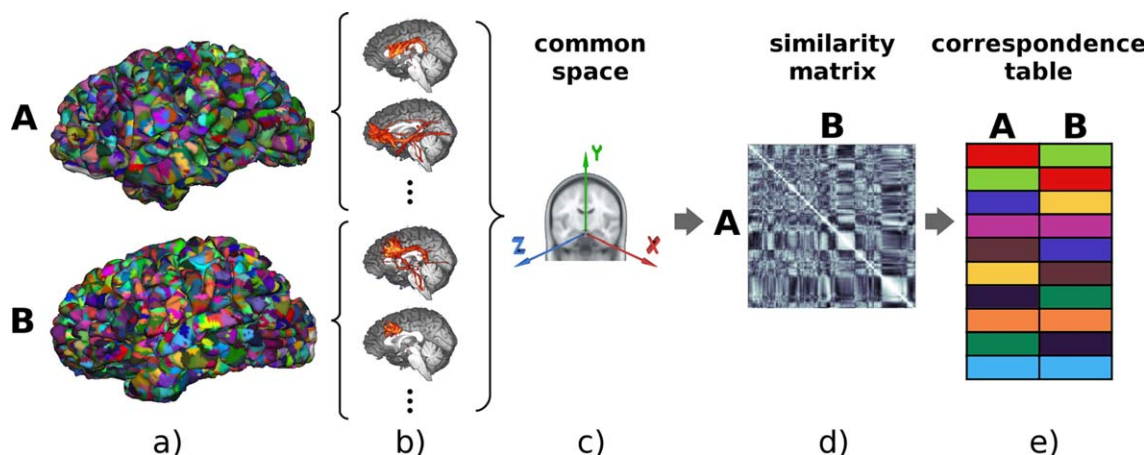
Figure 2.

Leaf-identification pipeline: maximum effective granularity partitions are obtained for each subject, i.e., Subjects A and B (**a**); a mean tractogram is computed for each cluster (**b**); all tracts are registered to a common space (**c**); pairwise tract similarity matrix is computed between the subjects (**d**); a greedy algorithm is used to extract the cluster correspondence table from the matrix (**e**). These clusters will become the new leaves of the trees. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

algorithm; Avants et al., 2008; Klein et al., 2009]. The mean tractograms are then transformed to the same common space using the deformation fields obtained from the FA image registration.

- For each pair of trees being compared, a tractogram distance matrix between their corresponding meta-leaves is obtained.
- Matching of the meta-leaves of the trees is done by applying a greedy algorithm to the distance matrix: The two tractograms with the highest similarity are matched and their entries are eliminated from the data. This step is iterated until there are no more entries in the matrix. In order to avoid poor matches, restrictions on minimum tractogram similarity and maximum Euclidean distance between cluster morphological centers are applied (minimum mean-tractogram similarity: 0.1 and minimum spatial distance between cluster centers: 2 cm) Clusters for which no suitable correspondence can be made are discarded and not considered in the comparison. There are other matching algorithms available, such as the Hungarian method [Kuhn 1955], which tries to optimize the matching in terms of global rather than local distance between matched elements. However, this also means higher computation time and resources. For a first implementation and proof of the method, we chose the simpler greedy matching with reduced computation time.

The leaf matching process is outlined in Figure 2.

Tree similarity measures

Two different tree similarity measures were implemented:

Tree cophenetic correlation coefficient (tCPCC). Farris [1969] introduced the cophenetic correlation coefficient (*CPCC*) to assess the degree to which a tree successfully encodes the similarity information by measuring the correlation between the pairwise distance matrix obtained directly from the data and a distance matrix derived from the tree structure. This principle can be adapted for tree comparison by instead correlating the distance values encoded by each of the trees for each pair of corresponding meta-leaves.

Weighted triples similarity (wTriples). An alternative tree comparison method, described in detail by Bansal et al. [2011], consists of comparing the joining order of all possible triples of leaves of each tree. The number of triples for which the joining order is exactly the same is divided by the total number of possible triples; obtaining a value ranging between 0 and 1.

The *tCPCC* and *wTriples* comparison methods are (partially) complementary: while the former stresses the similarity of the distance values encoded by both trees, the latter focuses on the similarity of the hierarchical topologies of both trees, regardless of the numerical values encoded. See Appendix for details on their mathematical implementation for this work.

Partition Selection

As argued above, the tree is suitable for assessing the structural map of the cortical sheet as a whole. However, in order to fully appreciate the function-anatomical organization of the cortex, we also need to map this information back onto the cortical surface. Because the tree is a multidimensional structure, it cannot be fully projected directly onto this two-dimensional space. Some strategies have

been proposed that allow including some degree of multi-granularity information into surface mapping, such as using similar color hues for subclusters of a bigger cluster (for example using reddish, greenish and bluish hues for subclusters of three main divisions) or hierarchical "space-blobs" [Cachia et al., 2003]. These approaches, however, are not suitable for the very high range of granularities and high number of nodes present in our trees. As an alternative, representative parcellations (being equivalent to a complete cut of the tree that severs all connections between the top node and any leaf) may be found that best approximate the information encoded in the tree. It is very unlikely that a single partition can represent the entire similarity structure of the data. Using a series of partitions at different granularity levels, which in this case would also be hierarchically nested, might be a better way to achieve it.

Many different methods for comparing and assessing partitions can be found in the literature [Halkidi et al., 2002; Rand, 1971; Theodoridis and Koutroubas, 1999]. However, these methods usually refer to the original data, which in our case would involve operations with high dimensional tractograms, making them computationally expensive and slow. Limiting the data used to that contained in the tree allows fast partition assessment algorithms to be implemented. There is also literature available on tree partitioning algorithms [Jain and Dubes, 1988; Langfelder et al., 2008; Zahn, 1971], but these methods did not translate into meaningful partitions in the case of the brain connectivity trees studied here. The most traditional approach to tree partitioning does, however, deserve introduction.

Minimum guaranteed intracluster similarity (horizontal cut)

By definition, if a horizontal cut is made through a dendrogram the partition obtained is the one that guarantees, for a given number of clusters, a lower bound for the intracluster similarity. Therefore, this cut yields regions with a certain minimum required consistency (or greater). In order to select a partition, either a number of desired clusters, or the distance level where the horizontal cut is made must be chosen.

Cluster spread vs. separation (SS) index

The horizontal cut method only takes into account the distance level of the clusters involved in the partition, that is, the encoded distance between the elements contained in those clusters, which relates to spread or scatter of the clusters. A more complete partition selection method should also consider the distance between such clusters related to their spread. Furthermore, the horizontal cut may only be used with a pre-defined granularity level and is unable to assess the quality of a partition. In order to tackle these shortcomings, we introduced a second algorithm presented below.

The overall spread of the clusters in a partition can be quantified through the formula:

$$Spread = \frac{1}{S_T} \sum_i^N S_i d_i \qquad (1)$$

where d_i and S_i are the distance level and size of cluster i, respectively. N is the number of clusters in the partition, and S_T the sum of all clusters sizes in the partition.

The distance level of the parent of a given node in the tree encodes the separation between the center of that cluster and that of its closest neighbor. The average separation between neighboring clusters for a given partition can then be expressed as:

$$Separation = \frac{1}{N} \sum_i^N d_{p(i)} \qquad (2)$$

where $d_{p(i)}$ is the distance level of the parent of node i.

Using these two formulas, a partition quality measure is obtained by calculating the ratio between the mean spread of clusters in the partition and the mean separation between neighboring clusters: the *spread-separation* (*SS*) index.

$$SS = \frac{S_T \sum_i^N d_{p(i)}}{N \sum_i^N d_i S_i} \qquad (3)$$

A higher value will indicate that, for that partition, the mean separation of clusters is high compared with the separation of elements within the clusters. This index can be used to find global or local maxima in the tree, thus revealing partitions of particular significance. Alternatively, it can be coupled with a required number of clusters, in order to find the best possible partition of a given desired granularity.

However, due to the data size and the extremely high number of possible partitions contained in the trees, an exhaustive assessment of all possible cuts through the tree would exceed any reasonable computational limits. In order to obtain partition selection methods fast enough to be integrated into an interactive tree exploration tool, a top-down hierarchical search algorithm was implemented here. This means that, starting at the partition defined by the first branching of the tree (or sub-tree), all possible subdivisions of each cluster going down up to four branching levels are considered, and the resulting partitions are evaluated. The best performing partition is identified, and the corresponding cluster from which division it had derived is subdivided down one level. The process is iterated until the desired number of clusters has been obtained or the maximum granularity partition has been reached.

Minimum cluster size difference

Due to the nature of the tractogram similarity measure used, areas that share long common pathways (like, for example, the longitudinal fasciculus) will tend to be more similar to their surrounding areas sharing these large connections than to those with shorter pathways or more local

connectivity fingerprints (such as the superior frontal lobe). Such highly cohesive areas tend to remain less partitioned by the spread-separation scheme than areas with local connectivity. Depending on the purpose of the partitioning, it may be useful to circumvent this side-effect by obtaining partitions guided by the connectivity structure encoded in the tree but with an emphasis on clusters of similar sizes. This can be accomplished by finding partitions that minimize the mean square size difference for a given number of clusters (using the same partition search algorithm as described for the previous method). The objective function to be minimized is expressed as:

$$SizeDiff = \frac{2}{N(N-1)} \sum_{i}^{N-1} \sum_{j=i+1}^{N} (S_i - S_j)^2 \qquad (4)$$

All algorithms described above were included as part of a fast interactive exploration and visualization tool for hierarchical characterization of brain connectivity, which simultaneously projects the selected partitions onto the brain surface. This tool was implemented as a module of the open-source *OpenWalnut* framework [www.openwalnut.org].

RESULTS

Choosing the Linkage Method

Hierarchical trees were built for both hemispheres of subjects A, B, and C (i.e., six datasets) using each of the graph linkages and the centroid-neighborhood method proposed. In order to assess their fit to the data, *CPCC* values (see *Appendix*) were computed for all obtained dendrograms. The results show that the Average and Centroid linkages perform well above the rest with no statistically significance difference between them (with values close to 0.8 over 1 against 0.65 from the next best performing method). Additional calculation of the computational load for each modality give the centroid method a clear advantage over the average method for very large datasets, with three orders of magnitude less operations needed. Therefore, the centroid method (with a 26-voxel neighborhood restriction) was selected for the rest of the study. The dendrogram preprocessing pipeline was applied to the centroid trees obtained, achieving a complexity reduction of more than 90% with a loss of information of less than 0.5% (0.15% on average), making it a remarkably efficient and useful tool for improving the performance of partition finding and tree comparison algorithms. Detailed accounts of the linkage selection process and dendrogram cleaning pipeline and parameters are included in the *Appendix*.

Comparing the Connectivity Structure Across Datasets

The information encoded in the cleaned trees can be used as a whole in order to detect structural differences between datasets. As described in the *Methods* section, mean tractograms were obtained for each meta-leaf of the processed trees and nonlinearly transformed to a common space, guided by FA registration. Here we morphed the data of subjects A and B into the space of subject C. For the within-subject comparisons across hemispheres, the tractograms of the right hemisphere were flipped and transformed into the left hemisphere; also guided by a previous FA registration. Next, the tractogram-distance matrices were obtained and the greedy leaf-matching algorithm was applied (see *Methods*). Using the resulting leaf-matching tables, the *tCPCC* and the *wTriples* similarity values were obtained. In order to test the reliability of the method and its robustness against noise, the whole process (starting at tractogram computation and tree building) was repeated with a noisier version of the same dataset, using only one, instead of three, repetitions of the MRI acquisition. Test–retest performance was also assessed using two datasets obtained from a fourth subject within a short period of time (1 week), referred to as D1 and D2.

In order to establish a baseline level for the matching values, a random matching scheme was set up, in which each meta-leaf of the first tree was matched at random to a meta-leaf of the second tree whose cluster center was not further away than 2 cm. Afterwards, *tCPCC* and *wTriples* values were obtained. This process was repeated 100 times for each possible subject combination and the average value was obtained. Distinct baseline values from both *tCPCC* and *wTriples* were computed for inter-subject comparisons, left versus right hemisphere comparisons and high versus low SNR comparisons.

The results are shown in Figure 3, where *tCPCC* and *wTriples* are plotted against the leaf matching quality (mean tractogram distances between matched clusters) between the two compared data sets. Several observations can be made:

- All tree comparison values obtained are well above their corresponding baseline levels, indicating that the matchings were not trivial, and that there are nonrandom structural similarities between the trees that can be detected.
- For *tCPCC*, the information loss by lower SNR and the variability between separate measurements of the same subject are smaller (i.e., the *tCPCC* is higher) than the differences between different hemispheres or subjects. This indicates that differences in leaf similarity, as encoded in the trees, are not generally obscured by noise and can be interpreted. In contrast, *wTriples*, which only measures tree topology (joining orders), seems to be much more susceptible to noise.
- The similarities between the same hemispheres in different subjects and those between different hemispheres in the same subjects are within the same order of magnitude (between-hemispheres slightly lower, but not significant).
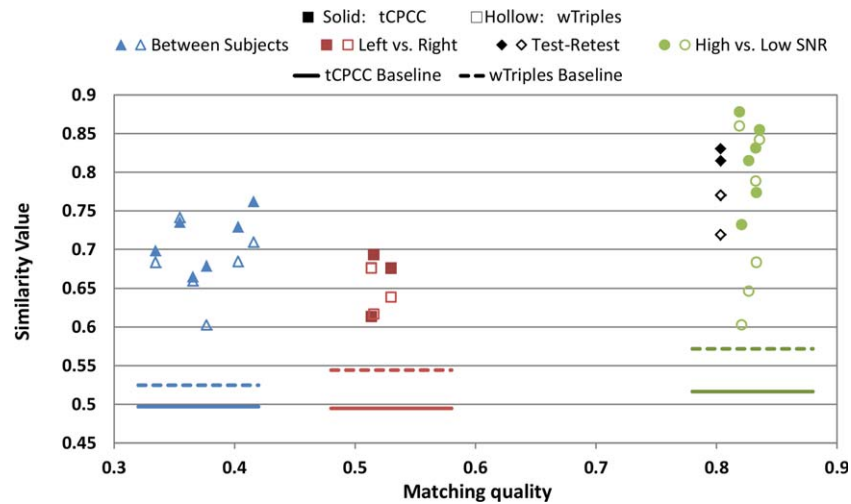
Figure 3.

Tree similarity values plotted against matching quality for tree comparisons. Baseline levels for the corresponding matchings are shown below their datapoints in the same color, solid lines correspond to *tCPCC* baselines, and dotted lines to *wTriples* ones. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

- Same-subject comparisons features much better leaf-matching quality compared with between-hemispheres comparisons, which in turn match better than between-subject comparisons.

Single Subject Partitioning

Using the horizontal cut and the spread-separation tree-partitioning methods (see *Methods*) nested whole-brain partitions were obtained at different granularity levels (defined in this case by a particular number of clusters). The spread-separation method yielded very similar results to those of the horizontal cut method. The nested partitions are exemplified in Figure 4, where we show the left hemisphere of subject A cut at four different granularity levels, exploring a wide range of hierarchical boundaries. At very low granularity (15 clusters) the parcellation seemed to reflect the rough course of major fiber bundles (e.g., red for the fronto-occipital fascicle, green for the arcuate fascicle, purple for the cingulum bundle, and cyan for the cortico-spinal tract). Increasing the granularity to 50 clusters caused further subdivisions, especially in the dorsolateral and dorsomedial frontal and parietal cortices, and also in the inferior frontal cortex and around the auditory cortex, reaching area sizes similar to Brodmann areas. Meanwhile, the cortex near the fronto-occipital fascicle, the superior part of the arcuate fascicle, and the cingulum bundle remained largely undivided. To obtain more fine-grained subdivisions in these regions, the threshold of the clustering criterion had to be lowered further, allowing for 100 clusters. Further increase of granularity continued

changing details, for example by further subdividing the inferior frontal gyrus.

In Figure 5, we focused on the subdivision of the left inferior frontal gyrus (IFG). At relatively low granularity (50 clusters), only some of the major boundaries between the opercular and triangular parts (subject A, B) and between the triangular and orbital parts of the IFG were revealed. At higher granularity, more subdivisions appeared, including those that are not covered by the classical tripartition (into opercular, triangular, and orbital parts). For the repetitive acquisitions in the same subject (D1 and D2), the subdivision was highly reproducible. Figure 6 exemplifies the similarity of our parcellation with cytoarchitectonic parcellation available from Jülich Research Centre (www.jubrain.fz-juelich.de/apps/cyto-viewer/cytoviewer-main.php).

It appears that, if tree-cutting is based on internal coherence and mutual separation of the clusters (i.e., horizontal cut or spread separation methods), uninteresting "background" connectivity by large fiber tracts cause, at any given level of granularity, some regions of the brain to remain largely undivided, while others were split into small sub-areas. This lead to the introduction of the minimized cluster size difference method (see *Methods* section). In Figure 7, the result for this partition method is depicted for the same subject featured in Figure 4. When comparing the results of the two partitioning methods, some clear differences are apparent. At low granularity (15 clusters), the large temporal-occipital-frontal cluster (in red, see Figure 4) broke up into smaller areas, especially on the medial brain surface, while in frontal and prefrontal cortex fewer clusters were formed. This trend is also evident at higher granularities. For example, at 250
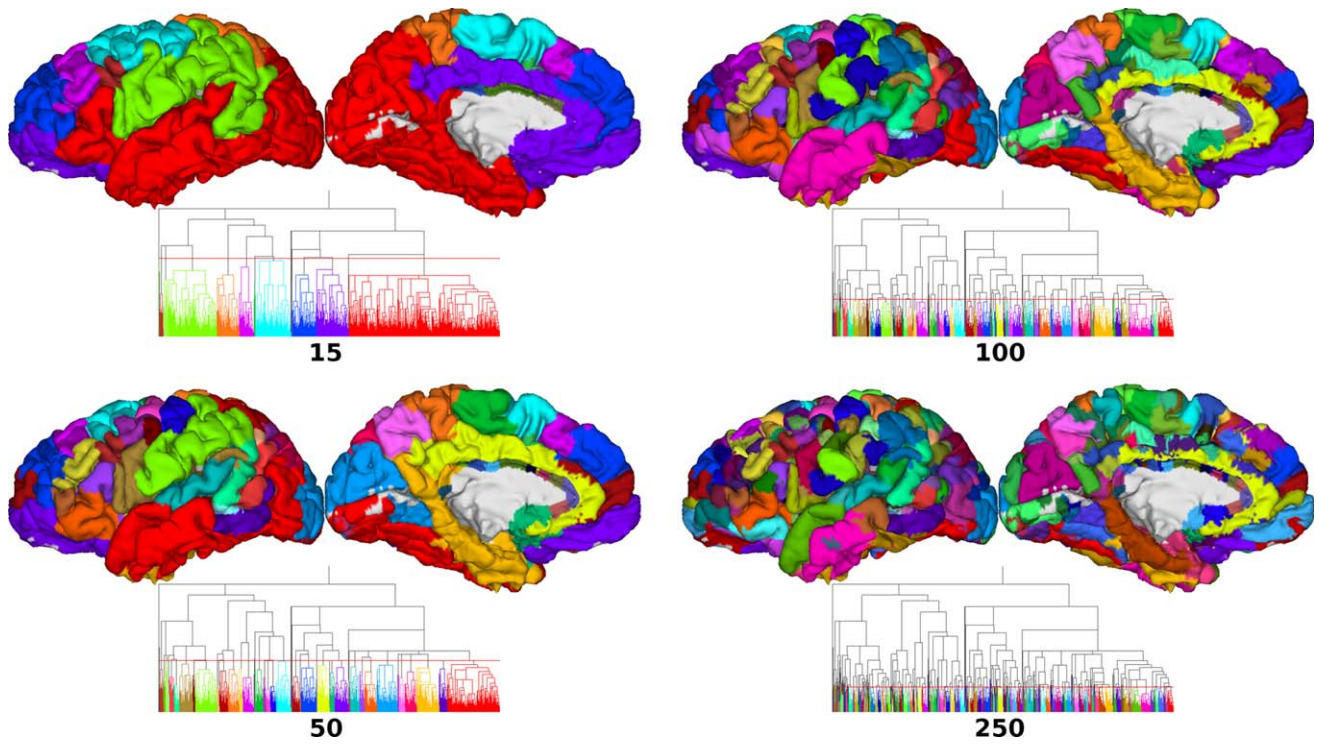
Figure 4.
Parcellations extracted from the hierarchical tree of the left hemisphere of subject A using the horizontal cut algorithm. The numbers indicate the predefined number of clusters. The red horizontal lines in the trees denote the cutting level. The spread-separation method yields almost identical results. See text for further explanation. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]
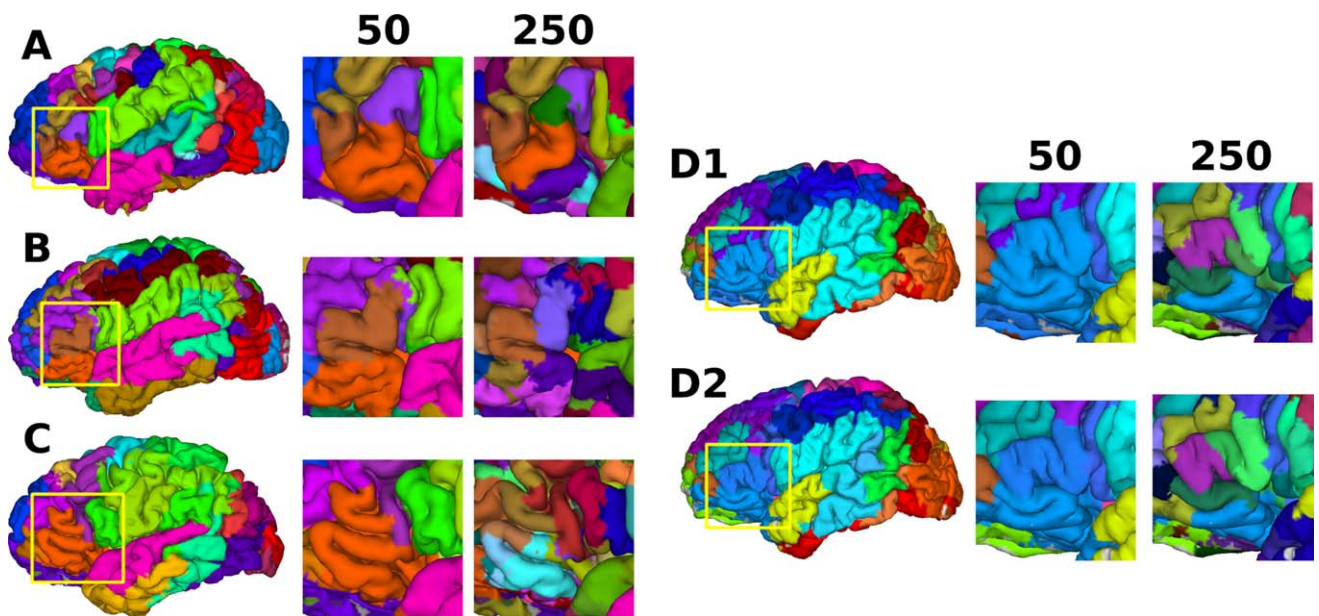

Figure 5.
Spread-separation subdivision of the inferior frontal gyrus at two different levels of granularity, for the left hemispheres of subjects A, B, and C (left), as well as the two acquisitions of subject D (right). See text for further explanation. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]
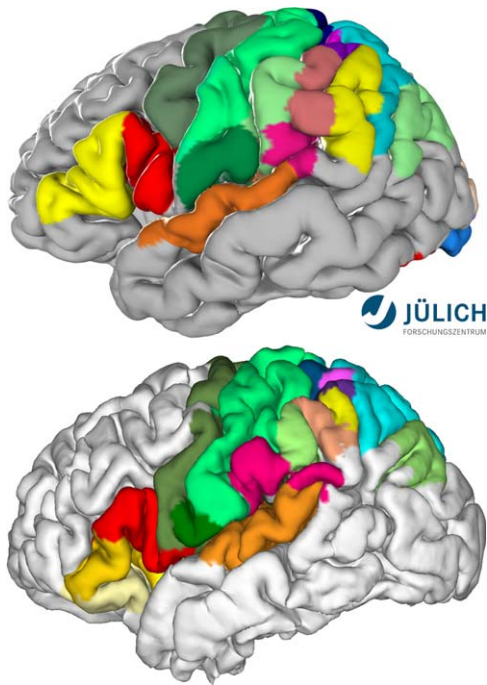
Figure 6.
Cytoarchitectonic parcellation provided by Jülich Research Centre (top), compared with the corresponding subtree of the left hemisphere of subject A at a global horizontal partition for 100 clusters (bottom; two clusters, one in the IFG, and other in the parietal cortex over the STG, have been further subdivided once to better show the corresponding matching). [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

clusters the occipital lobe was more subdivided and the frontal one was less subdivided than with the horizontal cut method.

Thus far, we have explored the partitioning methods that required the input of a global granularity level (here expressed as number of clusters, but it could also be the average size of of the clusters, or similar). However, the question remains: Which granularity levels might be the most representative ones for the tree? In order to reduce this arbitrariness, one can use the *SS* index (see *Methods* section) to select partitions. Using the *SS* partitioning, a series of parcellations can be obtained with maximum *SS* indices for each granularity level. In Figure 8 the *SS* indices were plotted as function of granularitiy for all data sets.

It can be seen that for small numbers of clusters the index rises steeply, meaning that in this range further subdivision usually leads to much better parcellations. In many data sets, this is followed by a shoulder (at about 50–200 clusters), where further subdivision does not greatly improve, or even slightly reduces, the quality of the parcellation (as measured by the *SS* index). Next, there follows a moderate increase, where subdivisions tend to (slightly to moderately) improve the *SS* index, until a maximum value is reached at about 200 to 600 clusters. From there, the curve steadily decreases, meaning that further subdivisions always lead to worse partitions. Consequently, the relevant range of partitions seems to start at the edge of the first shoulder and end at the maximum (where both mergings and subdivisions cause a moderate decrease of the *SS* index). Ultimately, the interesting range of partitions based on the diffusion data seems to be roughly 20 to 600 clusters.

Figure 9 shows the maximum *SS* index partitions for all subjects and hemispheres. These partitions have the maximum distinctness for the respective data sets, that is, the best ratio between intra-cluster inhomogeneity and between-cluster separation. These parcellations feature small parcels with an extent comparable to the width of a
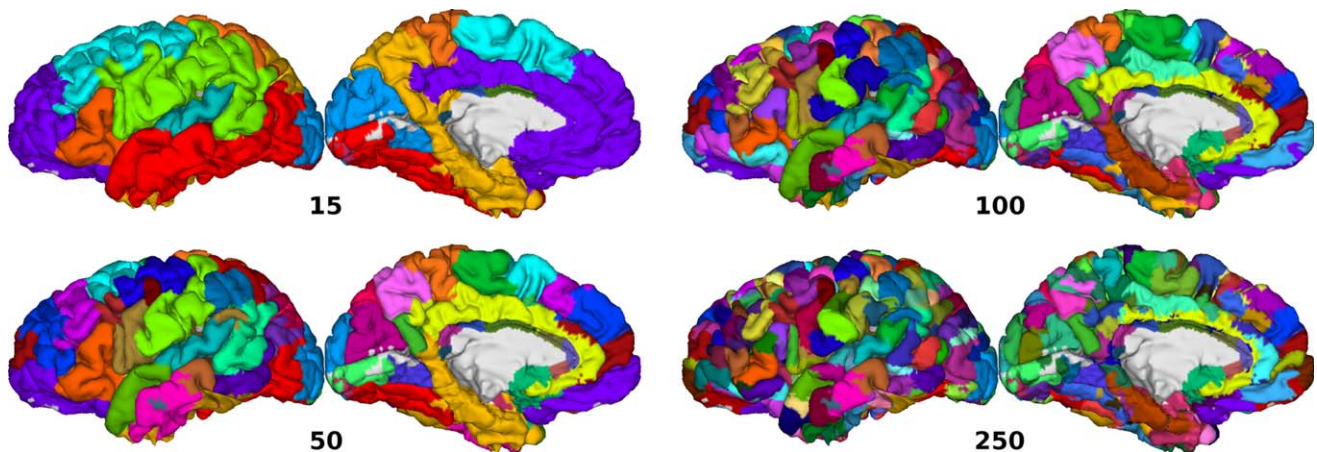

Figure 7.
Minimum size-difference partitioning for the same subject and cluster numbers as depicted in Figure 4. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]
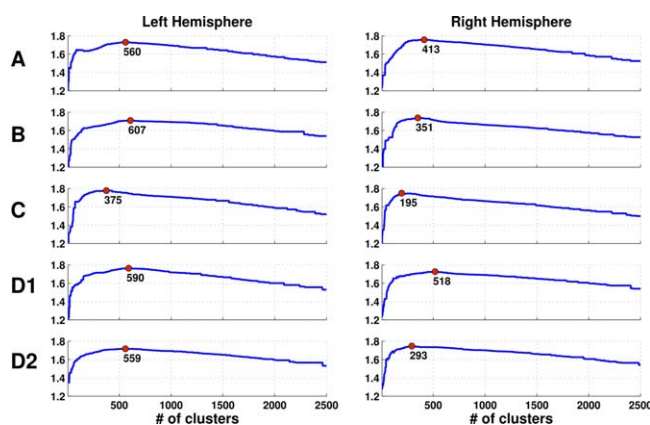
Figure 8.

SS indices obtained by the hierarchy search method, plotted against number of clusters. The red circles denote the maximums of the curves. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

major gyrus. It is evident that, at this level of granularity, the partitions of the two data sets from subject D are quite similar, while the partitions belonging to different hemispheres and/or subjects appear very different.

DISCUSSION

Tractography-Based Parcellation

As argued before, connectivity is among the most relevant structural cues for the characterization of the functio-anatomical identity of cortical tissue. Being the only method that can be applied to healthy human subjects, diffusion tractography is the method of choice for the reconstruction of these connectivity patterns [Anwander et al., 2007; Johansen-Berg et al., 2004]. For a thorough discussion of this issue, see Knösche and Tittgemeyer [2011].

The tractography based parcellation requires a robust tractography method. The local tensor model based on High Angular Resolution Diffusion Images (HARDI) allows a reproducible computation of the connectivity profile. The method is sensitive to small changes in connectivity between two voxels and is robust to noise which could affect the local model. Other tractography methods like the Probabilistic Index of Connectivity (PICo) based on the Persistent Angular Structure (PAS) [Parker and Alexander, 2005] or probabilistic tractography based on spherical deconvolution [Descoteaux et al. 2009] had shown to better resolve crossing fiber structures. The more complex local model might have been less robust to remaining noise in the diffusion data, which might have affected the local estimation of the fiber orientations [Yo et al. 2009]. While comparing a tensor based tractography with fiber tracking using spherical deconvolution Kristo et al. [2013] showed a higher reproducibility for the tensor based tractography. In this initial study we choose to use the more robust local model.

The fact that probabilistic tractography is employed ensures that, to a certain degree, fiber crossings and branching are taken into account. The parcellation method we proposed could be applied on any other tractography method. The comparison of the result using different local models and tractography algorithms will be subject of future investigations. In addition, all tractography algorithms including the one used here have a number of adjustable parameters, which potentially can affect the tractography result and the parcellation. For example, here we had to make choices on the number of streamlines, the scaling and the thresholding of the tractograms and the sharpening of the local diffusion profile [Anwander et al., 2007]. While a systematic parameter study on this and other tractography algorithms would certainly be very useful, the previous use of our approach in a number of parcellation studies yielding neuroanatomically plausible results provides some confidence [e.g., Anwander et al., 2007; Gorbach et al., 2011, 2012; Ruschel et al., in press; Schubotz et al., 2010].

In most implementations of tractography based parcellation the target space comprises the entire rest of the brain, including white matter [Anwander et al., 2007; Johansen-Berg et al., 2004; Mars et al., 2011; Schubotz et al., 2010; Tomassini et al., 2007]. A possible alternative is to restrict the target space to grey matter (or, for technical reasons, the white matter voxels just adjacent to grey matter) [e.g., Bach et al., 2011]. It is, however, not clear whether this really improves the situation. Most tractography methods are iterative algorithms that, especially over long distances, tend to accumulate errors and hence are subject to substantial blurring [Jones, 2010]. So, it is likely that differences between tracts, which are still quite evident in the intermediate white matter, become smoothed out at the distant cortical targets. On the other hand, using the entire brain as target space might also introduce biases of its own, as the tracts starting from two spatially distinct cortical elements are different by definition in their initial sections, even if they finally reach the same targets. This is especially true, if the tracts start in different gyri. How much this effect influences the result depends on the overall extent of the tractogram, that is, the relative weight of short and long range connections. So, the fact that parcellations often seem to reflect, to some degree, sulcal patterns (see Figs. 4–9), might have a methodological background. On the other hand, it is well known that in many cases macroanatomical landmarks, such as sulcal lines, are indeed likely to play a role as function-anatomical boundaries [Hasnain et al., 2001; Tahmasebi et al., 2012]. To what extent correlation between gyrification and tractography based parcellation is a product of methodological peculiarities or reflects neuroanatomical reality remains to be investigated.

Advantages and Limitations of Hierarchical Clustering

In this work, we propose a hierarchical clustering method for the analysis of high-resolution, whole-brain
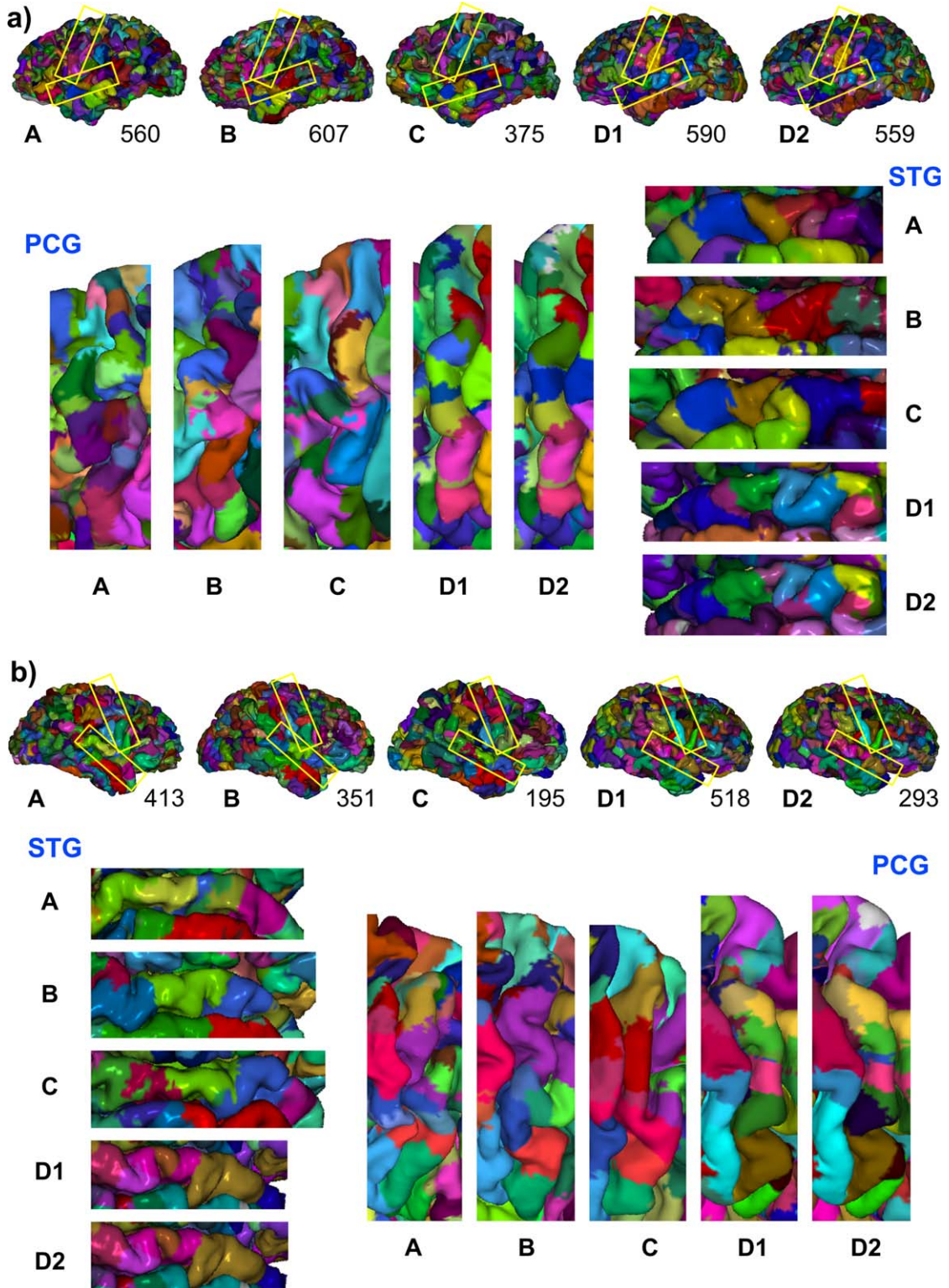
Figure 9.
Partitions with maximum SS index for all subjects' left (**A**) and right (**B**) hemispheres. The top subpanels show the whole brain parcellation, the bottom subpanels zoom into the superior temporal gyrus area and the precentral gyrus. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

anatomical connectivity data that provides an optimal data compression with minimal information loss. The method uses differences in connectivity patterns for drawing a functio-anatomical map of the cortex without the need to choose a particular granularity level. This way, almost all of the information on the connectivity pattern similarities is retained and all possible parcellations of the cortical sheet are not only stored, but also related to each other in a meaningful way. While this concept is not entirely new [Blumensath et al., 2013; Guevara et al., 2011], it is the first time that is is applied to whole-brain diffusion based anatomical connectivity data. Compared with classical single-partition connectivity-based brain parcellation methods [for a review, see, Knösche and Tittgemeyer, 2011], it offers a number of advantages.

First, it is important to compare functio-anatomical maps between subjects or between different datasets of the same subject (e.g., at different ages). With single-partition parcellation, one has to chose a particular level of granularity in order to obtain a parcellation. This level of granularity can be expressed, for example, by the number of desired clusters, by the differences between or the homogeneity within clusters, or by the sizes of the clusters. All these criteria can require different values in different datasets for defining the same functio-anatomical subdivision. It is therefore difficult to obtain comparable parcellations. Moreover, there might be more than one level of granularity relevant for the comparison. Using the whole information encoded in hierarchical trees, connectivity similarity (and therefore functio-anatomical organization of the cortex) can be compared efficiently without any explicit choices on granularities. Such comparisons can be potentially used to show changes or differences in the functio-anatomical organization of the brain in a great number of settings, including disease, development, aging and cognitive abilities. The particular advantage is that one can start at a general comparison (i.e., comparing the entire trees) without making any choices or assumptions, and then gradual zoom into certain parts of the trees (i.e., comparing subtrees) and/or particular levels of detail (i.e., pruning the lower level nodes).

Second, if larger parts of the cortex or the entire brain are to be parcellated, the definition of a granularity level, as required by non-hierarchical methods, becomes quite arbitrary. Even if comparison is not the goal, it is not easy to say, how many clusters are to be expected or how big they are. Also, the magnitude of difference between parcels depends on the brain region. For example, regions near large fiber tracts, such as the arcuate fascicle, tend to exhibit higher similarity in terms of their connectivity pattern, requiring lower thresholds for parcellation. Hierarchical parcellation circumvents the granularity choice. The obtained trees can be explored interactively in order to discover the functio-anatomical organization in different brain regions. Of course, it remains an important issue to extract actual partitions of the cortex from the tree (see below).

Third, the hierarchical trees encode the interrelation between different levels of description of the functio-anatomical cortex organization, from relatively local to very global. In fact, using very high resolution MRI data one could even imagine bridging the gap between microscopic and macroscopic levels [see Heidemann et al., 2012, for an intermediate stept into that direction]. This is of particular importance, if the parcellation is used as a basis for building a connectome. If the connectome is truly, as defined by Sporns [2011], "a comprehensive structural description of the network of elements and connections forming the human brain," it essentially has to span multiple levels of detail. Using the parcels of a hierarchical parcellation as the elements of the connectome could lead to a hierarchical connectome that not only describes the brain network at different levels of detail, but also encodes the relations between these levels. Note, however, that the construction of a true connectome relies on adequacy of the employed connectivity measures in terms of the true functio-anatomical structure of the brain. Certainly, non-invasive measures based on MRI, valuable as they may be, bear significant limitations in that respect. The parcellation resulting from hierarchical clustering could also be used as initial regions for global tractography methods like the recently proposed *plausibility tracking* method [Schreiber et al., 2014].

Nevertheless, hierarchical clustering also suffers from some principled limitations. Given its iterative agglomerative nature, established mergers cannot be undone. The procedure therefore has some sensitivity to local effects and errors may propagate, missing on the global optimum, when considering specific partitions. For this reasons, in scenarios dealing with small datasets or when only a single optimal partition is desired, optimization based methods such as *k*-means or model-based methods might be more adequate. However, for large datasets and a large number of expected clusters, these other methods may lead to exploding complexity and computation power requirements in order to achieve acceptable reliability (by design in the case of model-based methods, in order to maintain stability against local effects due to initial conditions in the case of *k*-means) [Kuncheva and Vetrov, 2006; Pham et al., 2005]. For these reasons, we strongly believe that in our scenario of whole brain parcellation, the advantages that hierarchical clustering offers (namely: multiple-nested-granularity, possibility for whole-structure comparison, and scalability with dataset size) greatly compensate for its limitations.

Meta-Leaf Matching

For the comparison of any cortical map between datasets, hierarchical parcellations being no exception, it is necessary to establish a correspondence between the cortical elements. In other words, we need to decide for each element (e.g., voxel) in one dataset, what is the functio-

anatomically equivalent element in the other dataset. This is not a big issue when comparing repeated measurements of the same subject, but due to the natural anatomical variability [Thompson et al., 1996] it poses quite a challenge if we want to compare across subjects or hemispheres. Attempts to obtain such a mapping on the basis of strutural MRI have resulted in numerous linear and non-linear registration algorithms (e.g., matching of freesurfer surfaces nodes) [Roca et al., 2010], but the results are not always satisfactory, in particular if the surfaces differ in terms of number and orientation of gyri and sulci [Ono et al., 1990]. Here, this problem concerns the meta-leaf identification between trees, which was achieved by maximizing mean-tractogram similarities using a greedy algorithm. This approach relies on the assumption that the connectivity pattern is a good reflection of the functio-anatomical identity of an cortical element—the same assumption that underlies the entire connectivity-based parcellation idea. For a more detailed discussion of the justification of this assumption, see Knösche and Tittgemeyer [2011]. Our analysis showed that the meta-leaf similarity method yields meaningful comparisons between trees. However, at this stage, intersubject matching is not always stable enough to quantitatively interpret small variations in them. The leaf matching is certainly one of the current challenges of the method. It remains to be investigated whether other matching strategies, like the "Hungarian" method [Kuhn 1955], yield an improvement. In general, however, it is not likely that by improved mathematical algorithms alone this issue is going to be resolved in a satisfactory way. Instead, the very notion of functio-anatomical equivalence needs to be refined. A comprehensive and reproducible definition of the equivalence of elements in two brains would provide solid ground from which to gauge any difference in structural properties or functional organization. Such a mapping would have to be unique, that is, each element in one brain must be assigned to exactly one element in the other brain, and vice versa. Furthermore, as the leaf matching criterion has of course a profound influence of the resulting tree comparison results, it has to be biologically meaningful. In other words, only if we have good reason to compare an element in one brain to just a particular element in the other brain (and not to any other), it makes sense to interpret their differences in, for example, connectivity or cytoarchitecture. Similar connectivity to the rest of the brain is certainly a good starting point for such an equivalence criterion, but it is surely not the ultimate solution. An interesting option might be guiding mesh matching with connectivity properties, as proposed (using much smaller pattern vectors) by Cathier and Mangin [2006] or Petrovic and Zollei [2011].

Tree Comparison

The hierarchical tree allows for comparison of the whole connectivity similarity structure across measurements, and not just particular partitions, which is not possible with the other methods. Note that the tree does actually contain all possible partitions together with their mutual relationships.

This comparison measure gives us the degree by which the structure of the connectivity similarity organization varies across different measurements. More specifically, the *tCPCC* measure focuses on the actual degree of similarity between connectivity patterns, while *wTriples* measures topological similarity (for example if the region most similar to a given selected area is the same in both measurements).

Unfortunately, compared with repeated measurements, the quality of meta-leaf matching across subjects or hemispheres inevitably decreases (see above), and so does the reliability of the comparison. There might be two possible solutions to this problem: either improving the quality of the matching by using more sophisticated methods, like combining surface topology information with connectivity pattern information (although this is unlikely to boost the quality to the same level as repeated measurements), or accepting that, due to the inter-subject variability, a perfect matching at high granularities is not possible, and trying to establish suitable levels at which the matching may be done with sufficient quality (one would have to be aware that the matching results obtained are only valid at those granularities).

Extraction of Partitions

Although a hierarchical tree in its entirety comprises the joint information of all possible partitions and their mutual relations, concrete anatomical interpretation requires the generation of actual partitions. As a compromise between single partitions and the entire tree, we characterized the hierarchical structure of the trees through series of partitions at different levels of granularity. Several partition schemes were implemented. Horizontal partitioning was shown to be a good approximation of the more sophisticated *spread separation (SS)* partitioning for a given granularity level. These partitions are very stable against noise and the boundaries have a high degree of reproducibility across subjects. In order to paliate the tendency of regions of the cortex that share large common tracts to remain in a single cluster across a higher range of granularities, a minimum size-difference clustering was implemented. This method effectively extracts more homogeneous parcellations.

Calculating the *SS* index for every granularity level, we showed that for each data set there is an entire range of similarly good partitions (approximately between 50 and 200 clusters). This fact raises general concerns about the search for a single optimal partition or even a series of a few partitions. Although one is able to single out one partition with the highest information content (in some sense) of all partitions, this information might still be completely insufficient to describe the entire structure. Hence, one has to try to find ways to (approximately) represent entire

classes of parcellations in an effective manner. As each bifurcation in the tree represents the separation between two clusters (i.e., a boundary), such a technique could aim at finding the most relevant or persistent boundaries rather than entire parcellations. An idea would be to look at the branch lengths of the nodes involved. The longer the branch (in absolute value or in relation to the node height), the more stable that region is in comparison to its neighboring ones. This way, important boundaries would be mapped on the cortex, rather than entire parcellations. However, this principle needs further investigation.

The extracted partitions could be used to do a connectome-based analysis of connectivity [Hagmann et al., 2008] or as a priori partition for white matter fiber analysis [Wassermann et al., 2010]. Within each method, partitions are always fully nested. This eases the interpretation of the boundary changes from one granularity level to the next. On the other hand, in an agglomerative method the information about the fuzzyness of the changes in connectivity similarity is not as well captured as in other approaches [Cerliani et al., 2012; Gorbach et al., 2011], although it might be extracted to a limited degree from the tree topology.

Relationship to Other Multigranularity Methods

As explained above, multigranularity methods like the one proposed here offer several general advantages over single-partition methods: they yield a more exhaustive representation of the real connectivity similarity structure; they are preferable for the analysis of larger regions (up to entire hemispheres or brains), due to the expectation that different boundaries may be relevant at different levels of granularity; they facilitate comparisons between data sets; and they allow for adaptive parcellation depending on the features that we would like to emphasize. Other researchers have approached multi-granularity in different ways. For example, Kahnt et al. [2012] generated a series of k-means based parcellations from resting-state fMRI data of the orbito-frontal cortex using different numbers of expected clusters. The fundamental difference between their approach and the one proposed in the current work lies in the fact that the hierarchical tree imposes a constraint on the relationship between the different parcellations, in that finer parcellations are nested in the coarser ones. Hence, in our method any finer subdivision complements, rather than competes with, the previous parcellation. Moreover, the embedding of the parcellations into a tree structure yields immediate clues about the distinctness and stability of certain boundaries, as well as to the topological relationship between different parcellations. An effort to bring multiple k-means parcellations at different granularities into a hierarchy has been presented for fMRI co-activation data by Clos et al. [2013], where hierarchically inconsistent voxels from the clusters obtained are removed resulting in nested partitions.

The work of Gorbach et al. [2011] takes a different approach to multi-granularity by obtaining a "space" of optimal parcellations from dMRI data through an information bottleneck method, minimizing the tradeoff between data compression and information preservation. For each desired granularity, the number of clusters is determined by a Lagrange multiplier parameter and an upper boundary for the number of clusters. In their approach, while boundaries are not necessarily nested across granularities, they seem more stable. The method may have an advantage over agglomerative methods at granularity levels where changes are gradual and boundaries fuzzy. It offers a solution between nested partitions and single partitioning at multiple levels. However, computational costs also escalate for growing datasets and granularities.

In comparison, our approach tries to characterize the whole connectivity similarity information in a compact tree, which is then easy to process. As demonstrated by the high CPCC values, most information of the connectivity similarity matrix (N^2 floating points, with N being the number of tractogram seed voxels) is sucessfully encoded with only a fraction of the size ($2N$ floating points plus $2N$ integers, easily stored as an ASCII text file). Furthermore, the number of tractogram similarities that must be computed in order to obtain the tree is 3 orders of magnitude lower than that needed to compute the matrix. This is an important advantage, given that tractogram similarity computation is a costly operation, if, like in our case, all the white matter is used as target space and high resolution (1 mm) is used (amounting to more than 15×10^5 floating point operations).

However, the use of multigranularity methods does not yet solve the problem of selecting relevant partitions. Cluster number selection remains an open problem in connectivity-based clustering literature. Various solutions have been proposed to solve it, such as visual inspection of reordered connectivity matrices [Johansen-Berg et al., 2004], consistency across subjects [Ruschel et al., in press], correspondence with cytoarchitectonic maps [Anwander et al., 2007], hierarchical consistency (when using optimization methods for different numbers of expected clusters) [Clos et al., 2013], variation of information [Clos et al., 2013; Kahnt et al., 2012], information-based model selection [Gorbach et al., 2012], consistency across modalities [Kelly et al. 2012] and the tree-based methods we propose here, which are especially suitable for whole brain parcellation. The hierarchical tree method is actually open to all these approaches, while offering a much richer stock of available partitions, among which to select.

Biological Validity

Here we made a proposal how to account for the structural organization of the cortex based on anatomical connectivity measures. A key question that remains is the one for the biological relevance of the obtained results. First of

all, our method is primarily a way to represent given information in a convenient way. Hence the validity and relevance of the parcellations hinges on the appropriateness of the underlying diffusion tractography. However, on top of this, also the construction of the tree and the selection of partitions need to be evaluated.

As this is a proof-of-principle study we only offer some preliminary evaluation of the neurobiological significance of the results, for example by comparing the inferior frontal gyrus parcellation with cytoarchitectonic maps. Much remains to be done in future studies. In particular, within-subject validation will be crucial as it avoids the inevitable uncertainties of comparing different brains. For example, functional localizer tasks in fMRI experiments could be used to gauge the functional significance of parcellations [Johansen-Berg et al., 2004; Schubotz et al., 2010]. Alternatively, resting-state functional connectivity [Kelly et al., 2012] and meta-analytic co-activation studies [Clos et al., 2013] also offer promising comparison possibilities. For example, one might apply the same method to structural and functional connectivity measurements. In vivo Brodmann mapping [Bazin et al., in press] based on quantitative T1 imaging might offer another option.

Outlook

As pointed out before, this study aims at proposing a novel technology for parcellating the brain and offering initial proof-of-principle validation. Obviously, much remains to be done. First, there are a number of methodological issues that require further attention. As detailed out above, these especially involve the tree comparison technique (especially the leaf matching) and the partition extraction method.

Second, we believe that this technology can be used to build a hierarchical function-anatomical atlas or a hierarchical connectome of the brain, which of course will require a much more numerous and representative cohort of brains. Here, the issue of neurobiological validation requires substantial attention. For example, it has to be investigated to what extent features that are not easily captured by agglomerative trees, such as gradation or non-nested hierarchies, are present in the brain and how our method reacts to them.

Third, although we have conceived and used our methods for the analysis of diffusion based anatomical connectivity, they should also be useful for the study of other kinds of multidimensional data, like resting-state functional connectivity. Whole brain parcellation methods have already been successfully used for the study of resting-state fMRI signals [Blumensath et al., 2013] and our approach might also bring new insights and possibilities to these approaches.

ACKNOWLEDGMENTS

REFERENCES

Amunts K, Lenzen M, Friederici AD, Schleicher A, Morosan P, Palomero-Gallagher N, Zilles K (2010): Broca's region: Novel organizational principles and multiple receptor mapping. PLoS Biol 8:e1000489.

Amunts K, Schleicher A, Zilles K (2007): Cytoarchitecture of the cerebral cortex–more than localization. Neuroimage 37:1061–1065.

Anwander A, Tittgemeyer M, von Cramon DY, Friederici AD, Knösche TR (2007): Connectivity-based parcellation of Broca's area. Cereb Cortex 17:816–825.

Avants B, Epstein C, Grossman M, Gee J (2008): Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain. Med Image Anal 12:16–41.

Bach D, Behren ST, Garrido L, Weiskopf N, Dolan R (2011): Deep and superficial amygdala nuclei projections revealed in vivo by probabilistic tractography. J Neurosci 31:618–623.

Bansal MS, Dong J, Fernandez-Baca D (2011): Comparing and aggregating partially resolved trees. Theor Comput Sci 412: 6634–6652.

Barbas H, Rempel-Clower N (1997): Cortical structure predicts the pattern of corticocortical connections. Cereb Cortex 7:635–646.

Basser PJ, Mattiello J, LeBihan D (1994): Estimation of the effective self-diffusion tensor from the NMR spin echo. J Magn Reson B 103:247–254.

Basser PJ, Pierpaoli C (1996): Microstructural and physiological features of tissues elucidated by quantitative-diffusion-tensor MRI. J Magn Reson B 111:209–219.

Bazin PL, Weiss M, Dinse J, Schäfer A, Trampel R, Turner R: A computational framework for ultra-high resolution cortical segmentation at 7Tesla. Neuroimage (in press, doi: 10.1016/j.neuroimage.2013.03.077).

Behrens TEJ, Johansen-Berg HJ, Woolrich MW, Smith SM, Wheeler-Kingshott CAM, Boulby PA, Barker GJ, Sillery EL, Sheehan K, Ciccarelli O, Thompson AJ, Brady M, Matthews PM (2003): Non-invasive mapping of connections between human thalamus and cortex using diffusion imaging. Nat Neurosci 6:750–757.

Beyer K, Goldstein J, Ramakrishnan R, Shaft U (1999): When is "Nearest Neighbor" Meaningful? In: Database Theory—ICDT'99. Heidelberg: Springer. pp 217–235.

Blumensath T, Jbabdi S, Glasser MF, Van Essen DC, Ugurbil K, Behrens TE, Smith SM (2013): Spatially constrained hierarchical parcellation of the brain with resting-state fMRI. Neuroimage 76:313–324.

Blundell C, Teh YW, Heller KA (2010): Bayesian rose trees. Proc 26th Conf Uncert Artif Intell. pp 65–72.

Brodmann K (1909). Vergleichende Lokalisationslehre der Großhirnrinde in ihren Prinzipien dargestellt auf Grund des Zellaufbaues. Barth, Leipzig.

Cachia A, Mangin JF, Riviere D, Kherif F, Boddaert N, Andrade A, Papadopoulos-Orfanos D, Poline JB, Bloch I, Zilbovicus M, Sonigo P, Brunelle F, Regis J (2003): A primal sketch of the cortex mean curvature: A morphogenesis based approach to study the variability of the folding patterns. IEEE Trans Med Imag 22:754–765.

Cathier P, Mangin JF (2006): Registration of cortical connectivity matrices. IEEE Int Conf Comp Vis Pat Rec Workshop, CVPRW'06. pp 66–73.

Caspers S, Eickhoff SB, Geyer S, Scheperjans F, Mohlberg H, Zilles K, Amunts K (2008): The human inferior parietal lobule in stereotaxic space. Brain Struct Funct 212:481–495.

Cerliani L, Thomas RM, Jbabdi S, Siero JCW, Nanetti L, Crippa A, Gazzola V, D'Arceuil H, Keysers C (2012): Probabilistic tractography recovers a rostrocaudal trajectory of connectivity variability in the human insular cortex. Hum Brain Mapp 33: 2005–2034.

Clos M, Amunts K, Laird AR, Fox PT, Eickhoff SB (2013): Tackling the multifunctional nature of Broca's region meta-analytically: Co-activation-based parcellation of area 44. Neuroimage 83:174–188.

Cordes D, Haughton V, Carew J, Arfanakis K, Maravilla K (2002): Hierarchical clustering to measure the connectivity in fMRI resting-state data. Magn Reson Imaging 20:305–317.

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.

Critchlow DE, Pearl DK, Qian CL (1996): The triples distance for rooted bifurcating phylogenetic trees. Syst Biol 45:323–334.

Descoteaux M, Deriche R, Knösche TR, Anwander A (2009): Deterministic and probabilistic tractography based on complex fibre orientation distributions. IEEE Trans Med Imaging 28:269–286.

Farris JS (1969): On the cophenetic correlation coefficient. Syst Biol 18:279–285.

Gorbach NS, Schütte C, Melzer C, Goldau M, Sujazow O, Jitsev J, Douglas T, Tittgemeyer M (2011): Hierarchical information-based clustering for connectivity-based cortex parcellation. Front Neuroinform 5:18.

Gorbach NS, Siep S, Jitsev J, Melzer C, Tittgemeyer M (2012): Information-theoretic connectivity-based cortex parcellation. In: Mach Learn Interp Neuroimaging, NIPS. Heidelberg: Springer. pp 186–193.

Gower JC (1990): Clustering axioms. CSNA Newslett 15:2–3.

Guevara P, Poupon C, Rivière D, Cointepas Y, Descoteaux M, Thirion B, Mangin JF (2011): Robust clustering of massive tractography datasets. Neuroimage 54:1975–1993.

Hagmann P, Cammoun L, Gigandet X, Meuli R, Honey CJ, Wedeen VJ, Sporns O (2008): Mapping the structural core of human cerebral cortex. PLoS Biol 6:e159.

Halkidi M, Batistakis Y, Vazirgiannis M (2002): Clustering validity checking methods: Part II. ACM Sigmod Record 31:19–27.

Hasnain MK, Fox PT, Woldorff MG (2001): Structure-function spatial covariance in the human visual cortex. Cereb Cortex 11:702–716.

Heidemann RM, Anwander A, Feiweier T, Knösche TR, Turner R (2012): k-space and q-space: combining ultra-high spatial and angular resolution in diffusion imaging using ZOOPPA at 7T. Neuroimage 60:967–978.

Jain A, Dubes R (1988): Algorithms for clustering data. Upper Saddle River: Prentice Hall.

Jbabdi S, Woolrich MW, Behrens TEJ (2009): Multiple-subjects connectivity-based parcellation using hierarchical Dirichlet process mixture models. Neuroimage 44:373–384.

Jenkinson M, Bannister P, Brady M, Smith S (2002): Improved optimization for the robust and accurate linear registration and motion correction of brain images. Neuroimage 17:825–841.

Johansen-Berg H, Behrens TE, Robson MD, Drobnjak I, Rushworth MF, Brady JM, Smith SM, Higham DJ, Matthews PM (2004): Changes in connectivity profiles define functionally distinct regions in human medial frontal cortex. Proc Natl Acad Sci USA 101:13335–13340.

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

Kahnt T, Chang LJ, Park SQ, Heinzle J, Haynes JD (2012): Connectivity-based parcellation of the human orbitofrontal cortex. J Neurosci 32:6240–6250.

Kelly C, Toro R, Di Martino A, Cox CL, Bellec P, Castellanos FX, Milham MP (2012): A convergent functional architecture of the insula emerges across imaging modalities. Neuroimage 61: 1129–1142.

Klein A, Andersson J, Ardekani B, Ashburner J, Avants B, Chiang M, Christensen G, Collins D, Gee J, Hellier P, Song J, Jenkinson M, Lepage C, Rueckert D, Thompson P, Vercauteren T, Woods R, Mann J, Parsey R (2009): Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration. Neuroimage 46:786–802.

Knösche TR, Tittgemeyer M (2011): The role of long-range connectivity for the characterization of the functional-anatomical organization of the cortex. Front Sys Neurosci 5:58.

Kristo G, Leemans A, Raemaekers M, Rutten GJ, Gelder B, Ramsey NF (2013): Reliability of two clinically relevant fiber pathways reconstructed with constrained spherical deconvolution. Magn Reson Med 70:1544–1556.

Kuhn HW (1955): The Hungarian method for the assignment problem. Nav Res Logist Q, 2:83–97.

Kuncheva LI, Vetrov DP (2006): Evaluation of stability of k-means cluster ensembles with respect to random initialization. IEEE Trans Pattern Anal Mach Intell 28:1798–1808.

Langfelder P, Zhang B, Horvath S (2008): Defining clusters from a hierarchical cluster tree: The dynamic tree cut package for R. Bioinformatics 24:719–720.

Liu X, Zhu XH, Qiu P, Chen W (2012): A correlation-matrix-based hierarchical clustering method for functional connectivity analysis. J Neurosci Methods 211:94–102.

Lohmann G, Mueller K, Bosch V, Mentzel H, Hessler S, Chen L, Zysset S, von Cramon DY (2001): Lipsia - A new software system for the evaluation of functional magnetic resonance images of the human brain. Comput Med Imaging Graph 25: 449–457.

Mars RB, Jbabdi S, Sallet J, O'Reilly JX, Croxson PL, Olivier E, Noonan MP, Bergmann C, Mitchell AS, Baxter MG, Behrens TEJ, Johansen-Berg H, Tomassini V, Miller KL, Rushworth MFS (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.

Morgan BJT, Ray APG (1995): Non-uniqueness and inversions in cluster analysis. Appl Stat 44:117–134.

Murtagh F (1983): A survey of recent advances in hierarchical-clustering algorithms. Comput J 26:354–359.

Murtagh F (1985): Multidimensional clustering algorithms. Vienna: Physica.

Ono M, Kubick S, Albernathey CD (1990): Atlas of the cerebral sulci. New York: Thieme.

Parker GJ, Alexander DC (2005): Probabilistic anatomical connectivity derived from the microscopic persistent angular structure of cerebral tissue. Philos Trans R Soc Lond B 360:893–902.

Passingham RE, Stephan KE, Kötter R (2002): The anatomical basis of functional localization in the cortex. Nat Rev Neurosci 3:606–616.

Petrovic A, Zollei L (2011). Evaluating volumetric brain registration performance using structural connectivity information. Med Image Comput Comput Assist Interv 14:524–531.

Pham DT, Dimov SS, Nguyen CD (2005): Selection of K in K-means clustering. Proc IME C J Mech Eng Sci 219:103–119.

Rand W (1971): Objective criteria for the evaluation of clustering methods. J Am Stat Assoc 66:846–850.

Restrepo G, Mesa H, Llanos EJ (2007): Three dissimilarity measures to contrast dendrograms. J Chem Inf Model 47:761–770.

Roca P, Rivière D, Guevara P, Poupon C, Mangin J (2009): Tractography-based parcellation of the cortex using a spatially-informed dimension reduction of the connectivity matrix. Med Image Comput Comput Assist Interv 12:935–942.

Roca P, Tucholka A, Riviere D, Guevara P, Poupon C, Mangin JF (2010): Inter-subject connectivity-based parcellation of a patch of cerebral cortex. Med Image Comput Comput Assist Interv 13:347–354.

Ruschel M, Knösche TR, Friederici A, Turner R, Geyer S, Anwander A: Connectivity architecture and subdivision of the human inferior parietal cortex revealed by diffusion MRI. Cereb Cortex (in press, doi: 10.1093/cercor/bht098).

Schubotz RI, Anwander A, Knösche TR, von Cramon DY, Tittgemeyer M (2010): Anatomical and functional parcellation of the human lateral premotor cortex. Neuroimage 50:396–408.

Schreiber J, Riffert T, Anwander A, Knösche TR (2014): Plausibility tracking: A method to evaluate anatomical connectivity and microstructural properties along fiber pathways. Neuroimage 90:163–178.

Sporns O (2011): The human connectome: A complex network. Ann NY Acad Sci 1224:109–125.

Stanberry L, Nandy R, Cordes D (2003): Cluster analysis of fMRI data using dendrogram sharpening. Hum Brain Mapp 20:201–219.

Tahmasebi AM, Davis MH, Wild CJ, Rodd JM, Hakyemez H, Abolmaesumi P, Johnsrude IS (2012): Is the link between anatomical structure and function equally strong at all cognitive levels of processing? Cereb Cortex 22:1593–1603.

Theodoridis S, Koutroubas K (1999): Pattern recognition. New York: Academic Press.

Thompson P, Schwartz C, Lin R, Khan A, Toga A (1996): Three-dimensional statistical analysis of sulcal variability in the human brain. J Neurosci 16:4261–4274.

Tomassini V, Jbabdi S, Klein JC, Behrens TEJ, Pozzilli C, Matthews PM, Rushworth MFS, Johansen-Berg H (2007): Diffusion-weighted imaging tractography-based parcellation of the human lateral premotor cortex identifies dorsal and ventral subregions with anatomical and functional specializations. J Neurosci 27:10259–10269.

Vogt O (1910): Die myeloarchitektonische Felderung des menschlichen Stirnhirns. J Psych Neurol 15:221–232.

Vogt O (1911): Die Myeloarchitektonik des Isocortex parietalis. J Psych Neurol 18:379–390.

Wang H, Wang W, Yang J, Yu P (2002): Clustering by pattern similarity in large data sets. Proc ACM SIGMOD Int Conf Manage Data. pp 394–405.

Wang E, Tungaraza R, Heaynor D, Grabowski T (2013): Multi-subject connectivity-based-parcellation of human IPL using Gaussian mixture models and hidden Markov random fields. IEEE 10th Int Symp on Biomed Imag (ISBI) pp 520–523.

Wassermann D, Bloy L, Kanterakis E, Verma R, Deriche R (2010): Unsupervised white matter fiber clustering and tract probability map generation: Applications of a Gaussian process framework for white matter fibers. Neuroimage 51:228.

Yo TS, Anwander A, Descoteaux M, Fillard P, Poupon C, Knösche TR (2009): Quantifying brain connectivity: A comparative tractography study. Med Image Comput Comput Assist Interv 12: 886–893.

Zahn CT (1971): Graph-theoretical methods for detecting and describing gestalt clusters. IEEE Trans Comp 20:68–86.

Zilles K (2004): Architecture of the human cerebral cortex - Regional and laminar organisation. In: Mai JK, Praxinas G, editors. The Human Nervous System. Amsterdam: Academic Press. pp 997–1055.

Zilles K, Amunts K (2009): Receptor mapping: architecture of the human cerebral cortex. Curr Opin Neurol 22:331–339.

Zilles K, Amunts K (2010): Centenary of Brodmann's map - Conception and fate. Nat Rev Neurosci 11:139–145.

Zilles K, Palomero-Gallagher N, Schleicher A (2004): Transmitter receptors and functional anatomy of the cerebral cortex. J Anat 205:417–432.

APPENDIX

Tractogram Distance Measure

In order to perform any kind of clustering a distance measure between the object points must first be defined. Here, this distance quantifies the similarity between the connectivity patterns of two seed points. It must satisfy the properties of symmetry ($d(x,y) = d(x,y)$ for any x,y), non-negativity ($d(x,y) \geq 0$ for any x,y) and identity of indiscernibles ($d(x,y) = 0$, if $x = y$). If the triangle inequality is also satisfied ($d(x,y) \leq d(x,z) + d(y,z)$ for any x,y,z) the distance measure is also a metric.

While the Euclidean distance is one of the most commonly used ones for low-dimensional data, it does not score well for scaling patterns or very high dimensionality [Beyer et al., 1999; Wang et al., 2002].

The correlation coefficient is a convenient way to measure the dependency between two variables (linearly) and it has been previously used as a similarity measure between tractograms [Anwander et al., 2007]. Correlation as such can also produce negative values, which cannot be sensibly interpreted for spatial connectivity patterns (two uncorrelated patterns are just as dissimilar as two negatively correlated ones). That is why we modified the measure by omitting the centering. However, the fact that our tractograms contain very many zeros causes the mean values to be very small. In consequence the differences between our measure and classical Pearson's correlation are minimal.

The distance measure is then defined by

$$d(x,y) = 1 - \frac{\Sigma_i(x_i \cdot y_i)}{\sqrt{\Sigma_i(x_i^2) \cdot \Sigma_i(y_i^2)}}, \qquad (A1)$$

where x_i is the ith element of tractogram x and $\sum$ is the summation operator. The working principle in this measure is the same as in the traditional correlation, widely established, with the difference that negative correlations

are disregarded and the discerning power is focused in positive correlations, which is better suited for comparison of anatomical tracts, that have no negative linear dependencies. Same as the traditional correlation, the proposed distance measure is not a metric since it does not satisfy the triangle inequality, but this does not lead to any shortcomings in clustering. Geometrically speaking, the proposed measure relates to the scaled projection of one vector on the other, while the correlation relates to the cosine of the angle between the vectors. Both measures are closely related.

In order to render the similarity measure robust to random artifacts in the probabilistic tractography, connectivity values smaller than 0.4 (less than 100 out of 100,000 seeded particles, as visitation values are log transformed and normalized) are set to 0 prior to computing the similarity [Anwander et al., 2007]. This value was chosen in order to eliminate only minimal noise and remain conservative (as any target voxel visited by more than 0.1% of the seeded particles will be considered), but the best threshold for probabilistic tractography is still an open question in literature [Jones, 2010].

Agglomerative Hierarchical Clustering

Graph linkage methods

In the graph linkage methods, distances between clusters are calculated from the individual distances between their component elements. There are four types of these linkages, governed by the following equations:

Single linkage: $\quad d(xy, z) = \min(d(x, z), d(y, z))$ $\quad$ (A2a)
Complete linkage: $\quad d(xy, z) = \max(d(x, z), d(y, z))$ $\quad$ (A2b)
Weighted linkage: $\quad d(xy, z) = (d(x, z) + d(y, z))/2$ $\quad$ (A2c)
Average linkage: $\quad d(xy, z) = (S_x \cdot d(x, z) + S_y \cdot d(y, z))/(S_x + S_y)$

$\quad$ (A2d)

where x and y are the clusters being merged, xy is the resulting new cluster, z is a cluster not being merged at that particular step, and S_i is the size or number of elements contained in cluster i. In the single linkage method, the new distance to a third cluster will be the smallest of the two distances to that third cluster before merging [Eq. (A2a)]; in the complete linkage method, it will be the greatest of those distances [Eq. (A2b)]; in the weighted linkage method, it will be the mean of the distances of the joining clusters [Eq. (A2c)], and in the average linkage method, it will be the mean of the distances of the joining clusters weighted by the number of elements each cluster holds, in other words, the new distance will be the average of all the pairwise distances between elements contained in clusters x and y with the elements of cluster z [Eq. (A2d)] [Murtagh, 1983].

These methods require the calculation of the pairwise distances between all elements in the dataset. In our study this translates to an extremely high amount of dis-

tance calculations ($\sim 6 \times 10^9$) where each of these requires 10^6 floating point operations (the number of seed voxels per hemisphere ranges from 65,000 to 100,000 and the size of each tractogram-the number of white matter voxels-ranges from 600,000 to 800,000 points, depending on brain size).

Centroid linkage method

In this linkage, when two clusters merge, the mean tractogram of the new cluster is computed, and the new distances to the rest of the clusters are recalculated.

$\quad$ Centroid linkage: $\quad xy = \Sigma_i (S_x \cdot x_i + S_y \cdot y_{i)}/(S_x + S_y)$ $\quad$ (A3)

where the symbols have the same meaning as in the graph methods. In principle, this involves an extra computing effort, as the new distances that must be calculated in every merging step involve high-dimensional mean tractograms. However, it can also be used to avoid the necessity of calculating the whole pairwise distance matrix by means of applying a neighborhood restriction.

Neighborhood restrictions implemented for the centroid method. With the neighborhood restriction, only tractograms with neighboring seed voxels are compared are and merged (or clusters where any of their included seed voxels are neighbors).

Several neighborhood levels may be chosen. The following neighborhoods were implemented in this study: 18 ($d_v = \sqrt{2}$), 26 ($d_v = \sqrt{3}$), 32 ($d_v = 2$), 92 ($d_v = 2\sqrt{2}$) and 124 (not defined by a value of d_v) where d_v stands for the maximum distance (in voxel units) of a neighbor voxel center from the seed voxel. A three-dimensional representation of the implemented neighborhoods can be seen in Figure A1.

As we worked with high resolution 1 mm images, there was no risk of adjacent voxels corresponding to the gm/wm interface of opposite gyri. In the case of the 92 and 124 neighborhoods, however, which expanded to nonadjacent voxels, there was a risk of considering an element as a neighbor that resides in a different gyrus. To avoid this, the algorithm was implemented as the convolution of two smaller neighborhoods kernels: 18 × 18 yields a 92 neighborhood, while 26 × 26 leads to 124 neighbors. In this sense, the smaller neighborhood was scanned, and if neighbors were detected, the respective neighborhood of each one of them was considered as well. This way, neighbors are considered as such only if they form a continuous sheet around the seed voxels. The results are analogous to what would be obtained through surface analysis with only a fraction of the cost.

Pseudocode for final centroid method (including neighborhood and initial size restrictions). Pseudocode for the final centroid algorithm is shown in below. For
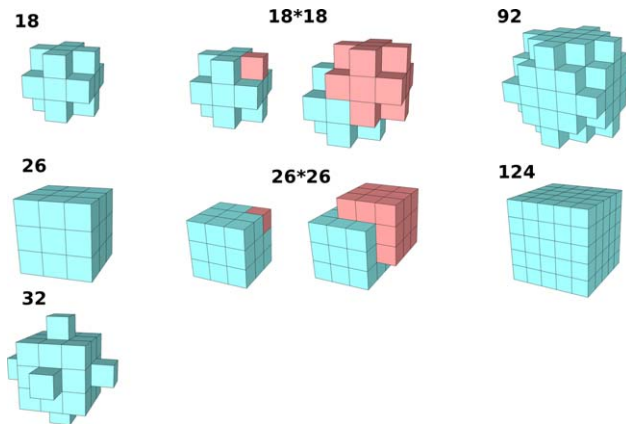
Figure A1.

Neighborhood models implemented: 18, 26, 32, 92, and 124. The 92 and 124 neighborhoods are obtained through the convolution of two 18 or 26 neighborhood kernels, respectively. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

simplicity, from this point on, this modified algorithm including neighborhood restriction and initial homogeneous merging stage will just be referred to as the centroid method or cXX where XX indicates the neighborhood level used.

```
smallestSize = 1;
activeSize = 1;

// start of homogeneous merging stage:
While ( cluster# < N )
{
   Find most similar adjacent cluster pair where
   size1 == smallestSize and size2 <= activeSize;

   If ( no pair matches the conditions )
   {
      If (there are no more clusters
          with size == smallestSize)
      {
         ++smallestSize;
         activeSize = smallestSize;
      }
      else {
         ++ activeSize;
      }
   }
   else {
      Merge clusters;
      Recalculate distances to neighbors;
   }
}

// start of unrestricted stage:
While (active clusters > 1 )
{
   Find most similar neighboring cluster pair;
   Merge clusters;
   Recalculate distances to neighbors;
}
```

Assessing the quality of the trees: the Cophenetic correlation coefficient. In order to measure the goodness of fit of the dendrograms generated (that is, how well does the dendrogram resemble the original similarity data) the Cophenetic Correlation Coefficient (*CPCC*) [Farris, 1969] was used. This measure quantifies how much information from the pairwise similarities between individual elements is present in the hierarchical tree, by calculating the degree of agreement between the distances encoded in the tree (named cophenetic distances, obtained by looking at the distance value of the merger where the desired elements are found in the same cluster for the first time) and the pairwise distances obtained from the original tractograms:

$$CPCC = \frac{\frac{1}{M}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}d_{ij}c_{ij} - \mu_D\mu_C}{\sigma_D\sigma_C}, \qquad (A4a)$$

$$\sigma_X = \sqrt{\frac{1}{M}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}d_{ij}^2 - \mu_X^2}, \qquad (A4b)$$

$$M = \frac{n(n-1)}{2}, \qquad (A4c)$$

$$\mu_X = \frac{1}{M}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}x_{ij}, \qquad (A4d)$$

where n is the total number of elements and d_{ij} and c_{ij} are the distance values between elements i and j, as computed from the tractograms or obtained from the tree, respectively (c_{ij} is then the y-axis value where the paths of leaves i and j meet in the tree). The range of *CPCC* is $[-1, 1]$. The higher the value, the better the fit between the tree and the data, a value of 1 indicating that the matrix and the tree contain exactly the same information (there is a linear dependence between both, which is not possible unless the distances between all the tractograms are equal) and a value of 0 meaning that the tree contains none of the original information (due to the nature of the hierarchical agglomerative method, negative *CPCC* values will not occur).

Choosing the Linkage Method

Pairwise tractogram distance matrices were obtained for both hemispheres of subjects A, B, and C (i.e., six datasets). Hierarchical trees were built over these matrices using each of the graph methods proposed. Trees were also built directly from the tractograms, using the centroid-neighborhood method for each of the different neighborhood levels. For the centroid trees, the number of clusters at where to stop the initial size-restricted merging stage was optimized in one of the datasets. This optimization looked to minimize information loss and provide a sufficiently high maximum granularity level while reducing tree complexity (and facilitating many steps of the tree processing). The optimized number of initial clusters was
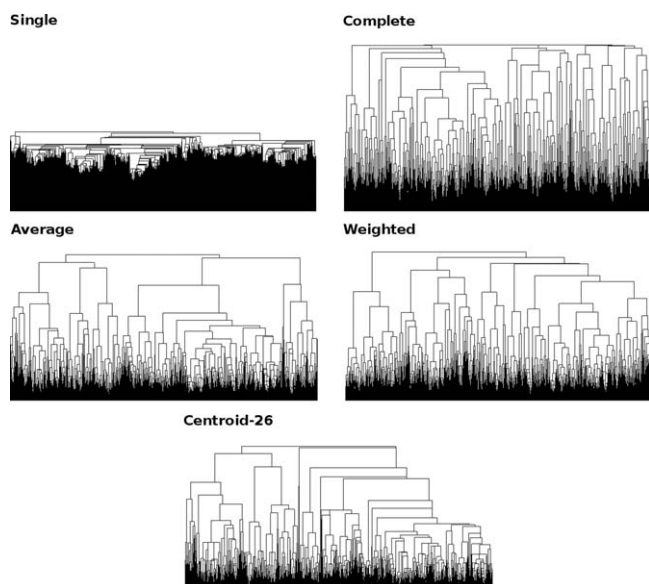
Figure A2.
Trees obtained from the left hemisphere data of subject A for each of the graph methods plus the centroid method with a 26 neighborhood. Note that a particular position on the x-axis does not identify a particular seed voxel; this may change in order to allow for the representation of the structure in tree form without any line crossings.

Figure A3.
Average *CPCC* values for trees obtained from each hemisphere of subjects A, B, and C, and from the three random tractogram sets. The first four pairs of columns refer to the single, complete, weighted, and average linkage methods. C18 to C124 refers to the centroid method with different degrees of neighborhood. See Methods section for more details. The error bars indicate the standard deviations. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

set at 5,000. This same value was applied to the remaining datasets with similar results. As will be shown below, the information loss was also minimal for the rest of the datasets. Additionally, in order to test the outcome of the tree building algorithms over unstructured data, a set of artificial tractograms (equal in number to those obtained from the real datasets) was generated in a way that they would yield a distance matrix of random values uniformly distributed between 0 and 1 (that is, a dataset without any hierarchical structure). This was achieved by creating tractograms representing points uniformly distributed over the surface of a sphere in *n*-dimensional space. However, in order to ensure this uniformity in a reasonable generation time, the dimension of the random tractograms was limited to $n = 10$. When testing the centroid method (which requires physical neighborhood information), each of the three random tractogram sets was assigned coordinates from a different real dataset.

It was not possible to detect significant differences in the overall topology of the trees obtained with the different methods by mere visual inspection, except perhaps that the distance values for the single and complete linkage methods tend to be much lower and much higher, respectively, than the ones for the other methods (Figure A2). Numerical analysis is, therefore, necessary to assess their fit to the data. For this purpose, *CPCC* values were computed for all obtained dendrograms. In order to set a baseline level for
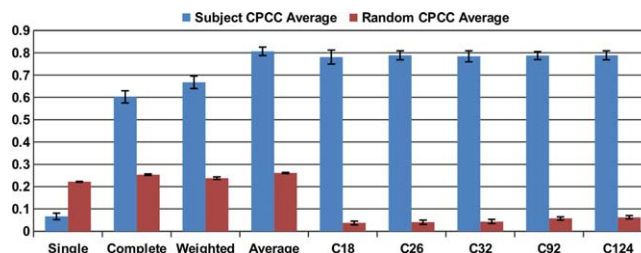
the *CPCC* values, trees were also built from unstructured datasets (using artificially generated tractograms that yield random uniformly distributed distance matrices, as explained in the previous section), and their *CPCC* values computed. The results are shown in Figure A3.

The results show that, for the real datasets the single linkage method performs worst, the complete and weighted linkage methods are not a very good match to the data either, and the average and centroid methods provide the best fit to the original data, obtaining high and very similar *CPCC* scores, with no statistically significance difference between them. Moreover, there was no significant improvement in quality using wider neighborhoods in the centroid method. In all cases the values obtained were well above their baseline levels, especially in the case of the centroid method.

Also, as a test of the information loss introduced by the homogeneous merging phase, *CPCC* values were computed for centroid trees built with equal parameters, but without merging restrictions (not shown). No significant change in the *CPCC* value was observed, meaning that the homogeneous merging stage with the selected parameter did not deteriorate the quality of the obtained tree (average *CPCC* difference was of 0.75% with a standard deviation of 0.65%).

The computational load incurred for obtaining each tree was empirically derived as the number of tractogram similarities computed, and the results are plotted in Figure A4. As can be seen, an average of $4.3 \times 10^4 \times N$ tractogram similarity operations were necessary to build up the graph linkage trees (value out of axis range), with N being the size of the dataset from which the trees were computed (the complexity of the graph methods is $N(N-1)/2$ and the datasets used are in the range of 6.5×10^4 to 10×10^4 points). On the other hand, centroid methods required only $15N$ to $50N$ operations, three orders of magnitude less than the graph methods.

It is clear from these results that from the methods considered, average and centroid linkages are the best fit to the data, with the latter having the further advantage of incurring far less computational load. Within the centroid methods, the computational load increased almost linearly with the number of neighboring voxels considered.

The 26 neighborhood centroid method (c26) was chosen as the optimal trade-off between the quality of the tree and the computational cost, and was the only method used for the remainder of the study.

Confounds and Challenges for Dendrogram Interpretation

The resulting dendrograms serve two purposes: on the one hand they are a compression of the pairwise similarities between connectional fingerprints, and on the other hand they also hold information on the similarities between clusters at every possible granularity and the hierarchical relationships between them, allowing for easy and quick partition generations. They are, however, complex structures and their interpretation and partition selection are not always straightforward. In addition, several factors might add confounds and complicate the analysis.

Artefactual datapoints

As in most types of clusterings, these can produce unwanted outliers that obscure the data and introduce errors in the analysis. In our particular case errors and spatial discontinuities in the mask of seed voxels might
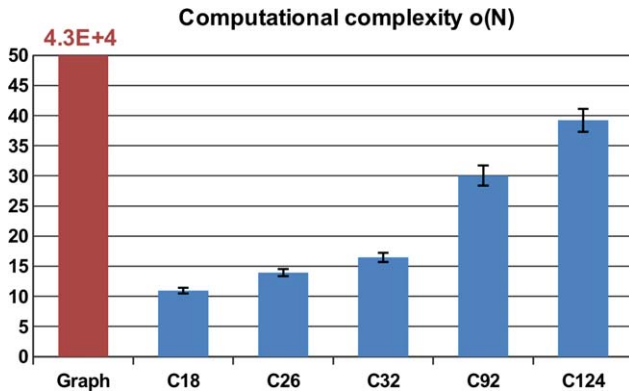

Figure A4.

Average computational complexity (expressed as the number of tractogram distance operations performed normalized by the size of the dataset N) of the tree building methods applied to the real datasets (graph linkage in red, centroid method with different neighborhood levels, 18 to 124, in blue). For interpretability, the bar for the graph linkage methods is truncated and the numerical value is indicated. Error bars show the standard deviation. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

result in unusable tractograms characterized by a very limited number of target voxels reached. This results in a very low similarity of these tractograms to the rest.

Non-monotonicity

In the most widely used linkage methods, the distance between a newly merged group of elements and the rest of the set are computed as a weighted average of the distance between elements (as in the graph methods, where the type of weighting defines the type of linkage). This means that this distance is always equal or greater than the distance between the groups that existed prior to the merge, resulting in a monotonic tree. In the centroid method, however, this is not always the case. As each group of elements is represented by a new representative centroid, this centroid could be closer to other elements than any of its components were before the merging [Morgan and Ray, 1995], which is called an inversion. In other words, it can happen that the intra-cluster distance exceeds the inter-cluster distance (see Figure A5a for a graphical clarification). These inversions or non-monotonic steps can appear when more than two points in the data have very similar distances to each other, and indicate areas with no clear binary cluster structure [Gower, 1990]. As a toy example, if we consider points in two-dimensioanl space positioned like vertices of a roughly equilateral triangle and use Euclidean distance, the centroid of two merging points will be closer to the third point than any of them were before. While these inversions do contain information about the distances encoded (when the tree is seen as a compression of the similarity matrix) they do not provide any additional information on the hierarchy structure, and they make interpretation of the hierarchy and tree analysis difficult and inconvenient [Murtagh, 1985].

Hierarchy-resolution limitation at highest granularities

The proposed method produces connectivity profiles with a very high spatial sampling resulting from seeding tractography at the white matter boundary with a voxel resolution of 1 mm. This produces an oversampling of the diffusion profiles compared with the limited spatial resolution of the diffusion acquisition and the uncertainty of the tractogram computation. As a result seed points with a very high similarity cannot be distinguished (for this reason, neighboring seed points with very high similarity are grouped together to base-areas as part of the proposed tree-building algorithm). The hierarchical relationships within these base-areas are characterized by several consecutive mergers with very small distance change indicating the non-separability of these regions and the irrelevance their internal structure for the hierarchical tree, while adding to the complexity of the tree.
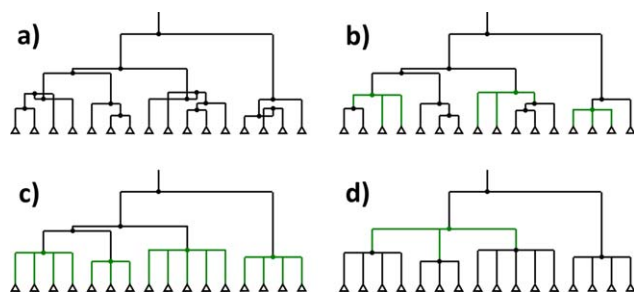
Figure A5.
Dendrogram pre-processing: example raw tree (**a**), monotonicity correction (**b**), limiting the highest granularity encoded (**c**), and collapse of non-binary structures (**d**). [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

Forced binary structure

As mentioned before, the iterative nature of the clustering process forces the dendrogram to always have binary bifurcations, whereas in reality the dataset may have structures nested in a non-binary way. This means that some of the nodes in the tree do not contribute to any real information about the similarity structure and are merely a byproduct of the pair-wise agglomerative method.

Dendrogram Preprocessing Pipeline

In order to address the aforementioned problems and ease the information extraction, the following tree preprocessing steps were developed and applied.

Outlier elimination

Isolated leaves resulting from faulty tractograms can easily be detected and eliminated without negative influence on the whole brain coverage. Data points with a distance value compared with their most similar neighbor higher than a threshold were discarded and removed from the analysis. This step was actually implemented as part of the tree building algorithm, in order to prevent the outliers from affecting the value of the centroids. Removing these outliers in general stabilizes the tree and the clustering result and simplifies its interpretation.

Monotonicity correction

As inversions occur when more than two elements are at similar distances from each other, it is possible to transform the non-monotonic trees of the centroid method into monotonic ones with little information loss. This is accomplished by merging every two nodes where an inversion occurs, creating a nonbinary branching with more than two nodes joining simultaneously into one (Figure A5b). This nonbinary structure more parsimoniously describes the original information present in the data. For each correction, the level value of the simplified node is calculated
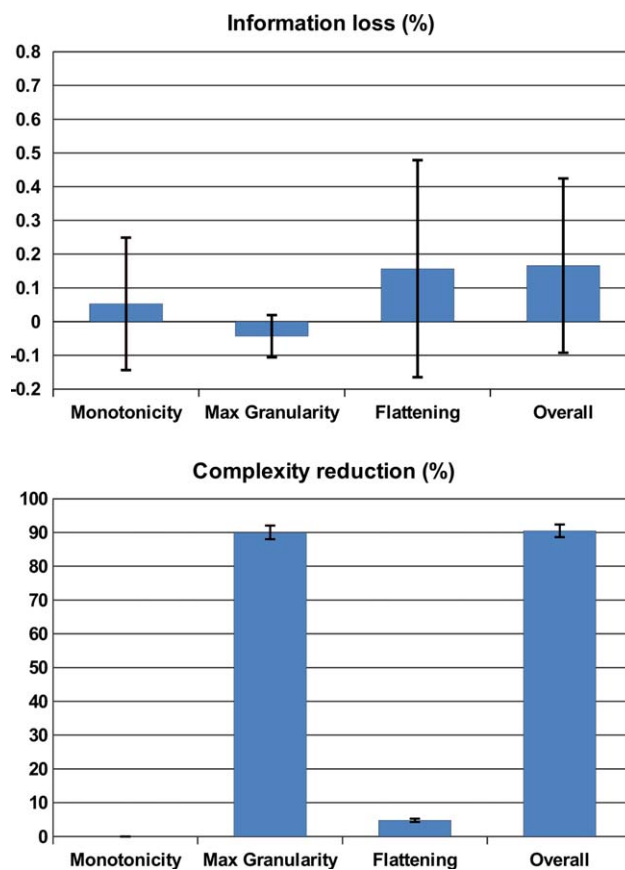

Figure A6.
Average tree information loss (top) and complexity reduction (bottom) of each step in the pre-processing pipeline, relative to the status before applying that particular step. The last column of each chart represents the overall added effect of the complete pipeline. Information loss is measured as being the relative decrease in *CPCC* index value in each step. Complexity reduction is measured as being the relative number of inner nodes eliminated. Error bars show the standard deviation. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

as the mean of the levels of the original nodes, weighted by their respective sizes in terms of number of leaves. Corrections are applied starting at the root node and working through the tree down to the leaf level.

Limiting maximum granularity

In terms of tree processing, the small differences between the leaves in the base-areas are ignored and the tree is transformed in a so-called rose tree, where the meta-leaves branch into single voxels (leaves, Figure A5c). The partition defined by these meta-leaves would then represent the maximum effective granularity achievable from the data. While rose-trees can be computed directly from data [Blundell et al., 2010], the computation costs are far greater than with the method proposed here.
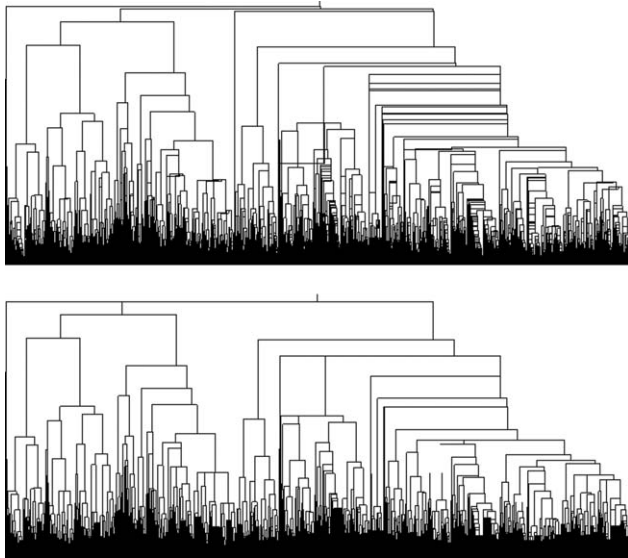
Figure A7.

Tree corresponding to the connectivity structure of the left hemisphere of subject A before (top) and after (bottom) tree preprocessing.

In our implementation, the meta-leaves are the homogenous clusters obtained during the first stage of the proposed centroid algorithm. All branchings within those nodes are then eliminated and their contained data points joined simultaneously at the original node level. Additionally, this grouping sharpens the connectivity profiles of the meta-leaves and allows for a better identification of connectivity similarities and differences between neighboring regions.
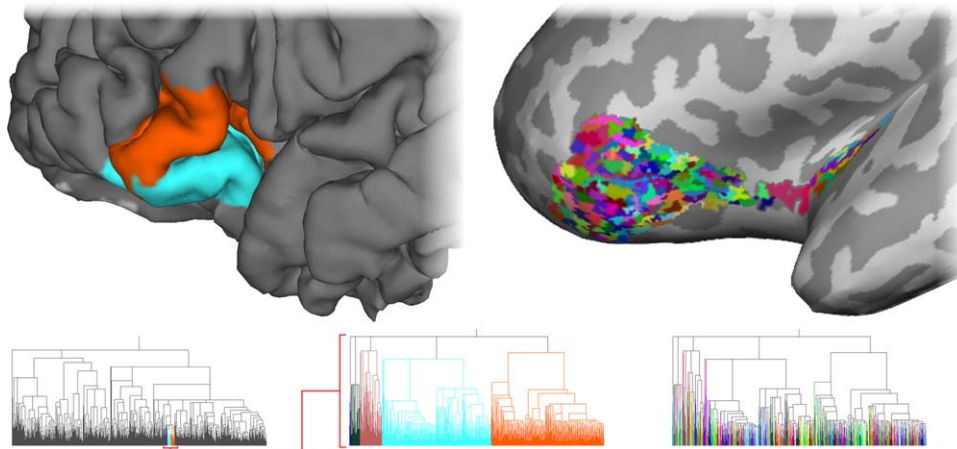
Collapse of non-binary structures

Cases where non-binary structures are present in the data are generally characterized in the tree by merges where the distance change is much smaller than the absolute distance level of the nodes being merged (when not resulting in an inversion). The dependency on the distance level accounts for the fact that the significance of distance change is the lower the higher a node stands in the tree hierarchy. A similar leveling concept to the one used with the non-monotonic steps was used here, flattening any merging with a distance change smaller than a certain proportion of the absolute distance value of the node considered. Constant and square dependencies were also considered, but the linear solution proved the best trade-off between complexity reduction and information loss. The resulting tree will be a better representation of the original data and will have a considerably reduced number of internal nodes, making it easier to identify natural divisions in the data (Figure A5d).

The preprocessing methods described in this section effectively reduce the number of branchings, which in turn reduces the tree complexity and possible confounds in the dendrogram, while still maintaining maximum usable information (shown quantitatively in the *Results* section). This also facilitates the task of the information extraction algorithms, which is introduced below.

Effects of Dendrogram Preprocessing

The tree preprocessing steps described above were applied to the c26 dendrograms of each hemisphere from subjects A, B and C. Parameter values were optimized for

Figure A8.

Detail of the clusters contained by the sub-tree covering the IFG region of the left hemisphere of subject A (upper left) and its position in the complete tree (lower left) as well as a view of the zoomed-in sub-tree (lower center). The meta-leaves contained in the mentioned sub-tree have been projected onto the inflated surface (upper right) and the zoomed-in sub-tree (lower right). [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

one of the datasets by testing multiple values and selecting those who performed best, achieving further complexity reduction without significantly adding any information loss. The optimized parameters were then applied to the remaining datasets and a similar effectiveness verified.

Firstly, those data points with distances greater than 0.1 to their nearest neighbor were considered as outliers and excluded, resulting in a rejection of an average of 0.5% of the data points (this step is actually integrated into the tree building process). Next, non-monotonicity was corrected and the maximum granularity was limited by merging all inner nodes of the 5,000 homogeneous sub-trees obtained during the first phase of tree construction, effectively transforming these nodes into non-binary meta-leaves; non-binary structures at all levels of the tree were detected and flattened using a parameter of $l = 0.05$ (nodes with branches shorter than 5% of the node height were eliminated). These parameters were empirically selected in order to obtain additional complexity reduction at higher levels of the tree (measuring complexity as the number of branchings or inner nodes in the tree) while keeping the total information loss in the same order range ($<1\%$). In order to quantitatively assess the complexity reduction and the information loss caused by the pre-processing, inner node count and CPCC values were obtained for the trees at each processing step, and their relative changes in relation to the previous states were evaluated (Figure A6). The results show that neither of the first two steps (monotonicity correction and limiting of maximum granularity) significantly reduced the amount of information contained in the trees, while the second step achieved a complexity reduction of almost 90%. The third step (flattening of non-binary structures) further reduced the complexity by 5%, while introducing an average of 0.2 % of information loss (without statistical significance). Overall, the whole pre-processing pipeline achieved a complexity reduction of more than 90% with a loss of information of less than 0.5% (0.15% on average), making it a remarkably efficient and useful tool for improving the performance of partition finding and tree comparison algorithms. It can also ease interpretation of the trees through visual inspection, although this still remains a challenging task. Visual changes on tree structure caused by the pre-processing are exemplified in Figure A7. An example of the obtained meta-leaves in one of the subjects is shown in Figure A8.

Measures for Dendrogram Comparison

Tree cophenetic correlation coefficient (tCPCC)

As different meta-leaves may have different sizes (in the sense of containing a different number of seed voxels), the CPCC factor [Farris 1969, Eq. (A4)] was modified in order to include a weighting with cluster size. This way the relevance of the distance value between two meta-leaves was proportional to the fraction of the total seed voxels contained in them. The mathematical formula for the tCPCC resulted as follows:

$$tCPCC = \frac{\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} x_{ij} S_{x_{ij}} y_{ij} S_{y_{ij}} - \mu_{X'}\mu_{Y'}}{\sigma_X \sigma_Y \sum_{i=1}^{n-1}\sum_{j=i+1}^{n}(S_{x_{ij}} S_{y_{ij}})}$$

$$\text{with } \sigma_X \sqrt{\frac{\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}(x_{ij}S_{x_{ij}})^2 - \mu_{x'}^2}{\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} S_{x_{ij}}^2}} \quad (A5)$$

$$\text{and } \mu_{X'} = \frac{\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} x_{ij}S_{x_{ij}}}{\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} S_{x_{ij}}}$$

where x_{ij} is the distance between meta-leaves i and j as encoded in tree X and Sx_{ij} is the sum of the sizes of meta-leaves i and j for tree X.

As with the tCPCC, a value of 1 would indicate that the distance values between single meta-leaves encoded by both trees are linearly dependent (meaning that both trees contain the same information encoded in their distance values), and a value of 0 means that the trees do not share any common information.

Weighted triples similarity (wTriples)

As with the tCPCC, a weighting was included into the basic formula (Bansal et al., 2011] to account for meta-leaf size, and the final formula was expressed as:

$$wTriples = \frac{\sum_{i=1}^{n-2}\sum_{j=i+1}^{n-1}\sum_{k=j+1}^{n} f_{top}(i,j,k)}{\sum_{i=1}^{n-2}\sum_{j=i+1}^{n-1}\sum_{k=j+1}^{n} S_{x_{ijk}} + S_{y_{ijk}}} \quad (A6)$$

$$\text{with } f_{top}(i,j,k)$$

$$= \begin{cases} S_{x_{ijk}} + S_{y_{ijk}} & \text{if topology (joining order) matches.} \\ 0 & \text{if topology does not match.} \end{cases}$$