

## Using connectomics for predictive assessment of brain parcellations

Kristoffer J. Albers<sup>a,1</sup>, Karen S. Ambrosen<sup>a,b,1</sup>, Matthew G. Liprot<sup>a,1</sup>, Tim B. Dyrby<sup>a,b</sup>,  
Mikkel N. Schmidt<sup>a</sup>, Morten Mørup<sup>a,a,\*</sup>

<sup>a</sup> Department of Applied Mathematics and Computer Science, Technical University of Denmark, Richard Petersens Plads, Building 324, DK-2800 Kgs. Lyngby, Denmark

<sup>b</sup> Danish Research Centre for Magnetic Resonance, Centre for Functional and Diagnostic Imaging and Research, Copenhagen University Hospital Amager and Hvidovre, Copenhagen, Denmark

### ARTICLE INFO

#### Keywords:

Brain parcellation  
Diffusion magnetic resonance imaging (dMRI)  
Functional magnetic resonance imaging (fMRI)  
Whole brain connectivity  
Human connectome  
Link prediction

### ABSTRACT

The organization of the human brain remains elusive, yet is of great importance to the mechanisms of integrative brain function. At the macroscale, its structural and functional interpretation is conventionally assessed at the level of cortical units. However, the definition and validation of such cortical parcellations are problematic due to the absence of a true gold standard. We propose a framework for quantitative evaluation of brain parcellations via statistical prediction of connectomics data. Specifically, we evaluate the extent in which the network representation at the level of cortical units (defined as parcels) accounts for high-resolution brain connectivity. Herein, we assess the pertinence and comparative ranking of ten existing parcellation atlases to account for functional (FC) and structural connectivity (SC) data based on data from the Human Connectome Project (HCP), and compare them to data-driven as well as spatially-homogeneous geometric parcellations including geodesic parcellations with similar size distributions as the atlases. We find substantial discrepancy in parcellation structures that well characterize FC and SC and differences in what well represents an individual's functional connectome when compared against the FC structure that is preserved across individuals. Surprisingly, simple spatial homogenous parcellations generally provide good representations of both FC and SC, but are inferior when their within-parcellation distribution of individual parcel sizes is matched to that of a valid atlas. This suggests that the choice of fine grained and coarse representations used by existing atlases are important. However, we find that resolution is more critical than the exact border location of parcels.

### 1. Introduction

The vast complexity of the human brain [Braitenberg and Schüz \(1991\)](#); [Murre and Sturdy \(1995\)](#) and the incomplete and noisy measurements available through neuroimaging modalities require a pragmatic approach to the analysis of the human connectome [Hagmann \(2005\)](#); [Sporns et al. \(2005\)](#). Segregation into anatomical or functional units provides interpretable and, in principle, noise-reduced network nodes whose inter-connections approximate the brain's organizational structure [Bullmore and Sporns \(2009\)](#); [Sporns \(2012\)](#). Much research is underway to delineate the structural and functional organization of the human brain [Arslan et al. \(2018\)](#); [Glasser et al. \(2016\)](#); [Smith \(2013\)](#); [Van Essen et al. \(2013b\)](#) but it remains unclear which parcellation best accounts for such organization and how this is quantified.

To provide a sound basis for analysis, the nodes provided by a given parcellation method must be robust across individuals, and fully represent the local infrastructure, microscopical properties and

connectional “fingerprint” — the unique pattern of inputs and outputs [Passingham et al. \(2002\)](#). For example, when defining cortical regions at the macroscale it has been suggested that specific functions of the areas, such as connectivity, reproducibility, convergence, multimodality, evolutionary coherence, and inter-subject variability, should all be taken into account [Amunts and Zilles \(2015\)](#); [Arslan et al. \(2018\)](#). However, despite these efforts there still remains a lack of gold standard evaluation strategies against which any particular parcellation can be tested.

The exact method of parcellation employed depends upon the application. Hence a wide variety of parcellation schemes are currently available. This includes cortical surface morphology ([Desikan-Killiany Desikan et al. \(2006\)](#), [Destrieux Fischl et al. \(2004\)](#), [AAL Gong et al. \(2009\)](#); [Tzourio-Mazoyer et al. \(2002\)](#)), functional activation ([Gordon Gordon et al. \(2014\)](#), [Power Power et al. \(2011\)](#), [Yeo Yeo et al. \(2011\)](#), [Baldassano Baldassano et al. \(2015\)](#), [Shen Shen et al. \(2013\)](#), [Craddock Craddock et al. \(2012\)](#), [Arslan Arslan et al. \(2015\)](#)), parcellations derived from structural connectivity (SC) [Ambrosen et al. \(2014\)](#); [Baldassano et al. \(2015\)](#);

\* Corresponding author.

E-mail address: [mmor@dtu.dk](mailto:mmor@dtu.dk) (M. Mørup).

<sup>1</sup> These authors contributed equally to this work

Parisot et al. (2016), and combinations thereof including Brainnetome Fan et al. (2016) (cortical surface morphology and SC), and HCP\_MMP1.0 Glasser et al. (2016) (function- and structure-related features). The various parcellation schemes exhibit considerable differences, e.g. number of parcels and parcel border locations. It thus remains a challenge to understand the relationship between different parcellation schemes Bohland et al. (2009), and no single parcellation appears to be universally accepted Arslan et al. (2018). The situation is exacerbated by studies showing that subsequent graph measures are sensitive to the chosen parcellation scheme, both for structural Hagmann et al. (2008); Wu et al. (2019) and functional Arslan et al. (2018); Fornito et al. (2010); Zalesky et al. (2010) analyses.

The differences between parcellations in terms of the size and spatial extent of their cortical units, and also their impact upon subsequent analyses, illustrate that it is important to validate their relevance to the application in question. Whilst reliability is often purported as a proxy for validation Dornas and Braun (2018); Fan et al. (2016); Glasser et al. (2016); Thirion et al. (2014), it is not sufficient because a method can be arbitrarily reliable yet poorly account for brain organization. In contrast, the use of data on brain organization that is independent of how a parcellation is derived can in principle permit validation of a parcellation. Thus, strategies investigating correspondence with neuroimaging data on task activation, cytoarchitecture, and myelination as well as characterizing network topology and their population differences have been considered. See Arslan et al. (2018) for a recent extensive review and evaluation of these strategies using data from the Human Connectome Project (HCP).

We propose a new alternative evaluation procedure for quantifying the quality of parcellations using statistical prediction based upon brain connectivity data. In contrast to cluster validity measures that quantify similarity within the cluster to the similarity of other clusters Arslan et al. (2018); Dornas and Braun (2018) or statistical prediction based on feature maps or time-series Thirion et al. (2014); Wang et al. (2018), our proposed approach quantifies how well the network organization is preserved within the representation induced by a parcellation. In particular, our statistical prediction framework poses quantification of parcellation quality as a link-prediction problem Ambrosen et al. (2014); Andersen et al. (2014); Clauset et al. (2008); Liben-Nowell and Kleinberg (2007). A parcellation is thereby assessed by its ability to characterize brain connectivity data that may derive from an orthogonal modality to that from which the parcellation was defined. In particular, the approach quantifies how well network structure is preserved in the considered modality by the network organization induced by the parcellation. The proposed assessment procedure relies on 1) the parcellation evaluated being independent of the brain connectivity data used for the evaluation, and 2) that there is no systematic biases in the brain connectivity data. It can be reasonable to assume that some modalities are independent of others, i.e., a parcellation derived from structural data may be independent of connectomes derived from functional data and vice versa. However, assuming 2) that there is no systematic biases in the estimated connectomes is not only difficult to test but also not expected to hold true in general. Our assessment procedure can therefore not be used as a gold standard for objectively evaluating which parcellation representation provides the best characterization of brain organization. However, it can be used to assess which parcellations best account for a given connectomics dataset.

Herein, we have used high-resolution SC and FC data from the Human Connectome Project (HCP) Feinberg et al. (2010); Moeller et al. (2010); Setsompop et al. (2012); Van Essen et al. (2012); Xu et al. (2012) to validate the suite of ten prominent parcellations available from Arslan et al. (2018). We contrast the predictive performance of these atlases to spatially-homogeneous clustering procedures including the geometric parcellation considered in Thirion et al. (2014) that do not comply with anatomical information beyond spatial adjacency. We further contrast performance to geodesic clustering constrained to parcel size distributions similar to those of the considered atlases.

The overall aim of this study is to quantify representations that most adequately account for the organization of high resolution structural and functional connectomes obtained from non-invasive MRI measurements. We first verify the predictive validation framework on synthetic data in which connectomes are generated according to ground truth parcellation structures and the predictive evaluation framework used to recover the most plausible parcellation. Subsequently, using our predictive validation framework we systematically evaluate:

- The comparative ranking of existing prominent atlases in accounting for inter subject SC as well as inter and intra subject FC.
- The relative performance of data driven parcellations considering their different levels of resolutions.
- Performances by simple baseline parcellations obtained from geometric clustering of vertex locations (i.e., uninformed by anatomy, FC, and SC) considering both Euclidean distances and geodesic distances accounting for surface structure.
- The impact of parcel size distributions on performance.

## 2. Materials and methods

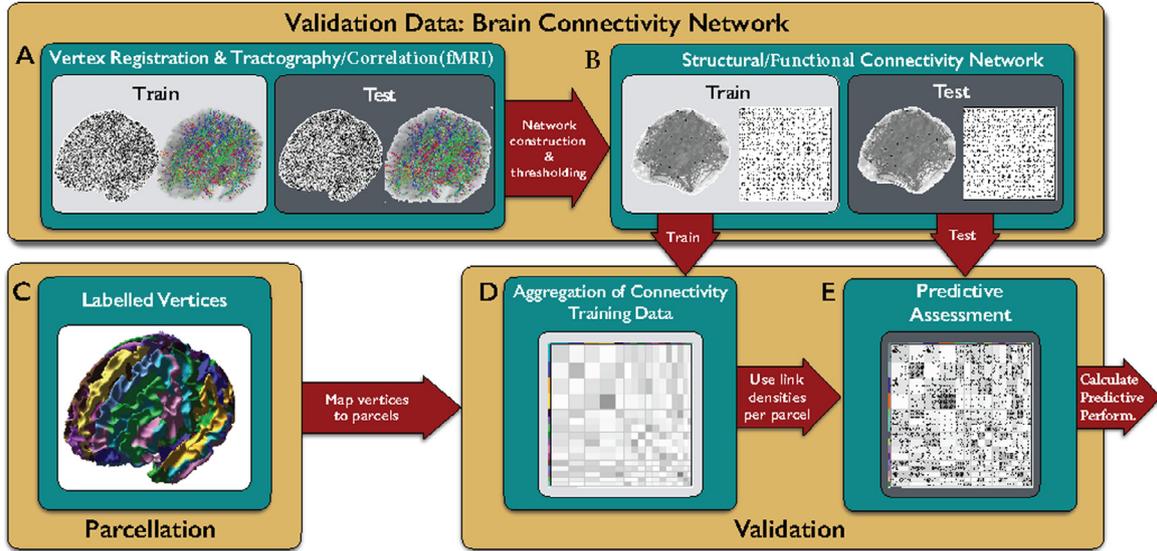
The MRI data used in the preparation of this work were obtained from the WU-Minn Human Connectome Project (HCP) database (<https://ida.loni.usc.edu/login.jsp>) in the "500 subjects" release Feinberg et al. (2010); Moeller et al. (2010); Setsompop et al. (2012); Van Essen et al. (2012); Xu et al. (2012). A total of 250 random subjects were used.

### 2.1. Diffusion imaging, tractography and construction of structural connectivity graphs

Acquisition parameters are described in full for dMRI in Feinberg et al. (2010); Moeller et al. (2010); Setsompop et al. (2012); Sotiropoulos et al. (2013); Xu et al. (2012), and for the structural scans in Milchenko and Marcus (2013), and are listed in brief here. The dMRI was acquired with a multiband factor of 3, covering 270 directions distributed over 3 diffusion shells of b-values 1000, 2000 and 3000 s/mm<sup>2</sup>, plus 18  $b = 0$  (non-diffusion weighted) scans. The nominal voxel size was 1.25 mm isotropic. Both  $T_1$ -weighted and  $T_2$ -weighted structural scans at 0.7 mm isotropic resolution were also acquired.

All pre-processing of the data, including correction for sequence-dependent artefacts such as eddy-current distortion, was performed by the "minimal preprocessing pipeline" provided by the HCP project Glasser et al. (2013). This included the generation of native pial and white-matter surfaces, and their coregistration to a standard vertex mesh. This provides a one-to-one correspondence between the surface vertices of every subject, and hence permits vertex-wise analysis of tractography results across the HCP population.

Tractography was performed using a GPU implementation of FSL's BedpostX Hernández et al. (2013) and ProtrackX2 Behrens et al. (2007, 2003b); Jenkinson et al. (2012). BedpostX parameters included a specification of up to 3 fibres per voxel, and a deconvolution model using zeppelins Behrens et al. (2007, 2003b). ProtrackX2 was run in "matrix3" mode, with all voxels in the white matter (as specified by a structural imaging mask) as seed points. The GM-WM surface boundary and all subcortical grey-matter voxels were specified as tractography target masks. Streamlines were kept and entered into the resultant connectivity matrix if they succeeded in traversing opposite directions from a seed voxel and reaching two different target surface vertices or subcortical voxels. One thousand streamlines were generated from every seed voxel as used in Ambrosen et al. (2014). The result of tractography is therefore a symmetric connectivity matrix of size  $[(\text{number of target surface vertices}) + (\text{number of target subcortical voxels})]^2$ . In this study, only the surface vertices are analysed. Thus, SC graphs ( $n = 59,412$  vertices) covering the cerebral cortex of both hemispheres were generated.



**Fig. 1.** Illustration of the proposed predictive validation framework. A) The native surfaces of all subjects are co-registered to a standard vertex mesh to obtain one-to-one correspondence between the surface vertices of every subject Glasser et al. (2013). DMRI data: Tractography is performed between all vertices of the surface by initialising 1000 streamlines in all white-matter voxels resulting in a weighted symmetric SC network for each subject. fMRI data: pairwise correlation is calculated between all vertices. B) The networks are thresholded to obtain binary links (connections in left panel, dots in right panel). C) The considered parcellation. D) The training networks are permuted according to the parcellation  $\mathbf{z}$  in question and the aggregated statistics used to define the link densities  $\eta$  between and within parcels calculated according to the training network. E) The predictive performance is assessed by calculating the predictive performance using these link densities defined according to the training network (grey background) to predict the links of the test network (overlaid dots).

## 2.2. Construction of resting state functional connectivity graphs

Resting state functional connectivity data were obtained from the structurally denoised ICA-FIX cleaned versions for the 250 subjects. For further details see Griffanti et al. (2014); Salimi-Khorshidi et al. (2014); Smith et al. (2013). The graphs were formed by averaging the Pearson correlation matrix, estimated using both of the sessions that were acquired for each individual subject. Within each session, data were acquired with both left-right and right-left phase encoding directions, resulting in an averaging of four correlograms per subject, each estimated from 1200 time frames. For SC we only consider inter-subject prediction as we have one scan available pr. subject. However, as we have two sessions for the functional data we also consider within subject performance using the estimated functional connectome of the first session as training data to predict the functional connectome of the second session as test data.

## 2.3. Functional and structural connectomes

Each subject's SC and FC graphs were binarised by thresholding at the value closest to forming a connectivity density of 1%, forming a symmetric binary  $n \times n$  adjacency matrix  $\mathbf{A}$  such that  $A_{ij} = 1$  and  $A_{ij} = 0$  respectively denotes the existence or absence of a structural or functional connection in either direction between  $i$  and  $j$ . Here  $i$  and  $j$  are used to represent the nodes at the vertex level.

## 2.4. The predictive assessment framework

Fig. 1 outlines the proposed predictive assessment procedure. Input to the procedure is a parcellation  $\mathbf{z}$  ( $z_i = m$  indicates that node  $i$  belongs to parcel number  $m$ ) and the SC (or FC) networks of the training and test data to be assessed. From the training network, the connectivity density between parcel  $l$  and  $m$ , defined by  $\eta_{lm}$ , is estimated and used to predict the connectivity structure between parcel  $l$  and  $m$  in the test network.

Let  $N_{lm}^+ = \frac{1}{1+\delta_{l=m}} \sum_{i \neq j} A_{ij}^{(train)} \delta_{z_i=l} \delta_{z_j=m}$  and  $N_{lm}^- = \frac{1}{1+\delta_{l=m}} \sum_{i \neq j} (1 - A_{ij}^{(train)}) \delta_{z_i=l} \delta_{z_j=m}$  respectively be the number of (aggregated) links and non-links in the training networks between nodes in cluster  $l$  and nodes

in cluster  $m$ , and  $\delta$  is the Kronecker delta function. Assuming that each link in the graph is generated independently given the within and between parcel densities ( $\eta_{lm}$ ), we have that

$$p(\mathbf{A}^{(train)} | \eta, \mathbf{z}) = \prod_{i>j} \eta_{z_i, z_j}^{A_{ij}} (1 - \eta_{z_i, z_j})^{(1-A_{ij})} \quad (1)$$

$$= \prod_{l \geq m} \eta_{lm}^{N_{lm}^+} (1 - \eta_{lm})^{N_{lm}^-}. \quad (2)$$

We will use Bayesian inference in order to robustly infer  $\eta_{lm}$ . For the Bernoulli likelihood above the conjugate prior is the beta distribution defined as  $Beta(\alpha, \beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta^{\alpha-1} (1-\theta)^{\beta-1}$  in which  $\Gamma(x)$  is the gamma function (which notably for integer values corresponds to the factorial function  $(x-1)!$ ). Imposing this conjugate prior we obtain for the posterior distribution

$$p(\eta_{lm} | \mathbf{A}^{(train)}, \mathbf{z}) = \frac{p(\mathbf{A}^{(train)} | \eta_{lm}, \mathbf{z}) p(\eta_{lm})}{\int p(\mathbf{A}^{(train)} | \eta_{lm}, \mathbf{z}) p(\eta_{lm}) d\eta_{lm}} \quad (3)$$

$$= Beta(N_{lm}^+ + \alpha, N_{lm}^- + \beta). \quad (4)$$

Let  $\langle x \rangle = \int x p(x) dx$  denote the expectation of  $x$  with respect to distribution  $p(x)$  and  $x^{(MAP)} = \arg \max_{\eta_{lm}} p(\eta_{lm} | \mathbf{A}^{(train)}, \mathbf{z})$  denote the (point) estimate maximizing the posterior. Accordingly, we have

$$\begin{aligned} \eta_{lm}^{(MAP)} &= \frac{N_{lm}^+ + \alpha - 1}{N_{lm}^+ + N_{lm}^- + \alpha + \beta - 2}, \\ \log \langle \eta_{lm} \rangle &= \log(N_{lm}^+ + \alpha) - \log(N_{lm}^+ + N_{lm}^- + \alpha + \beta), \\ \log \langle (1 - \eta_{lm}) \rangle &= \log(N_{lm}^- + \beta) - \log(N_{lm}^+ + N_{lm}^- + \alpha + \beta), \\ \langle \log(\eta_{lm}) \rangle &= \psi(N_{lm}^+ + \alpha) - \psi(N_{lm}^+ + N_{lm}^- + \alpha + \beta), \\ \langle \log(1 - \eta_{lm}) \rangle &= \psi(N_{lm}^- + \beta) - \psi(N_{lm}^+ + N_{lm}^- + \alpha + \beta), \end{aligned}$$

where  $\psi$  is the digamma function. Notably, the prior provides robust estimates providing non-zero statistics alleviating potential issues of  $\log(0)$  and  $\psi(0)$ . These parcellation-dependent connectivity statistics are then used to predict the SC (or FC) of the test graphs. We presently consider

three prominent measures of predictive performance namely area under curve of the receiver operator characteristics [Clauset et al. \(2008\)](#); [Miller et al. \(2009\)](#), predictive log-likelihood ( $L$ ) [Palla et al., 2014](#) and predictive log-loss ( $LL$ ) [Ambrosen et al. \(2013\)](#). Specifically, using AUC,  $L$ , and  $LL$  we investigate the extent to which the between-parcel densities estimated using the training-graph are able to predict the occurrences of links and non-links in the test-graph. Each of these three measures are outlined below.

#### 2.4.1. Area under curve of the receiver operator characteristic

A common metric to quantify predictive performance when predicting links in networks is the area under curve (AUC) of the receiver operator characteristics (ROC) curve [Clauset et al. \(2008\)](#); [Miller et al. \(2009\)](#). The ROC tracks the ratio of false positive (x-axis) and true positive (y-axis) for changing thresholds of predicting links. Notably, the AUC of the ROC is invariant to class imbalance as the x- and y-axis of the ROC curve respectively are normalized by the total number of non-links and links. We use the maximum a posteriori estimate,  $\eta^{(MAP)}$ , defining the most plausible inter-parcel link density as scores for quantifying the presence or absence of links, and contrast the actual test connectivity  $A_{ij}^{(test)}$  to the corresponding density score obtained from the training network  $\eta_{z_i, z_j}^{(MAP)}$ . Specifically, we impose in the estimate of  $\eta_{z_i, z_j}^{(MAP)}$  a uniform prior  $\alpha = \beta = 1$  such that  $Beta(1, 1) = 1$ . As a result,  $\eta_{z_i, z_j}^{(MAP)} = \frac{N_{z_i, z_j}^+}{N_{z_i, z_j}^+ + N_{z_i, z_j}^-}$  corresponds to the maximum likelihood of the connectivity density, i.e.,  $\arg \max_{\eta_{lm}} p(A^{(train)} | \eta_{lm}, \mathbf{z})$ . Notably, the AUC score does not rely on log or  $\psi$  functions that need to be estimated robustly which enables the estimate for this choice of prior to solely rely on the empirically observed connectivity density. An AUC score of 0.5 indicates that the scoring procedure is no better than chance, whereas an AUC score of 1 indicates that a threshold value of the scoring procedure exists which provides a perfect separation of links from non-links in the test graph  $A^{(test)}$ .

#### 2.4.2. Predictive likelihood

We further consider the predictive log-likelihood ( $L$ ). In the analysis we impose a non-informative Jeffrey's prior using  $\alpha = \beta = 0.5$  (for a discussion of choices for the prior see also [Zhu and Lu \(2004\)](#)).

$$L = \log \prod_{i>j} \int P(A_{ij}^{(test)} | \eta, \mathbf{z}) P(\eta | A^{(train)}, \mathbf{z}) d\eta \quad (5)$$

$$= \sum_{i>j} A_{ij}^{(test)} \log(\langle \eta_{z_i, z_j} \rangle) + (1 - A_{ij}^{(test)}) \log(\langle 1 - \eta_{z_i, z_j} \rangle). \quad (6)$$

This quantifies how probable it is to observe the test graph  $A^{(test)}$  given the expected probability of observing links and non-links within and between parcels in  $\mathbf{z}$  according to the training graph  $A^{(train)}$ .

#### 2.4.3. Predictive log-loss

A related measure is the expected log-loss ( $LL$ ) in which the expectation is taken with respect to the log-likelihood of the test graph  $A^{(test)}$  thereby as opposed to the predictive likelihood taking expectations of the logarithm instead of logarithm of expectations:

$$LL = \int \log \left( \prod_{i>j} P(A_{ij}^{(test)} | \eta, \mathbf{z}) \right) P(\eta | A^{(train)}, \mathbf{z}) d\eta \quad (7)$$

$$= \sum_{i>j} A_{ij}^{(test)} \langle \log(\eta_{z_i, z_j}) \rangle + (1 - A_{ij}^{(test)}) \langle \log(1 - \eta_{z_i, z_j}) \rangle. \quad (8)$$

In the analysis we also here impose a non-informative Jeffrey's prior using  $\alpha = \beta = 0.5$ .

As the AUC is only concerned with the relative ranking of links and non-links it is expected to be more robust to outliers when used for predictive assessment than the predictive likelihood ( $L$ ) and predictive

log-loss ( $LL$ ) measures. As the difference between  $L$  and  $LL$  stems from the difference between utilizing the log function as opposed to the  $\psi$  function we expect the behavior of  $L$  and  $LL$  to be similar. However, we expect the  $LL$  to be more sensitive to outliers, i.e. links occurring in low inter-parcel density regions and non-links occurring in high density regions as  $\log(x)$  is less sensitive when compared to  $\psi(x)$  for small values of  $x$ .

Notably, the predictive framework operates in the space of the full connectome. A benefit of this is that the predictive assessment considers the same data regardless of number of parcels used by the deployed atlases. Notably, FC is often analyzed from the perspective of average within parcel time courses as opposed to the explicitly created connectome presently considered. However, no such time-courses exist for SC and thus it is natural to evaluate SC in terms of the full connectivity structure. Typically, FC has been evaluated using various measures of within cluster homogeneity measures vs. between cluster homogeneity [Arslan et al. \(2018\)](#); [Craddock et al. \(2012\)](#). Such homogeneity measures indeed rely on the evaluation of all the vertices as also used in the present evaluation rather than considering only the mean-time series. Importantly, relying solely on the derived mean time series when evaluating performance can be systematically biased by the number of parcels employed, whereas by considering the data in full such as evaluating using the entire connectome naturally enables to contrast the performance of parcels differing in their numbers. Furthermore, cluster homogeneity measures rely on some measure of similarity whereas the proposed predictive assessment directly operates on the connectivity structure of connectomes and thus does not require the specification of a node similarity measure. Finally, homogeneity measures rely on in sample estimates whereas the predictive assessment explicitly utilizes discrepancies between training and test graphs.

Importantly, the predictive assessment quantifies the extent to which the simplified representation in terms of atlas parcels is capable of preserving the connectivity structure of the full connectome. This provides a means of quantifying how well consistent information in regards to the full connectome is preserved in the simplified aggregated network representation at the level of parcels.

#### 2.5. Considered brain atlases

We applied the predictive validation framework on the precomputed group level parcellations available from [Arslan et al. \(2018\)](#), which includes atlases based on cortical surface morphology: Desikan (70 parcels) [Desikan et al. \(2006\)](#), Destrieux (150 parcels) [Fischl et al. \(2004\)](#), AAL (82 parcels) [Gong et al. \(2009\)](#); Tzourio-Mazoyer et al. (2002) (independent of SC and FC). Five based on resting-state fMRI: Gordon (333 parcels) [Gordon et al. \(2014\)](#), Power (130 parcels) [Power et al. \(2011\)](#), Yeo (96 parcels) [Yeo et al. \(2011\)](#), Baldassano (171 parcels defined using the HCP FC data) [Baldassano et al. \(2015\)](#), Shen (200 parcels) [Shen et al. \(2013\)](#) and atlas based on multiple modalities: Fan (210 parcels) [Fan et al. \(2016\)](#) (cortical surface morphology, FC and SC), Human Connectome Project (HCP) multi-modality parcellation (Glasser, 360 parcels) [Glasser et al. \(2016\)](#) which includes fMRI (both resting-state and task-based from the HCP), cortical thickness and myelin mapping. See [Arslan et al. \(2018\)](#) for further details of the considered parcellations.

#### 2.6. Homogeneous vs size-matched geometric parcellations

We contrasted the atlas performances to two geodesic parcellation schemes based on agglomerative hierarchical clustering. The geodesic parcellations were generated using average linkage agglomerative hierarchical clustering with the geodesic distance as measure of proximity. The geodesic distance was given as the shortest path along the cortical surface between the vertices of the average surface ( $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$ ).

Thereby parcels are defined solely based on their 3-dimensional Cartesian coordinates forming spatially-homogeneous random parcels uninformed by anatomy, SC, or FC. This approach is similar to the geometric parcellation considered in [Thirion et al. \(2014\)](#) but with the additional constraint of accounting for the cortical surface and using agglomerative hierarchical clustering. This is in contrast to the Euclidean distances and k-means clustering employed in [Thirion et al. \(2014\)](#) and also included herein (labelled "Geometric"), which uses the group level parcellation available from [Arslan et al. \(2018\)](#). To quantify merits of using geodesics accounting for surface structure we included the corresponding agglomerative hierarchical clustering using Euclidean distances. The agglomerative hierarchical clustering was performed using MATLAB (R2016a, The MathWorks Inc., Natick, MA, 2000).

The agglomerative hierarchical clustering procedure produces parcels of roughly the same size. Thus, in order to assess the influence of inhomogeneous size distributions used by the considered atlases we further considered a geodesic clustering in which the cluster size distributions were similar to the atlases. This was accomplished sorting the sizes of each parcel within each atlas in descending order and identifying the first point in the dendrogram in which the largest cluster was less than the current largest size parcel of the atlas. These vertices were then assigned as corresponding to this atlas parcel and the nodes removed from the dendrogram. The procedure was then recursed considering the next ordered largest parcel in the atlas and the remaining nodes in the dendrogram. This resulted in a parcellation in which each extracted parcel was smaller than the size matched atlas parcel. The unassigned vertices were subsequently assigned neighboring parcels by alternatingly considering parcels absolutely lacking most vertices to reach their target size or parcels lacking most vertices as measured relative to their target parcel size.

The considered atlases and their corresponding homogeneous and size distributed geodesic representations are given in [Fig. 2](#). Notably, and by design, the geodesic geometric parcellation does not comply with anatomy and has a size distribution substantially differing from the atlases (red dots in right most panel), whereas the size-matched geodesic clustering procedure is in close agreement with the size distribution of the atlases considered (blue dots in the right most panel).

## 2.7. Data driven parcellations

Apart from the atlases above we further included data driven group-wise parcellations based on clustering fMRI time-series given by k-means parcellations (K-Means-2) described in [Thirion et al. \(2014\)](#), the N-Cuts procedure of [Craddock et al. \(2012\)](#) (N-Cuts-2), Arslan group level parcellation [Arslan et al. \(2015\)](#), Blumensath group level parcellation [Blumensath et al. \(2013\)](#), Bellec group level parcellation [Bellec et al. \(2006\)](#), and the Ward agglomerative Euclidean clustering (Ward-2) all provided in [Arslan et al. \(2018\)](#) and based on HCP rs-fMRI data considering Dataset 1 of the Unrelated 100 in the HCP database (for further details see [Arslan et al. \(2018\)](#)). These data driven parcellations are defined at multiple resolutions and [Fig. 3](#) specifies their resolutions available as well as parcel size distributions given as boxplots. Notably, the hierarchical geometric procedures using Euclidean and geodesic distances admit the evaluation of any resolution level by considering different levels of the cuts in the generated dendrograms. We have thus further included their representations at all levels of resolutions given by the above parcellations. Inspecting [Fig. 3](#) we observe that Bellec has relatively homogeneous parcel sizes which we attribute to a threshold on allowed size applied in this procedure to avoid overly large parcels [Bellec et al. \(2006\)](#).

## 3. Results

To establish the predictive assessment procedure we first consider a synthetic study in which the connectivity structure of the training and

test connectomes are generated in compliance with the atlases. We subsequently used the predictive assessment using SC and FC data from the Human Connectome project.

### 3.1. Synthetic study

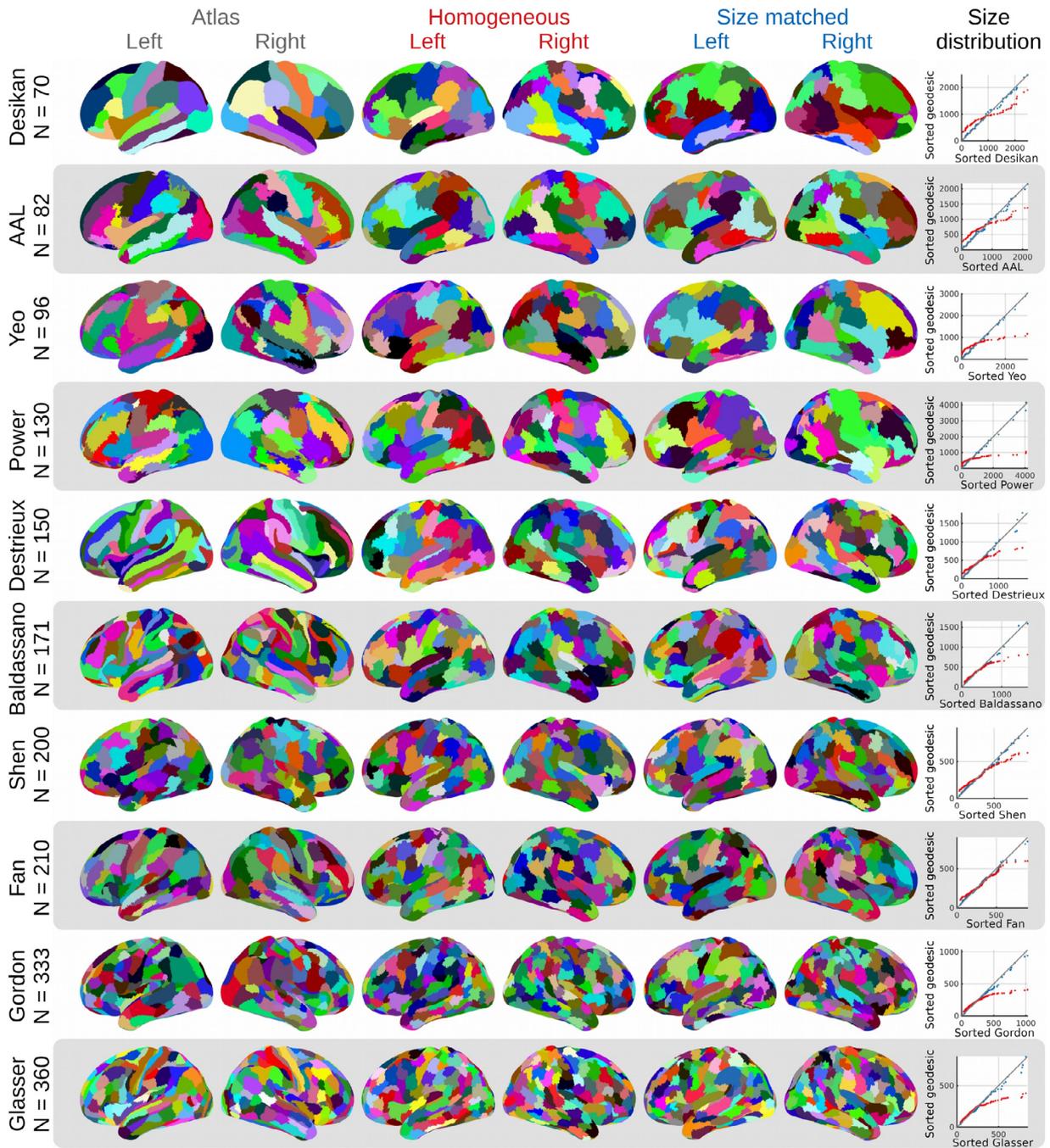
For each of the ten considered group level atlases of [Arslan et al. \(2018\)](#) we synthetically generated 5 connectomes according to an organization complying respectively with each of the atlases. The data was generated such that the relative probability of generating a link between nodes in the two parcels  $l$  and  $m$  were given by the probability  $\eta_{l,m} \sim \text{Beta}(0.5, 0.5)$ , i.e. the bimodal arcsine distribution. Notably, the bimodality provides high support for low and high density regions (i.e.  $\eta_{l,m}$  close to zero or close to one), thus, well reflecting that some regions are highly interconnected whereas others have very few connections between them. The five graphs for the given parcellation  $p$  were then generated according to the induced inter-parcel connectivity structure  $\eta^{(p)}$ , and at an approximate density of 1% such that the links and non-links between parcels relatively complied with the specified link-densities between parcels. The predictive assessment was then employed iteratively, firstly using the first graph as training and second as test, and then secondly using the second graph as training and third for test, and so on. Finally the fifth graph was used as training and the first graph as test. This provided in total five predictive assessments. When considering the synthetic datasets generated for parcel  $p$ , we evaluated the performance of all parcellations including both the parcellation used to generate the data and the parcellations differing from the one used to generate the data. We further included the geometric parcellation using Euclidean and Geodesic distances employing the same number of parcels as used in the parcellation used to generate the synthetic dataset. For each of the 10 atlases we thus generated five connectomes complying with the respective atlas structure and evaluated the performance of 12 parcellations including the atlas used to generate the data.

[Fig. 4](#) details the normalized mutual information (NMI) between the considered parcellations as well as the results of the predictive assessment framework on the synthetically generated connectomes averaged across the five predictions. NMI is computed as

$$\text{NMI}(\mathbf{z}^{(a)}, \mathbf{z}^{(b)}) = 2\text{MI}(\mathbf{z}^{(a)}, \mathbf{z}^{(b)}) / (\text{H}(\mathbf{z}^{(a)}) + \text{H}(\mathbf{z}^{(b)})), \quad (9)$$

where  $\text{MI}(\mathbf{z}^{(a)}, \mathbf{z}^{(b)}) = \sum_l \sum_m P(l, m) \log\left(\frac{P(l, m)}{P(l)P(m)}\right)$  denotes the mutual information between parcellation  $\mathbf{z}^{(a)}$  and  $\mathbf{z}^{(b)}$  where  $P(l, m) = \frac{1}{n} \sum_i \delta_{z_i^{(a)}, l} \delta_{z_i^{(b)}, m}$ ,  $\text{H}(\mathbf{z}^{(x)}) = \text{MI}(\mathbf{z}^{(x)}, \mathbf{z}^{(x)}) = -\sum_l P(l) \log(P(l))$  is the entropy of parcellation  $x$  such that  $P(l) = \frac{1}{n} \sum_i \delta_{z_i^{(x)}, l}$ . The normalization ensures NMI is upper bounded by 1.

Whereas the NMI explicitly quantifies correspondence of the parcellations knowing the correct underlying parcellation, the predictive assessment uses the observed connectomics data to assess the validity of parcellations. As such, the predictive assessment in contrast to NMI is asymmetric as the likelihood changes depending on the parcellation used to generate the data. The atlas used to define the organization of links and non-links in each connectome (row index) clearly outperforms the atlases that were not used to generate the connectomes (given by off diagonal elements) demonstrating that the predictive framework correctly points to the atlas that best accounts for the connectivity structure in the data. Interestingly, the homogeneous geodesic and Euclidean parcellations identified using agglomerative clustering on the vertex coordinates and having the same number of parcels as the underlying atlases generally perform on par with the incorrect atlases. Whereas the results in [Fig. 4](#) is not surprising, they establish that the proposed predictive evaluation framework in the ideal setting when connectomes comply with a known ground truth parcellation structure is able to pinpoint the most suitable parcellation (i.e., the parcellation that was used to generate the data). Importantly, not knowing what parcellation generated the data the predictive evaluation framework can evaluate how plausi-



**Fig. 2.** Employed parcellation (left most columns) and the geodesic geometric parcellations extracted using agglomerative hierarchical clustering (mid columns) and the procedure to generate parcels of similar sizes to the considered atlases. Right most panel the sorted parcel sizes of the considered atlas (x-axis) plotted against the sorted sizes of the geodesic agglomerative (red dots) and geodesic size matched (blue dots) procedures (y-axis).

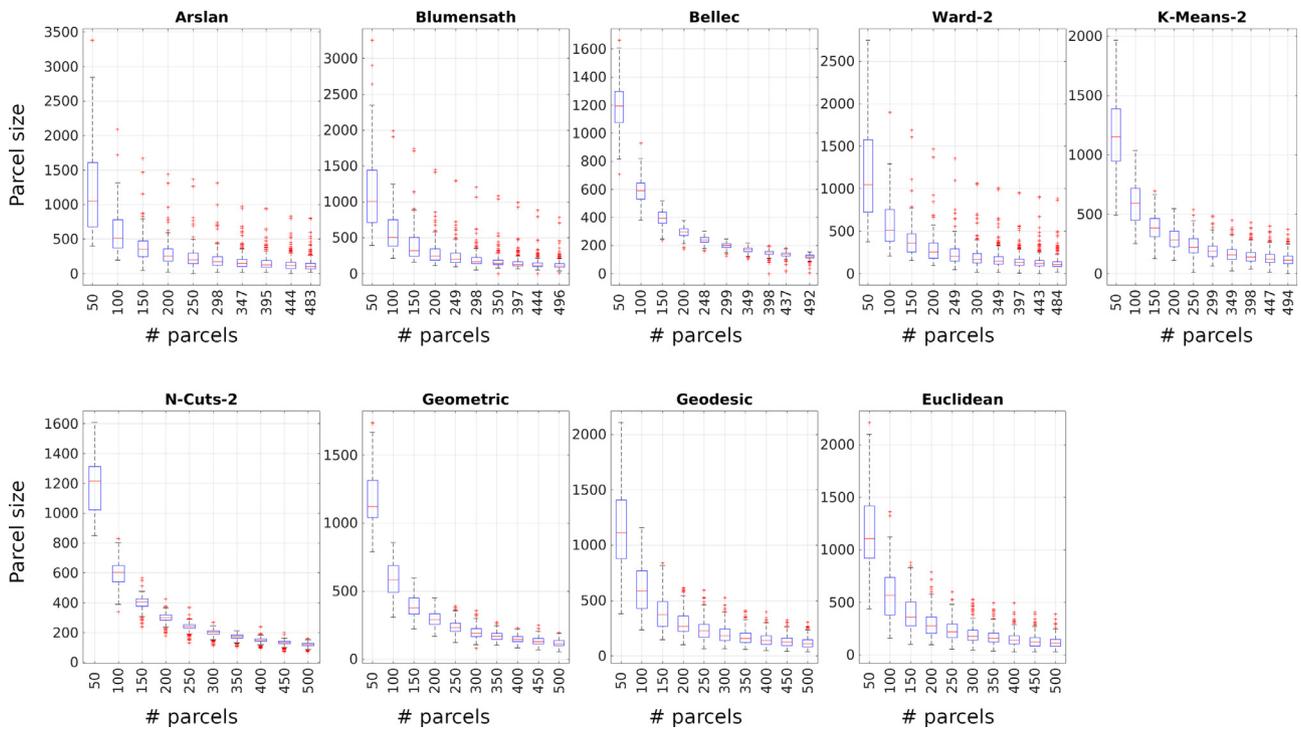
ble different parcellations are in terms of accounting for the observed structure of links and non-links between the vertices of the connectome.

To further investigate the predictive assessment and the three different performance measures AUC,  $L$  and  $LL$  we systematically analyze the influence of subdividing or merging parcels, misaligning vertices and changing the prior in Fig. 5.

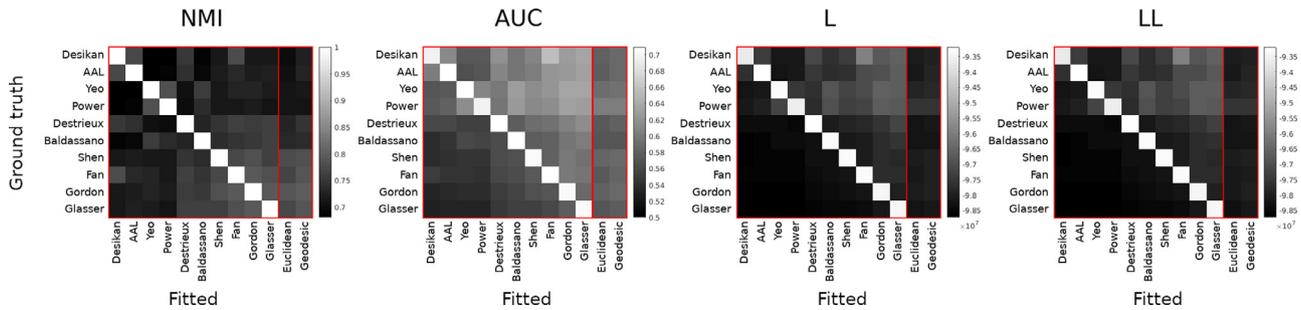
In the top panel of Fig. 5 we generated synthetic datasets according to the geodesic hierarchical clustering using 100, 250 and 500 parcels and investigated performance as function of resolution according to the geodesic hierarchical clustering. For all the considered predictive performance measures we observe that they point to the correct resolution whereas merging parcels appear to have a more detrimental effect than

subdividing parcels. We further observe that  $LL$  appears slightly more sensitive than AUC and  $L$ .

In the middle panel of Fig. 5 we investigated the impact of having 1% to 20% of vertices misaligned between the training and test graph when generating the graphs using 200 parcels based on the geodesic hierarchical clustering solution. We observe that the predictive performance declines as the misalignment increases. The AUC appears robust to misalignment pointing to a 200 parcel solution as most adequate for all levels of misalignment considered. However, we observe that both  $L$  and  $LL$  completely fails in identifying the correct resolution facing misalignment of 20% of the vertices and that  $L$  as opposed to  $LL$  keeps increasing when using more than 200 parcels when facing 10% mis-



**Fig. 3.** Considered data driven parcellations including the agglomerative hierarchical clustering procedures using average linkage considering Geodesic and Euclidean distances. x-axis denotes number of parcels and y-axis the size of parcels. The size distribution at each level of number of parcels employed are given as box plots.



**Fig. 4.** Synthetic analysis of the predictive assessment procedure. Left most panel, assessment of parcellations using normalized mutual information (NMI) based on knowing the true underlying parcellation. Right three panels: The predictive assessment in which the connectivity structure of the observed connectomes  $A^{(train)}$  and  $A^{(test)}$  generated according to the parcellation structure of each atlas is used to assess the validity of parcellations, respectively considering the AUC, log predictive likelihood ( $L$ ) and predictive log-loss ( $LL$ ).

aligned vertices. This points to  $LL$  performing better than  $L$  in the realistic scenario where all vertices are not perfectly aligned.

Finally, in the bottom panel of Fig. 5 we investigate the impact of the choice of prior used for the predictive assessment utilising the Jeffrey’s prior  $Beta(0.5, 0.5)$ , a uniform  $Beta(1, 1)$  prior, as well as a weak prior  $Beta(0.01, 0.99)$  correctly imposing a network density of 1% and a stronger  $Beta(0.1, 9.9)$  prior also imposing a network density of 1%. We observe that the predictive assessment for the AUC and  $L$  appears uninfluenced by these choices of priors. Whereas the  $LL$  appears uninfluenced by the choice of the priors  $Beta(1, 1)$ ,  $Beta(0.5, 0.5)$ ,  $Beta(0.1, 9.9)$  the weak 1% density  $Beta(0.01, 0.99)$  prior deviates substantially from the use of the other priors as the number of parcels and thus impact of the prior increases. Regardless the choice of prior all predictive assessments considered correctly points to the 200 parcel solution.

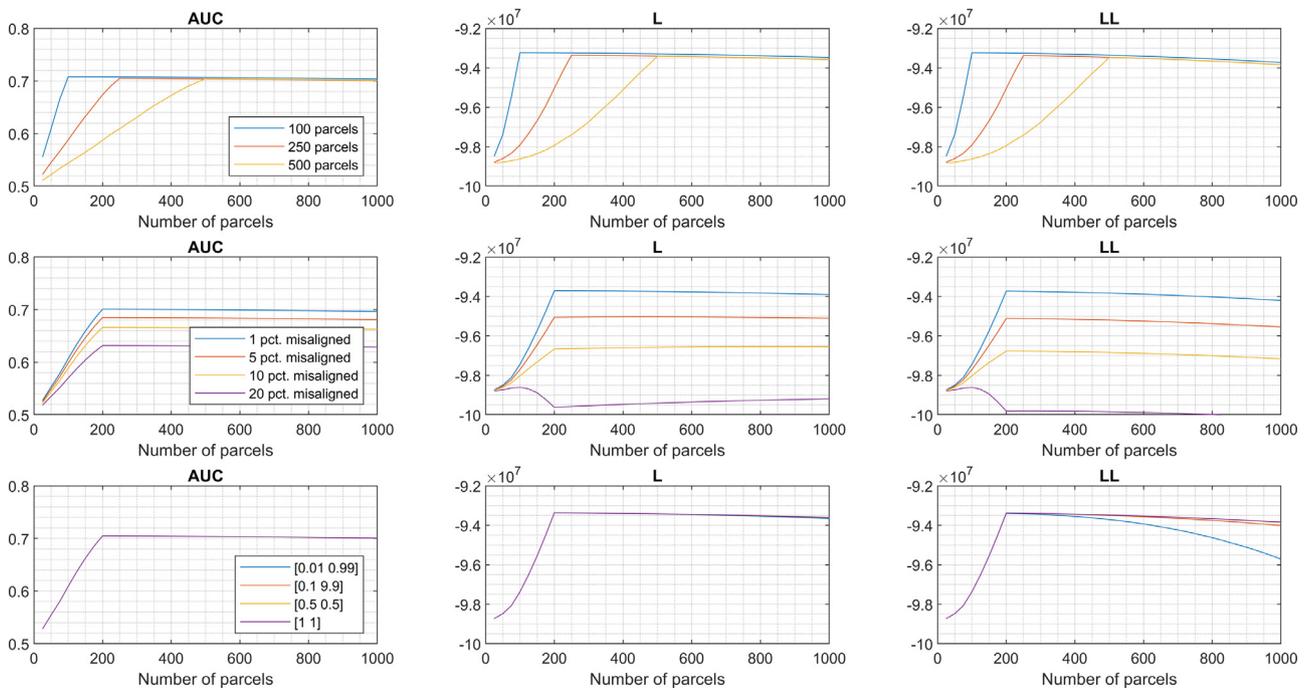
### 3.2. HCP structural and functional connectivity networks

Similar to the synthetic data analyses for the real SC and FC networks each subject is used once to predict one other subject’s connectome such that subject 1 predicts subject 2, subject 2 predicts subject 3 and so on

such that subject 250 predicts subject 1. As such, all subjects are used both as training and testing once. For the FC analyses we apart from inter subject predictions also considered intra subject analyses. Here we used the rescan of a subject as test data such that each subject’s first scan was used as FC training data to predict the subject’s own rescan FC data. Fig. 6 shows the predictive assessment of parcellations applied to the per-subject FC and SC graphs.

#### 3.2.1. Functional connectivity

We found as to be expected (Fig. 6, upper panels) that the predictive performance was consistently and substantially better when predicting the connectivity structure within subjects (intra FC assessment) than between subjects (inter FC assessment). Furthermore, for the intra-subject predictions, Baldassano was the best performing atlas, followed by Glasser. Notably, the training procedure used to derive the Baldassano atlas included resting-state fMRI data from the HCP, and the same data was also used to inform the generation of the Glasser atlas. For the intra-subject predictions we also note that with the exception of N-Cuts-2 all the rs-fMRI informed data driven parcellations, (i.e, Arslan, Blumensath, Bellec, Ward-2 and K-means-2), perform substantially better



**Fig. 5.** Investigation of impact of resolution (top row), misalignment (middle row) and choice of prior parameters (bottom row) for the predictive assessment. Standard deviation of the mean is not included as shaded region as they are negligible (less than 0.05% of the observed values).

than the geometric parcellations. However, the geometric parcellations, (i.e. Geometric and Geodesic) generally perform well, and indeed are on par or better than most of the atlases, particularly for the intra-subject predictive assessment. We further observe that the agglomeration based on geodesic distances performs slightly better than the corresponding parcellation based on Euclidean distances for the intra-subject analyses. However, accounting for the uneven sizes of parcels used by the atlases (blue dots) we observe that the atlases in general (for exceptions see Destrieux and Shen) outperform the size matched inhomogeneous geometric parcellations (black dots). This indicates that the atlas-defined trade-off in terms of fine-grained and coarse resolutions across the cortex better complies with the functional connectivity structure. For the inter-subject predictions, the overall predictive performance is substantially lower than the intra-subject predictive performance. Notably, the predictive performance reaches its maximum at a relatively low number of parcels (<200) for all three predictive performance metrics. One of the best performing atlas is here also Baldassano and the size-matched geometric parcellations generally performing worse than the atlases. We do not for the inter-subject assessment observe consistent differences in performance using Euclidean or Geodesic agglomerative clustering.

### 3.2.2. Structural connectivity

For the SC graphs (Fig. 6, lower panels) we observe that Fan is the best performing atlas, followed by Gordon. Notably, Fan is the only one of the employed atlases whose derivation incorporated SC information. Interestingly, we again observe that the geometric parcellations perform on par or better than most atlases. However, when accounting for the uneven size distributions of the employed atlases a substantial drop in performance is in general observed resulting in worse performance of the geodesic geometric procedure than the atlases. Again this points to a favourable trade-off in the employed atlases of how fine and coarse structures are positioned along the cortex. Notably, in accounting for SC, Bellec provides the most favourable parcellation across all three predictive measures and across most resolutions. Again we also observe that Geodesic outperforms Euclidean.

To assess the influence of network thresholding on the predictive assessment we in Fig. 7 include an analysis based on a subsample of 25

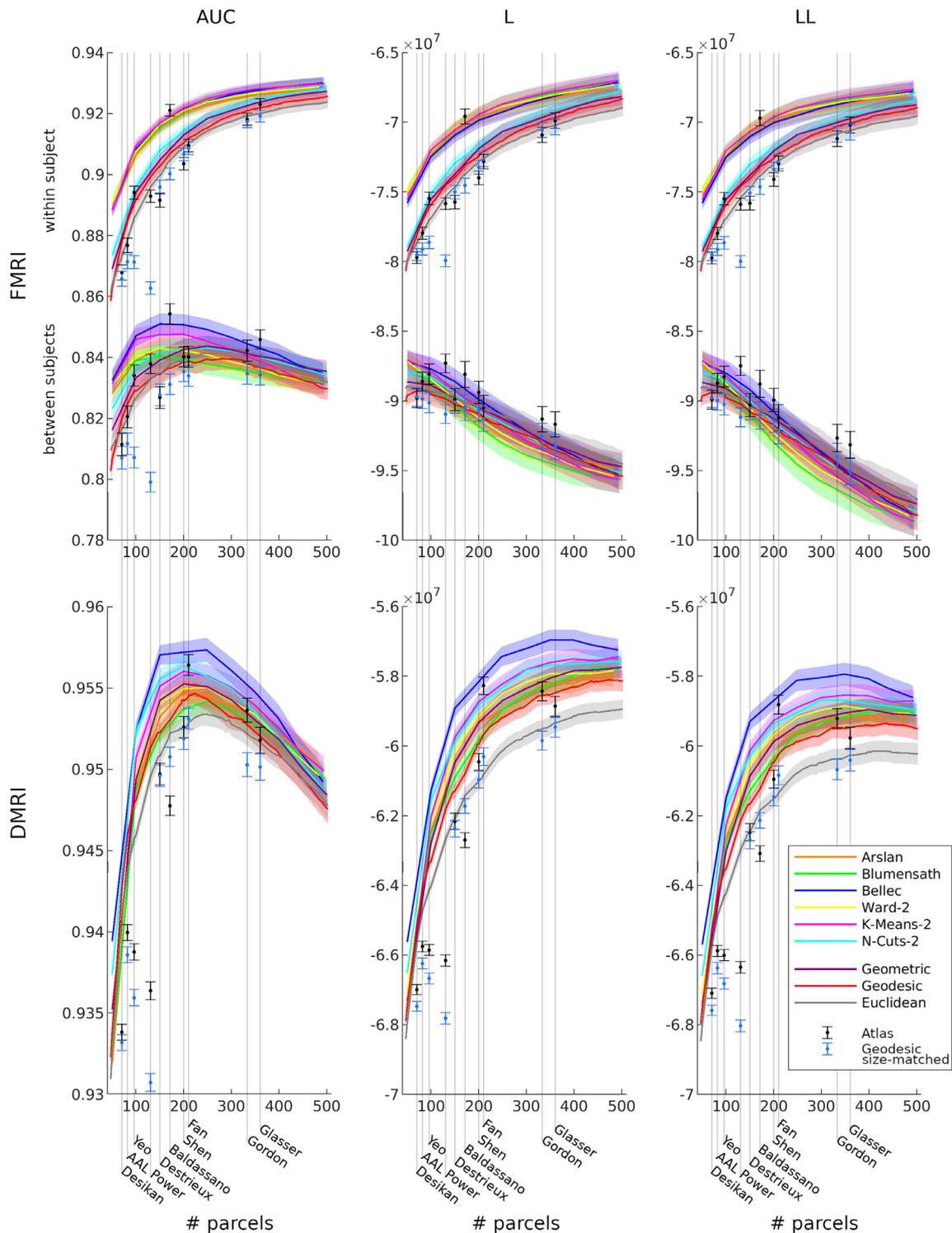
structural connectivity graphs. The SC graphs are respectively thresholded at 0.5%, 1%, 2% and 4%. We observe that the comparative rankings of the different parcellations are reasonably preserved across threshold levels. However, we also observe a substantial impact on the supported network resolutions, in particular, when considering predictive assessment using  $L$  and  $LL$  such that increasing the threshold reduces the supported parcellation resolution.

## 4. Discussion

We have presented a validation framework that permits quantitative assessment of any given parcellation scheme in the absence of a gold standard reference (ground truth parcellation). The framework uses statistical prediction to validate a parcellation by its ability to characterize the structure of brain connectivity data. Using this framework we evaluated several existing parcellations (several based solely upon FC data but only one (Fan) based upon SC data), in their ability to characterize the organization of FC and SC data from the Human Connectome Project.

The synthetic study demonstrated the validity of the predictive framework as it correctly identified the correct parcellation used to generate each connectome. Notably, the predictive assessment could correctly identify the true atlas from the generated connectomes providing similar correspondences between parcellations to correspondences assessed by normalized mutual information (NMI). Importantly, NMI requires knowing the true atlas structure as opposed to the predictive assessment procedure relying only on having observed the connectomes induced by the atlas structure. From the synthetic study it was also observed that in the absence of the correct parcellation the homogeneous geodesic parcellation performed on par with many of the atlases not used to generate the data. Thus, when incorrectly characterizing the underlying connectivity organization a homogeneous representation provided a good alternative representation.

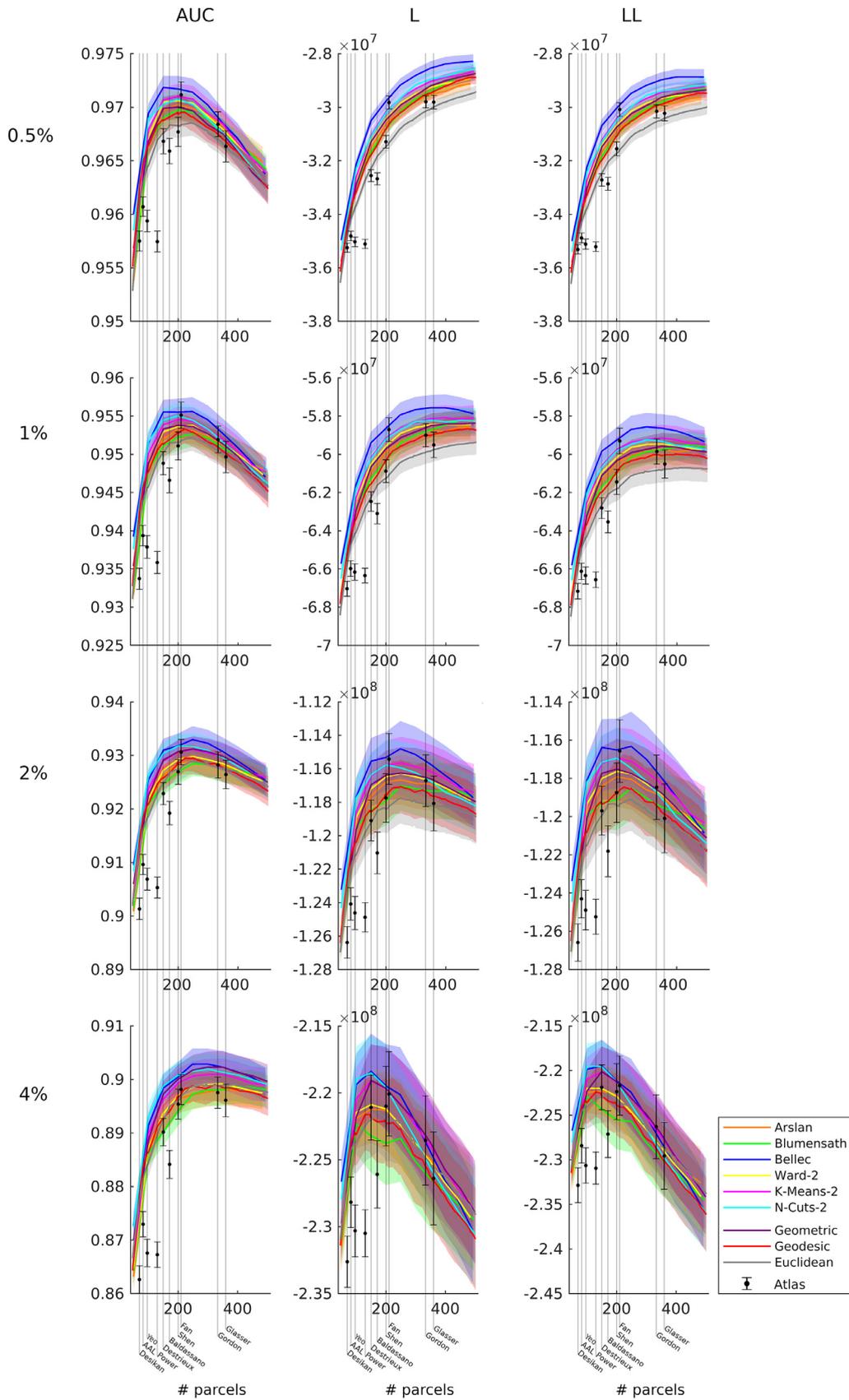
On the real FC and SC data we observed that no atlas representation is in general superior both for intra and inter subject functional connectivity and structural connectivity at the same time. It is reassuring that the atlases based on HCP fMRI data, i.e. Baldassano and Glasser



**Fig. 6.** Predictive assessment using FC (top panel) and SC (bottom panel) respectively considering the AUC, log predictive likelihood ( $L$ ) and predictive log-loss ( $LL$ ). Shaded regions and error bars denote standard deviation of the mean. (For the results, methods separated by more than 1.25 times their standard deviations can be rejected as performing the same at a 5% level (using a one-sided test assuming performances are normally distributed and the ratio between largest and smallest standard deviation of the two methods compared are below 2). Black dots are the actual atlases performance and blue dots the performance of the corresponding geodesic size-matched parcellation. The vertical gray lines indicate the location of each atlas in terms of resolution (number of parcels) employed.

were both assessed to perform well in accounting for FC. For SC however, other atlases, namely Fan and Gordon, were assessed to perform better. Notably, the Fan atlas, which also employed structural connectivity during its derivation, was found to perform best. We attribute this to different biases in the FC and SC data (see also below) and previously re-

ported discrepancies in functional and structural connectivity estimates Greicius et al. (2009); Honey et al. (2009). See also Røge et al. (2017a); Suárez et al. (2020) and references therein. Notably, in the recent study of Messé (2020) structure-function relationship have been found to de-



**Fig. 7.** Predictive assessment considering a subset of 25 SC graphs thresholded respectively at 0.5%, 1%, 2% and 4% considering the AUC, log predictive likelihood ( $L$ ) and predictive log-loss ( $LL$ ). Shaded regions and error bars denote standard deviation of the mean across the 25 predictions. Black dots are the actual atlases performance. The vertical gray lines indicate the location of each atlas in terms of resolution (number of parcels) employed.

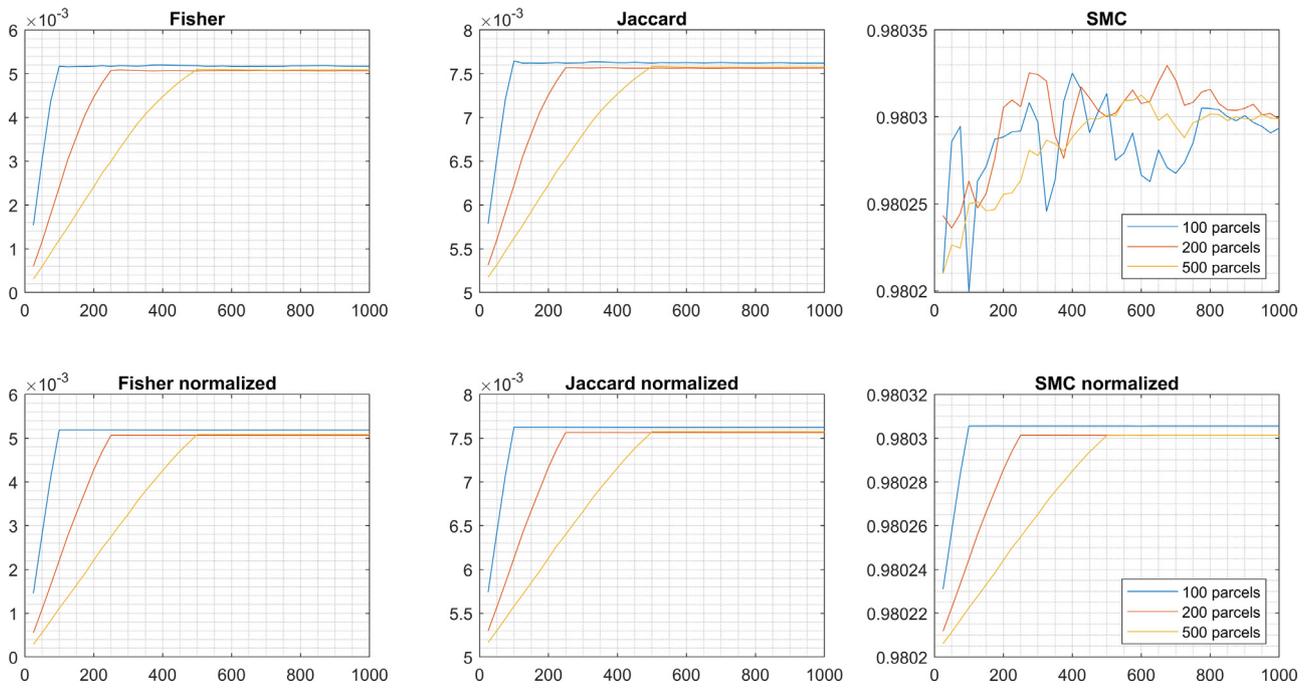


Fig. 8. Assessment using homogeneity considering the synthetic study in which networks are generated according to a parcellation having 100, 250 and 500 parcels.

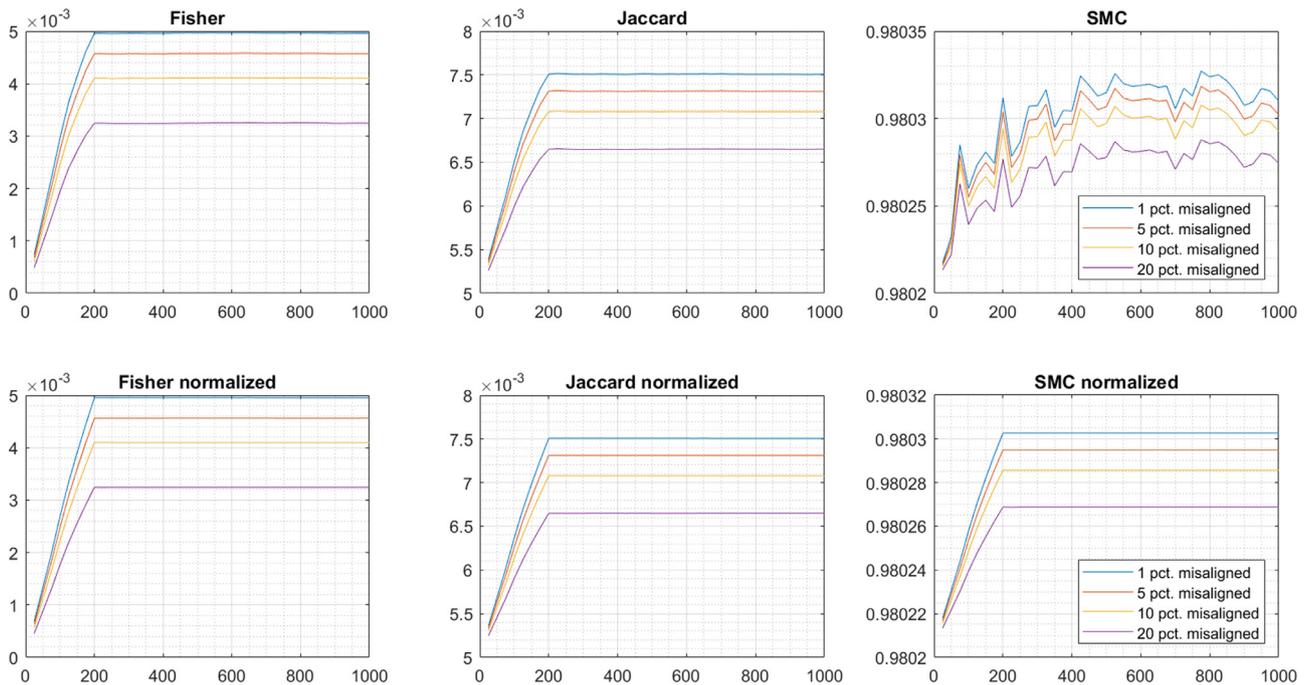


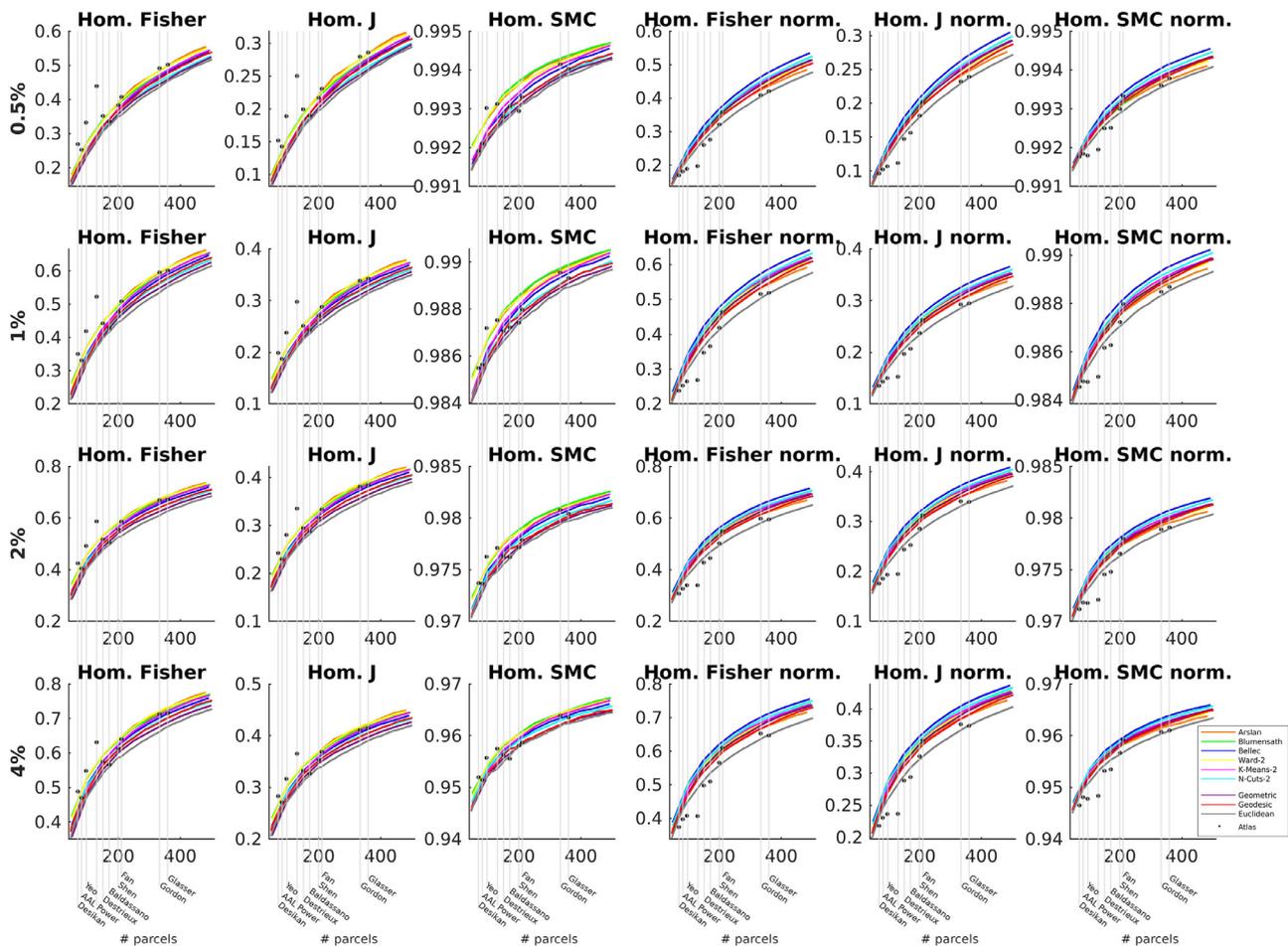
Fig. 9. Assessment using homogeneity considering the synthetic study in which networks generated are misaligned by 1%, 5%, 10% and 20% with respect to the ground truth parcellation having 200 parcels.

pend on the parcellation employed, and our results further indicate that SC and FC support different atlas representations.

The inter-subject structural connectivity predictive assessment had a substantially higher AUC and predictive likelihood when compared to the predictive performance of functional connectivity, scoring even higher than the intra-subject predictions in FC. We attribute this to the structural connectivity estimates being more reliable than the functional connectivity estimates. We emphasize, however, that performance of atlases in terms of the predictive scores should be interpreted relatively and not absolutely, as the offset (vertical shift) observed between the

FC and SC performance is a result of the consistency differences of the derived FC and SC graphs. Such an offset explains, for example, how a consistent, yet incorrectly estimated, connectome may be easier to predict than a noisy but unbiased connectome.

The geometric parcellations that are uninformed by anatomy, SC and FC performed surprisingly well in general performing better than most atlases at their corresponding level of resolution. We attribute this to their relative homogeneous size distributions as observed in Fig. 2 right panel as well as Fig. 3. We investigated the impact of size distribution by retaining the parcel size distribution, yet with an incorrect repre-

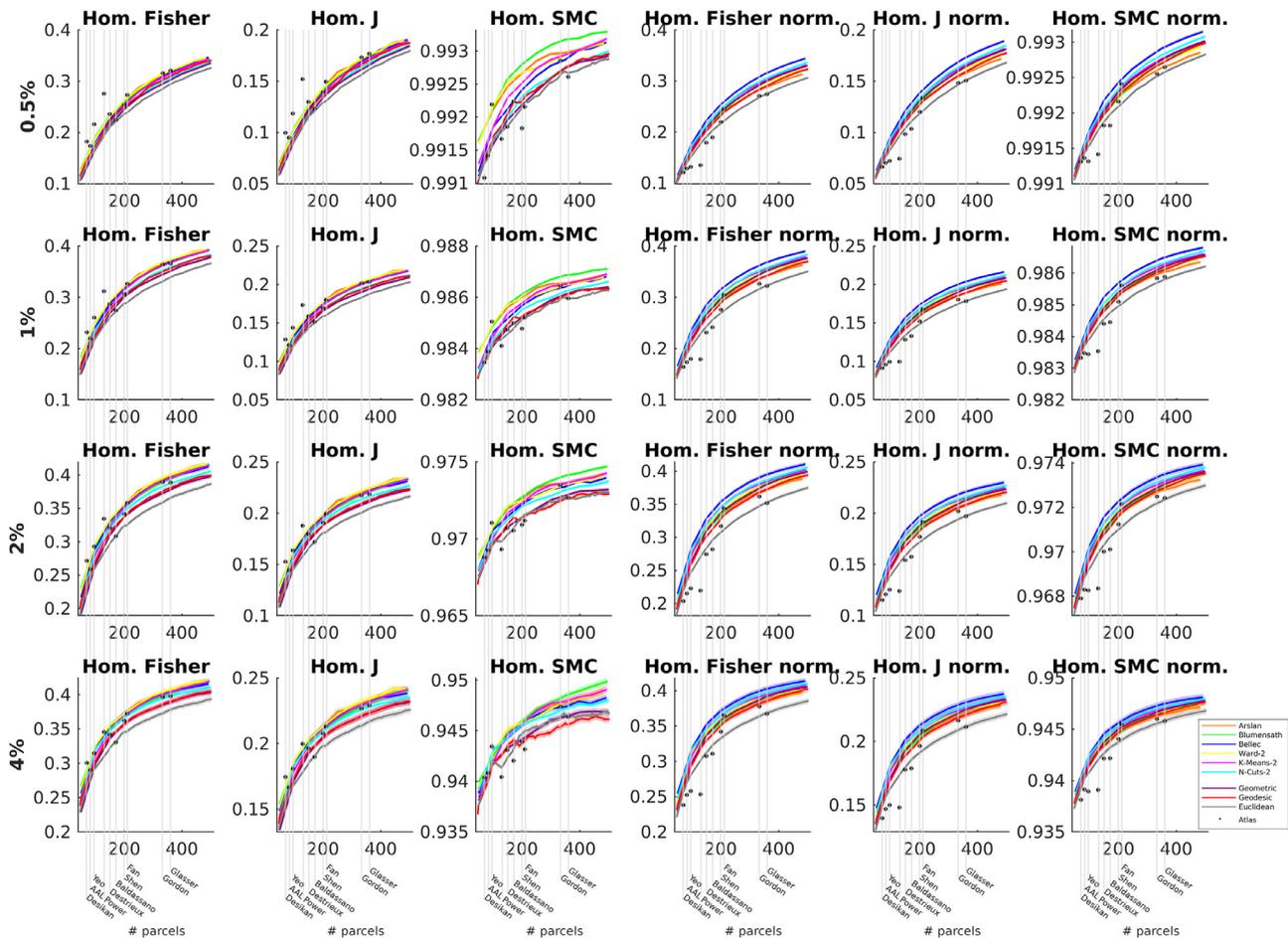


**Fig. 10.** Assessment using homogeneity considering a subset of 25 SC graphs thresholded respectively at 0.5%, 1%, 2% and 4%. Homogeneity is based on similarity defined by Fisher transformed correlation (Fisher), Jaccard ( $J$ ) and the Simple Matching Coefficient ( $SMC$ ) respectively. Left three plots are based on averaging homogeneity across parcels whereas the rightmost three plots are based on normalizing the overall homogeneity according to parcel size. Shaded regions and error bars denote standard deviation of the mean across the 25 predictions. Black dots are the actual atlases performance. The vertical gray lines indicate the location of each atlas in terms of resolution (number of parcels) employed. (As two of the SC graphs contained nodes with no links for a threshold level of 0.5% these networks were discarded and this analysis is based on the remaining 23 SC graphs.)

sensation of the cortical regions in terms of their underlying network resolution. We here observed that an incorrect representation of the underlying resolution levels deteriorated performance when compared to the homogeneous geodesic representation. For the data-driven parcellations we further observed the most favorable performance using Bellec for the SC and inter-subject FC analyses. Inspecting the size distribution of Bellec in Fig. 3 we observe Bellec as having very even sized parcels as indicated by the in general relative narrow box plots. We thus attribute the superior performance of Bellec to not using too coarse representations by avoiding parcels of relatively very large sizes. Taken together this provide supports, in general, for the importance as to how the resolution (relatively large or small parcel sizes) is defined in terms of the underlying network organization and when not adequately complying with the structure of the connectomes it is favorable to use even sized parcels. The utility of such homogeneous representations has previously been reported in the context of structural Hagmann et al. (2008) and functional Thirion et al. (2014) MRI data. Finally, we also observed that, for the SC and intra-subject FC predictions, accounting for the cortical surface structure using Geodesic distances generally performed better than the corresponding use of Euclidean distances.

The recent popularity of data-driven in-vivo parcellation approaches, as exemplified by Behrens et al. (2003a); Bellec et al. (2015, 2006); Blumensath et al. (2013); Clos et al. (2013); Craddock et al. (2013); Eickhoff et al. (2015); Fan et al. (2016); Glasser et al. (2016);

Shen et al. (2013); Thirion et al. (2014); Yeo et al. (2011), highlights the growing demand for alternatives to conventional atlases, e.g. Desikan et al. (2006); Destrieux et al. (2010); Lancaster et al. (2000); Nieuwenhuys (2013); Nieuwenhuys et al. (2015); Tzourio-Mazoyer et al. (2002). However, it is necessary to recall that such data-driven methods are primarily uni-modal (though with some exceptions, e.g. Fan et al. (2016); Glasser et al. (2016); Parisot et al. (2017)), and as such their resultant parcellations do not conform to the traditional definition of cortical areas, where within-area homogeneity and inter-area uniqueness are required to be congruent across the three commonly-employed modalities of structure, function and connectivity Amunts and Zilles (2015); Felleman and Van Essen (1991); Van Essen et al. (1992). While there is much evidence that boundaries between purported cortical areas are often well-defined within, and congruent across, modalities, this is primarily limited to the well-investigated sub-cortical, sensory- and motor-regions, e.g. Behrens et al. (2003a); Bzdok et al. (2013); Eickhoff et al. (2015); Kelly et al. (2012). In contrast, the more abstract processing units in the higher associative areas within the prefrontal, parietal, and to some extent the temporal cortices have been challenging to delineate across modalities Cerliani et al. (2017); Eickhoff et al. (2017, 2015). Any uni-modal parcellation can therefore only "yield maps of the brain that can be seen as spatial hypotheses on functional or structural segregation - a hypothesis that may and should be tested by integrative,



**Fig. 11.** Assessment using homogeneity in which within parcel similarity is quantified between two graphs  $A^{(train)}$  and  $A^{(test)}$  considering a subset of 25 SC graphs thresholded respectively at 0.5%, 1%, 2% and 4%. Homogeneity is based on similarity between two graphs defined by Fisher transformed correlation (Fisher), Jaccard ( $J$ ) and the Simple Matching Coefficient ( $SMC$ ) respectively. Left three plots are based on averaging homogeneity across parcels whereas the rightmost three plots are based on normalizing the overall homogeneity according to parcel size. Shaded regions and error bars denote standard deviation of the mean across the 25 predictions. Black dots are the actual atlases performance. The vertical gray lines indicate the location of each atlas in terms of resolution (number of parcels) employed. (As two of the SC graphs contained nodes with no links for a threshold level of 0.5% results including these networks were discarded and the analysis is based on 21 pairs of SC graphs.)

multi-modal investigations” Eickhoff et al. (2015). The methods and results presented herein can be interpreted as a quantitative assessment of such a parcellation-via-structure hypothesis, revealing how well potentially modality-independent parcellations predict (and therefore possess congruency with) the organization of structural and functional connectomes as presently quantified from MRI.

#### 4.1. Number of parcels supported by SC and FC data

The predictive framework also provides an indication of the level of resolution that the SC and FC data supports. In particular, we found that for inter-subject predictive assessment the SC supported a finer resolution than FC. Currently, there is no consensus or sufficient evidence to support how many parcels should be used. Even the authors of the atlas with the greatest number of parcels (360) used in this study (Glasser et al. Glasser et al. (2016)) state that their parcellation may still underestimate the true number of parcels at the macroscale, as their subdividing of areas such as the primary visual cortex are coarser than reported previously Glasser et al. (2016). Other studies also report higher cluster numbers for data-driven approaches, e.g. Røge et al. (2017b); Thirion et al. (2014) for fMRI and our intra-subject predictive assessment based on fMRI also points to the utility of a large number of clusters. Although the proposed predictive framework can also be used

to quantify the level of resolution best accounting for the connectivity structure, care must be taken when interpreting the estimated numbers of clusters supported by the FC and SC data. As such, we find that it is dependent upon inter vs. intra subject variability in FC and is likely to be influenced by the quality and processing of the dMRI and fMRI data in general. The optimum may also be influenced by biases in the SC and FC data, as discussed below in Sections 4.2- 4.3, and be a natural consequence of the presence of fuzzy boundaries due to microstructure gradients between purported cortical regions, more common in the higher association areas Carmichael and Price (1994); Eickhoff et al. (2017). These effects can potentially lead to an incorrect estimate of the number of functional and structural units. Importantly, we also observe that the optimal numbers of parcels depend on metric used for the predictive evaluation. As such, we generally observe that the AUC metric favors less parcels than  $L$  and  $LL$  that behave very similar. We attribute this to the  $L$  and  $LL$  metrics potentially being driven by outliers. I.e., the presence of links where the estimated associated probability of observing links is very low. In contrast, the AUC metric is concerned only about the relative ranking of the observed links and non-links in terms of the connectivity density of the training data and thereby can be considered more robust to outliers. On synthetic data we systematically investigated the impact of parcel resolution, misalignment and choice of prior when using the three metrics AUC,  $L$ , and  $LL$ . We observed that all

metrics disfavored more strongly the use of few as opposed to too many parcels and that the  $LL$  generally was more sensitive than the AUC and  $L$ , whereas the AUC was more robust in particular facing severe misalignments of vertices. Whereas the three predictive measures overall behaved similarly they have different inherent trade-offs in terms of their sensitivity to resolution, noise (i.e., misalignment), and choice of prior specification and we thus argue to consider them all when assessing the predictive performance.

#### 4.2. Biases in the surface registration between subjects

A possible limitation to the results reported herein is the accuracy of the initial vertex-to-vertex registration framework, as provided by the HCP pipeline Glasser et al. (2013). As this is driven by surface topology Fischl (2012), there exists the possibility that the subsequent vertex alignment is biased towards anatomical landmarks (and therefore provides atlases based upon surface morphology with an inherent prediction boost). As anatomy may not be an optimal predictor of SC Amunts et al. (1999); Amunts and Zilles (2015); Devlin and Poldrack (2007), this means that the assumed vertex-to-vertex correspondence may not fully reflect the nature of the SC data. Hence, such a bias would exhibit itself as noise in the vertex labeling, which would in turn propagate to the adjacency matrix (graph). As a consequence, it would be more difficult to produce large homogeneous parcels of vertices which all possess similar patterns of FC and SC. This would make larger clusters less likely, and so our predictive framework could therefore support an over-parcellation. Advanced vertex registration procedures, such as that employed in Glasser et al. (2016); Robinson et al. (2014), may improve matters as the imposed predictive bias will be balanced between multiple modalities. On synthetic data we observed the predictive assessment to be robust to vertex misalignments. However, we also observed that substantial misalignment (i.e., 20% of vertices misaligned between training and test graphs) had a detrimental effect on the performance of  $L$  and  $LL$ .

#### 4.3. Modality specific biases

SC is established from dMRI data by integrating the derived local estimates of fibre bundle orientations obtained with standard tractography methods Behrens et al. (2003a). However, just as for all other methods that estimate connectivity, tractography has its own challenges and limitations, e.g. gyral crown bias Donahue et al. (2016); Reveley et al. (2015); Van Essen et al. (2013a), which could affect the precise location of parcel borders, path length dependencies Liptrot et al. (2014) and other factors such as the impact of scanning parameters Ambrosen et al. (2020) which together are known to impose unknown levels of Type I and Type II errors on the estimated connections Jones (2010); Jones et al. (2013); Le Bihan et al. (2006); Maier-Hein et al. (2017); Morris et al. (2008). These confounds, biases and shortcomings of tractography are as yet not fully quantifiable due to the lack of a gold-standard reference Dyrby et al. (2018); Knösche et al. (2015), and indeed are not detectable as they will be present in both training and test datasets. Previous studies have reported a compliance between SC (e.g. Wang et al., 2015) with other modalities, though the extent of such inter-modality congruence is likely to vary across the cortex Cerliani et al. (2017).

For FC we observe substantial predictive performance differences between inter-subject as opposed to intra-subject prediction which we attribute to substantial inter-subject variability. fMRI is inherently noisy and influenced by motion artifacts, cardiac and respiratory noise, whereas the blood oxygen level dependent (BOLD) signal is only an indirect measure of neural activity Murphy et al. (2013). These confounds makes the fMRI signal very noisy and can introduce biases in both the derived training and test functional connectomes. In particular, subject motion is attributed to have a detrimental effect on reliability when compared to scanner noise and between session coregistration

Gorgolewski et al. (2013). Although preprocessing pipelines attempt to address these confounds they are not perfect and their choices impact results Churchill et al. (2012).

Unfortunately, no matter which connectivity modality is employed within our prediction framework it will not be able to compensate for modality-specific biases present in both training and test populations. However, the framework introduced herein can easily be extended to include multi-modality data beyond SC and FC, such as tracer studies, or histological reconstruction of axonal trajectories Amunts and Zilles (2015); Dyrby et al. (2018). As demonstrated by Glasser et al. (2016), the incorporation of many independent data-sources can mollify the effects of their individual biases and here parcellations that are performing well across modalities by the proposed predictive assessment procedure may be less influenced by modality specific biases.

#### 4.4. Thresholding of FC and SC networks

As with any graph model of connectivity, false positives and false negatives will occur as the incorrect presence or absence of links. Herein, as is common practice Drakesmith et al. (2015); Hagemann et al. (2008, 2007), we attempt to remove many of the false positive connections by thresholding the FC and SC graphs at a threshold of 1% density. However, this uniformly-applied strategy also increases the false negative rate. Unfortunately, whilst the false positive rate can in principle be reduced to zero by increasing the threshold, the minimum false negative rate, achieved at null thresholding, will be non-zero and can only be improved by better data acquisition and processing strategies. As such, it must be noted that the choice of threshold level determines the balance between specificity and sensitivity, and no optimal threshold exists Dyrby et al. (2007); Knösche et al. (2015); Qi et al. (2015); Zalesky et al. (2016). In Ambrosen et al. (2020) we investigated the influence of different thresholds and found that aggregated statistics considering structural connectivity showed robustness to threshold. Notably, SC and FC are likely to have different noise content, thus one might expect that to make them equivalent (in terms of SNR) then they would need different thresholds. As the computations required for the considered threshold level is massive we presently considered the impact of threshold for a subset of 25 SC graphs. Increasing the threshold impacts the SNR of the derived connectomes and we observed in particular for the  $L$  and  $LL$  predictive measures that increasing the threshold including more noisy links resulted in reduced support for many parcels. However, the comparative ranking of the parcellations were reasonably preserved which we attribute to the aggregated statistics of inter-parcel densities of the parcellations to be robust. In particular, we found the AUC measure to be more robust to threshold levels which we attribute to the measure solely relying on the ranking of links and non-links according to inter parcel densities. Contrary to the  $L$  and  $LL$  measures that are influenced by network density the AUC relying on the rate of false and true positives are invariant to class-imbalance (i.e., network density). We however observe that the AUC declines as we increase the threshold. We attribute this to the graph becoming less predictable and thus more noisy when increasing the threshold.

#### 4.5. Alternative representations of graphs

While it is possible to run the analysis on volumetric data we used the vertex based HCP representation for the assessment of all atlases. Whereas this required volumetric parcels be translated to vertex representations as done in Arslan et al. (2018) this translation can introduce noise and obscure parcels thereby make the performance of volumetric parcels, i.e., Power, Yeo, Fan, Shen, and AAL less favorable. However, we cannot compare performance assessed using volumetric data to the performance using vertex based data as this will change the network used for the predictions. We therefore used the vertex based representation which can give atlases that are vertex based an advantage.

An obvious extension to this work would also be the incorporation of weighted graphs. However, it is often not clear how relevant weightings could be generated. For example for SC graphs weights could be streamline counts Hagmann et al. (2010a, 2008, 2007); Honey et al. (2009), path-based apparent diffusion coefficient Hagmann et al. (2010b) or fractional anisotropy van den Heuvel and Sporns (2011), none of which has direct correspondence with connection strength Jones et al. (2013). For functional connectivity the correlation values, or Fisher transformations thereof, could be modelled directly as proposed in Baldassano et al. (2015); Herlau et al. (2012b) circumventing the need of thresholding the graphs. Nevertheless, should a universally-agreed way of generating weights become wide-spread, the framework presented herein is fully capable of utilising weighted graphs when using the predictive likelihood performance metrics. Finally, whereas a few atlases have included both cortical and sub-cortical regions (e.g. Power et al. (2011); Yeo et al. (2011)), the aim of this study focused on a direct comparison with existing parcellations containing cortical-only parcels. However, the prediction framework demonstrated herein is equally applicable to the inclusion of sub-cortical regions.

The present framework considered connectivity density estimates between parcels as the only statistics used for the predictive assessment of connectomes. This model structure assumes that connectivities across parcels are independent given their connectivity strength, complying with the assumptions imposed in classic stochastic block-modeling Kemp et al. (2006); Nowicki and Snijders (2001); Schmidt and Mørup (2013); Snijders and Nowicki (1997). We note that the proposed predictive assessment procedure naturally extends to more advanced block-model structures such as the degree-corrected stochastic block-model Herlau et al. (2014); Karrer and Newman (2011) and models accounting for hierarchical structure Blundell and Teh (2013); Clauset et al. (2008); Herlau et al. (2012a); Roy et al. (2007).

As opposed to homogeneity measures relying on in-sample estimates the predictive assessment procedure considers how network structure is preserved across training and test graphs. Importantly, subdividing parcels and in general including more parcels improves upon performance as smaller parcels tend to be more homogeneous. Interestingly, as noted in Arslan et al. (2018) homogeneity typically has the shortcoming that inhomogeneous parcellation composed of several small parcels and few large parcels will perform better than parcellations homogeneous in size distribution. We observe the opposite effect utilizing the proposed predictive assessment procedure and it thus seem to not exhibit such biases observed utilizing traditional homogeneity measures. To reduce such bias in homogeneity measures it has been proposed to weight homogeneity by parcel size Arslan et al. (2018). Notably, the influence of parcel is directly accounted for in the predictive assessment as larger parcels contribute more than small parcels in the predictive assessment as more links and non-links are predicted for the larger parcels. Assessing parcellations using homogeneity considering the synthetic study in Fig. 5 and analysis of a subset of 25 SC graphs in Fig. 7 can be found in the supplementary material. We here include also a homogeneity analysis based on measuring within parcel similarity between two SC graphs.

## 5. Conclusion

We proposed a predictive assessment procedure that uses connectomics data to quantify the validity of brain parcellations. The efficacy of the procedure was verified on synthetic connectomics data complying in connectivity structure with prominent brain parcellations. Using FC and SC data from the HCP, the assessment procedure provided a comparative ranking of parcellations.

We observed that no parcellations were best across the three connectomics data sets considered (i.e., intra FC, inter FC and inter SC). Notably, we observed that the Baldassano and Fan atlases, which were explicitly trained on FC and on SC respectively, were also the best at accounting for FC and SC. For the data driven parcellations we found that for the intra FC analyses the data-driven parcellations were more favor-

able than the geometric parcellations. For inter FC an SC we observed that Bellec had the most favorable performance which we attribute to the homogeneous size distribution employed when here defining the parcels. Notably, these data-driven parcellations were based on HCP rs-fMRI data thus potentially biased to have favorable performance for the FC data. However, we also observe good performance of these data driven approaches when considering SC that does not rely on the rs-fMRI data. We further observed for the intra FC and inter SC analyses that geometric parcellations using geodesics accounting for the surface structure in general performed better than using Euclidean distances for the agglomerative clustering procedure. Surprisingly, we found that homogeneous geometric parcellations generally performed on par or better than most of the considered atlases despite not relying on anatomy nor HCP rs-fMRI data as the data-driven procedures. However, when accounting for the inhomogeneity of parcel sizes the atlases generally had a favorable representation of coarser and finer delineations of structures. This points to the importance of not using overly large parcels that may incorrectly aggregate the connectome and the utility of size constraints such as imposed in Bellec. Whereas the procedure pointed to the numbers of parcels to use to best account for SC and FC data this depend on the assessment measure (i.e., AUC or  $L$  and  $LL$ ) and is intrinsically related to the signal-to-noise ratio of the connectomics data. As such we observed the support of substantially more parcels when considering intra subject FC assessment as opposed to inter-subject FC assessment.

In summary, we find that the most suitable parcellation depends on data modality and analyses scope, i.e. intra vs. inter-subject analysis. We further observe that resolution is important and more so than the specific geometry of the parcellations in general. We presently validated all the parcellations using FC and SC data, but the proposed validation framework is generic and therefore applicable to any other brain connectivity mapping approach. As data-quality, and the number of data-sources and data-derived approaches to connectomics, and thereby also parcellation schemes, will only increase in the future, we foresee that the prediction framework presented herein will prove to be an important tool in assessing their quality.

## 6. Compliance with Ethical Standards

**Funding:** This study was funded by Lundbeckfond, grant number R105-9813. The Tesla K40 GPU card used for data preprocessing was donated by the NVIDIA Foundation. **Conflict of Interest:** The authors declare that they have no conflict of interest.

**Ethical approval:** This article does not contain any studies with human participants or animals performed by any of the authors. As such, ethical approval is not required.

**Informed consent:** For this type of study (i.e. retrospective) formal consent is not required.

**Data and code availability:** Data sharing is not applicable to this article as no new data were created or analyzed in this study. Code for the predictive assessment and analysis of simulated data is available at <https://brainconnectivity.compute.dtu.dk/>.

## Credit authorship contribution statement

**Kristoffer J. Albers:** Conceptualization, Methodology, Software, Formal analysis, Investigation, Data curation, Writing - original draft, Writing - review & editing. **Karen S. Ambrosen:** Conceptualization, Methodology, Software, Formal analysis, Investigation, Data curation, Writing - original draft, Writing - review & editing. **Matthew G. Liptrot:** Conceptualization, Methodology, Software, Formal analysis, Investigation, Data curation, Writing - original draft, Writing - review & editing. **Tim B. Dyrby:** Conceptualization, Methodology, Writing - review & editing. **Mikkel N. Schmidt:** Conceptualization, Methodology, Writing - review & editing. **Morten Mørup:** Conceptualization, Methodology, Software, Formal analysis, Investigation, Writing - original draft, Writing - review & editing, Project administration, Funding acquisition.

## Acknowledgments

This project was supported by the Lundbeck Foundation, grant no. R105-9813. The Tesla K40 GPU card used for the BedpostX calculations was donated by the NVIDIA Corporation. Data were provided by the Human Connectome Project, WU-Minn Consortium (Principal Investigators: David Van Essen and Kamil Ugurbil; 1U54MH091657) funded by the 16 NIH Institutes and Centers that support the NIH Blueprint for Neuroscience Research; and by the McDonnell Center for Systems Neuroscience at Washington University. We would like to thank the authors of [Arslan et al., 2018] for making the parcellations considered easy accessible 1225 and providing visualizing code, as utilized in figure 2, available at [biomedica.doc.ic.ac.uk/brain-parcellation-survey/](http://biomedica.doc.ic.ac.uk/brain-parcellation-survey/).

## Supplementary material

Supplementary material associated with this article can be found, in the online version, at [10.1016/j.neuroimage.2021.118170](https://doi.org/10.1016/j.neuroimage.2021.118170)

## References

- Ambrosen, K.S., Albers, K.J., Dyrby, T.B., Schmidt, M.N., Mørup, M., 2014. Nonparametric bayesian clustering of structural whole brain connectivity in full image resolution. In: 2014 International Workshop on Pattern Recognition in Neuroimaging. IEEE, pp. 1–4.
- Ambrosen, K.S., Eskildsen, S.F., Hinne, M., Krug, K., Lundell, H., Schmidt, M.N., van Gerven, M.A., Mørup, M., Dyrby, T.B., 2020. Validation of structural brain connectivity networks: The impact of scanning parameters. *NeuroImage* 204, 116207.
- Ambrosen, K.S., Herlau, T., Dyrby, T., Schmidt, M.N., Mørup, M., 2013. Comparing structural brain connectivity by the infinite relational model. In: 2013 International Workshop on Pattern Recognition in Neuroimaging (PRNI). IEEE, pp. 50–53.
- Amunts, K., Schleicher, A., Bürgel, U., Mohlberg, H., Uylings, H.B.M., Zilles, K., 1999. Broca's region revisited: Cytoarchitecture and intersubject variability. *J. Comp. Neurol.* 412 (2), 319–341.
- Amunts, K., Zilles, K., 2015. Architectonic mapping of the human brain beyond brodmann. *Neuron* 88 (6), 1086–1107.
- Andersen, K.W., Madsen, K.H., Siebner, H.R., Schmidt, M.N., Mørup, M., Hansen, L.K., 2014. Non-parametric bayesian graph models reveal community structure in resting state fmri. *NeuroImage* 100, 301–315.
- Arslan, S., Ktena, S.I., Makropoulos, A., Robinson, E.C., Rueckert, D., Parisot, S., 2018. Human brain mapping: a systematic comparison of parcellation methods for the human cerebral cortex. *NeuroImage* 170, 5–30.
- Arslan, S., Parisot, S., Rueckert, D., 2015. Joint spectral decomposition for the parcellation of the human cerebral cortex using resting-state fmri. In: International Conference on Information Processing in Medical Imaging. Springer, pp. 85–97.
- Baldassano, C., Beck, D.M., Fei-Fei, L., 2015. Parcellating connectivity in spatial maps. *PeerJ* 3, e784.
- Behrens, T., Berg, H.J., Jbabdi, S., Rushworth, M., Woolrich, M., 2007. Probabilistic diffusion tractography with multiple fibre orientations: what can we gain? *NeuroImage* 34 (1), 144–155.
- Behrens, T., Johansen-Berg, H., Woolrich, M., Smith, S., Wheeler-Kingshott, C., Boulby, P., Barker, G., Sillery, E., Sheehan, K., Ciccarelli, O., et al., 2003. Non-invasive mapping of connections between human thalamus and cortex using diffusion imaging. *Nat. Neurosci.* 6 (7), 750–757.
- Behrens, T., Woolrich, M., Jenkinson, M., Johansen-Berg, H., Nunes, R., Clare, S., Matthews, P., Brady, J., Smith, S., 2003. Characterization and propagation of uncertainty in diffusion-weighted mr imaging. *Magn. Reson. Med.* 50 (5), 1077–1088.
- Bellec, P., Benhajali, Y., Carbonell, F., Dansereau, C., Albouy, G., Pelland, M., Craddock, C., Collignon, O., Doyon, J., Stip, E., Orban, P., 2015. Impact of the resolution of brain parcels on connectome-wide association studies in fmri. *NeuroImage* 123, 212–228.
- Bellec, P., Perlberg, V., Jbabdi, S., Pélégrini-Issac, M., Anton, J.-L., Doyon, J., Benali, H., 2006. Identification of large-scale networks in the brain using fmri. *NeuroImage* 29 (4), 1231–1243.
- Blumensath, T., Jbabdi, S., Glasser, M.F., Van Essen, D.C., Ugurbil, K., Behrens, T.E., Smith, S.M., 2013. Spatially constrained hierarchical parcellation of the brain with resting-state fmri. *NeuroImage* 76, 313–324.
- Blundell, C., Teh, Y.W., 2013. Bayesian hierarchical community discovery. In: Advances in Neural Information Processing Systems, pp. 1601–1609.
- Bohland, J.W., Bokil, H., Allen, C.B., Mitra, P.P., 2009. The brain atlas concordance problem: quantitative comparison of anatomical parcellations. *PLoS One* 4 (9), e7200.
- Braitenberg, V., Schüz, A., 1991. *Anatomy of the Cortex: Statistics and Geometry*. Springer-Verlag Publishing.
- Bullmore, E., Sporns, O., 2009. Complex brain networks: graph theoretical analysis of structural and functional systems. *Nat. Rev. Neurosci.* 10 (3), 186–198.
- Bzdok, D., Laird, A.R., Zilles, K., Fox, P.T., Eickhoff, S.B., 2013. An investigation of the structural, connective, and functional subspecialization in the human amygdala. *Hum. Brain Mapp.* 34 (12), 3247–3266.
- Carmichael, S.T., Price, J.L., 1994. Architectonic subdivision of the orbital and medial prefrontal cortex in the macaque monkey. *J. Comp. Neurol.* 346 (3), 366–402.
- Cerliani, L., D'Arceuil, H., Thiebaut de Schotten, M., 2017. Connectivity-based parcellation of the macaque frontal cortex, and its relation with the cytoarchitectonic distribution described in current atlases. *Brain Struct. Funct.* 222 (3), 1331–1349.
- Churchill, N.W., Oder, A., Abdi, H., Tam, F., Lee, W., Thomas, C., Ween, J.E., Graham, S.J., Strother, S.C., 2012. Optimizing preprocessing and analysis pipelines for single-subject fmri i. standard temporal motion and physiological noise correction methods. *Hum. Brain Mapp.* 33 (3), 609–627.
- Clauset, A., Moore, C., Newman, M.E., 2008. Hierarchical structure and the prediction of missing links in networks. *Nature* 453 (7191), 98–101.
- Clos, M., Amunts, K., Laird, A.R., Fox, P.T., Eickhoff, S.B., 2013. Tackling the multifunctional nature of Broca's region meta-analytically: Co-activation-based parcellation of area 44. *NeuroImage* VL - 83 IS -, 174–188.
- Craddock, R.C., James, G.A., Holtzheimer, P.E., Hu, X.P., Mayberg, H.S., 2012. A whole brain fmri atlas generated via spatially constrained spectral clustering. *Hum. Brain Mapp.* 33 (8), 1914–1928.
- Craddock, R.C., Jbabdi, S., Yan, C.-G., Vogelstein, J.T., Castellanos, F.X., Di Martino, A., Kelly, C., Heberlein, K., Colcombe, S., Milham, M.P., 2013. Imaging human connectomes at the macroscale. *Nat. Methods* 10 (6), 524–539.
- Desikan, R.S., Ségonne, F., Fischl, B., Quinn, B.T., Dickerson, B.C., Blacker, D., Buckner, R.L., Dale, A.M., Maguire, R.P., Hyman, B.T., et al., 2006. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *NeuroImage* 31 (3), 968–980.
- Destrieux, C., Fischl, B., Dale, A., Halgren, E., 2010. Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. *NeuroImage* 53 (1), 1–15.
- Devlin, J.T., Poldrack, R.A., 2007. In praise of tedious anatomy. *NeuroImage* 37 (4), 1033–1041.
- Donahue, C.J., Sotiropoulos, S.N., Jbabdi, S., Hernandez-Fernandez, M., Behrens, T.E., Dyrby, T.B., Coalson, T., Kennedy, H., Knoblauch, K., Van Essen, D.C., et al., 2016. Using diffusion tractography to predict cortical connection strength and distance: a quantitative comparison with tracers in the monkey. *J. Neurosci.* 36 (25), 6758–6770.
- Dornas, J.V., Braun, J., 2018. Finer parcellation reveals detailed correlational structure of resting-state fmri signals. *J. Neurosci. Methods* 294, 15–33.
- Drakesmith, M., Caeyenberghs, K., Dutt, A., Lewis, G., David, A., Jones, D., 2015. Overcoming the effects of false positives and threshold bias in graph theoretical analyses of neuroimaging data. *NeuroImage* 118, 313–333.
- Dyrby, T.B., Innocenti, G.M., Bech, M., Lundell, H., 2018. Validation strategies for the interpretation of microstructure imaging using diffusion mri. *NeuroImage* 182, 62–79.
- Dyrby, T.B., Søgaard, L.V., Parker, G.J., Alexander, D.C., Lind, N.M., Baaré, W.F., Hay-Schmidt, A., Eriksen, N., Pakkenberg, B., Paulson, O.B., et al., 2007. Validation of in vitro probabilistic tractography. *NeuroImage* 37 (4), 1267–1277.
- Eickhoff, S.B., Constable, R.T., Yeo, B.T.T., 2017. Topographic organization of the cerebral cortex and brain cartography. *NeuroImage*.
- Eickhoff, S.B., Thirion, B., Varoquaux, G., Bzdok, D., 2015. Connectivity-based parcellation: Critique and implications. *Hum. Brain Mapp.* 36 (12), 4771–4792.
- Fan, L., Li, H., Zhuo, J., Zhang, Y., Wang, J., Chen, L., Yang, Z., Chu, C., Xie, S., Laird, A.R., et al., 2016. The human brainnetome atlas: a new brain atlas based on connective architecture. *Cerebr. Cortex* bhv157.
- Feinberg, D.A., Moeller, S., Smith, S.M., Auerbach, E., Ramanna, S., Glasser, M.F., Miller, K.L., Ugurbil, K., Yacoub, E., 2010. Multiplexed echo planar imaging for sub-second whole brain fmri and fast diffusion imaging. *PLoS One* 5 (12), e15710.
- Felleman, D.J., Van Essen, D.C., 1991. Distributed hierarchical processing in the primate cerebral cortex. *Cerebral cortex* (New York, N.Y. : 1991) 1 (1), 1–47.
- Fischl, B., 2012. Freesurfer. *NeuroImage* 62 (2), 774–781.
- Fischl, B., van der Kouwe, A., Destrieux, C., Halgren, E., Ségonne, F., Salat, D.H., Busa, E., Seidman, L.J., Goldstein, J., Kennedy, D., et al., 2004. Automatically parcellating the human cerebral cortex. *Cerebr. cortex* 14 (1), 11–22.
- Fornito, A., Zalesky, A., Bullmore, E.T., 2010. Network scaling effects in graph analytical studies of human resting-state fmri data. *Front. Syst. Neurosci.* 4.
- Glasser, M.F., Coalson, T.S., Robinson, E.C., Hacker, C.D., Harwell, J., Yacoub, E., Ugurbil, K., Andersson, J., Beckmann, C.F., Jenkinson, M., Smith, S.M., Van Essen, D.C., 2016. A multi-modal parcellation of human cerebral cortex. *Nature* 536 (7615), 171–178.
- Glasser, M.F., Sotiropoulos, S.N., Wilson, J.A., Coalson, T.S., Fischl, B., Andersson, J.L., Xu, J., Jbabdi, S., Webster, M., Polimeni, J.R., et al., 2013. The minimal preprocessing pipelines for the human connectome project. *NeuroImage* 80, 105–124.
- Gong, G., He, Y., Concha, L., Lebel, C., Gross, D.W., Evans, A.C., Beaulieu, C., 2009. Mapping anatomical connectivity patterns of human cerebral cortex using in vivo diffusion tensor imaging tractography. *Cerebr. Cortex* 19 (3), 524–536.
- Gordon, E.M., Laumann, T.O., Adeyemo, B., Huckins, J.F., Kelley, W.M., Petersen, S.E., 2014. Generation and evaluation of a cortical area parcellation from resting-state correlations. *Cerebr. Cortex* bhv239.
- Gorgolewski, K.J., Storkey, A.J., Bastin, M.E., Whittle, I., Pernet, C., 2013. Single subject fmri test-retest reliability metrics and confounding factors. *NeuroImage* 69, 231–243. doi:10.1016/j.neuroimage.2012.10.085.
- Greicius, M.D., Supekar, K., Menon, V., Dougherty, R.F., 2009. Resting-state functional connectivity reflects structural connectivity in the default mode network. *Cerebr. Cortex* 19 (1), 72–78.
- Griffanti, L., Salimi-Khorshidi, G., Beckmann, C.F., Auerbach, E.J., Douaud, G., Sexton, C.E., Zsoldos, E., Ebmeier, K.P., Filippini, N., Mackay, C.E., et al., 2014. Ica-based artefact removal and accelerated fmri acquisition for improved resting state network imaging. *NeuroImage* 95, 232–247.
- Hagmann, P., 2005. From diffusion MRI to brain connectomics [phd thesis]. Lausanne: Ecole Polytechnique Fdrale de Lausanne (EPFL). 127 p.

- Hagmann, P., Cammoun, L., Gigandet, X., Gerhard, S., Grant, P.E., Wedeen, V., Meuli, R., Thiran, J.-P., Honey, C.J., Sporns, O., 2010. Mr connectomics: principles and challenges. *J. Neurosci. Methods* 194 (1), 34–45.
- Hagmann, P., Cammoun, L., Gigandet, X., Meuli, R., Honey, C.J., Wedeen, V.J., Sporns, O., 2008. Mapping the structural core of human cerebral cortex. *PLoS Biol.* 6 (7), e159.
- Hagmann, P., Kurant, M., Gigandet, X., Thiran, P., Wedeen, V.J., Meuli, R., Thiran, J.-P., 2007. Mapping human whole-brain structural networks with diffusion mri. *PLoS One* 2 (7), e597.
- Hagmann, P., SPORNS, O., Madan, N., Cammoun, L., Pienaar, R., Wedeen, V.J., Meuli, R., Thiran, J.P., Grant, P.E., 2010. White matter maturation reshapes structural connectivity in the late developing human brain. *Proc. Natl. Acad. Sci.* 107 (44), 19067–19072.
- Herlau, T., Mørup, M., Schmidt, M.N., Hansen, L.K., 2012. Detecting hierarchical structure in networks. In: 2012 3rd International Workshop on Cognitive Information Processing (CIP). IEEE, pp. 1–6.
- Herlau, T., Mørup, M., Schmidt, M.N., Hansen, L.K., 2012. Modelling dense relational data. IEEE International Workshop on Machine Learning for Signal Processing (MLSP), Santander, Spain.
- Herlau, T., Schmidt, M.N., Mørup, M., 2014. Infinite-degree-corrected stochastic block model. *Phys. Rev. E* 90 (3), 032819.
- Hernández, M., Guerrero, G.D., Cecilia, J.M., García, J.M., Inuggi, A., Jbabdi, S., Behrens, T.E., Sotiropoulos, S.N., 2013. Accelerating fibre orientation estimation from diffusion weighted magnetic resonance imaging using gpus. *PLoS One* 8 (4), e61892.
- van den Heuvel, M.P., Sporns, O., 2011. Rich-Club Organization of the Human Connectome. *J. Neurosci. : Off. J. Soc. Neurosci.* 31 (44), 15775–15786.
- Honey, C., Sporns, O., Cammoun, L., Gigandet, X., Thiran, J.-P., Meuli, R., Hagmann, P., 2009. Predicting human resting-state functional connectivity from structural connectivity. *Proc. Natl. Acad. Sci.* 106 (6), 2035–2040.
- Jenkinson, M., Beckmann, C.F., Behrens, T.E., Woolrich, M.W., Smith, S.M., 2012. *Fsl*. *Neuroimage* 62 (2), 782–790.
- Jones, D.K., 2010. Challenges and limitations of quantifying brain connectivity in vivo with diffusion MRI. *Imaging Med.* 2 (3), 341–355.
- Jones, D.K., Knösche, T.R., Turner, R., 2013. White matter integrity, fiber count, and other fallacies: the do's and don'ts of diffusion mri. *Neuroimage* 73, 239–254.
- Karrer, B., Newman, M.E., 2011. Stochastic blockmodels and community structure in networks. *Phys. Rev. E* 83 (1), 016107.
- Kelly, C., Toro, R., Di Martino, A., Cox, C.L., Bellec, P., Castellanos, F.X., Milham, M.P., 2012. A convergent functional architecture of the insula emerges across imaging modalities. *Neuroimage* 61 (4), 1129–1142.
- Kemp, C., Tenenbaum, J.B., Griffiths, T.L., Yamada, T., Ueda, N., 2006. Learning systems of concepts with an infinite relational model. In: AAAI, Vol. 3, p. 5.
- Knösche, T.R., Anwender, A., Liptrot, M., Dyrby, T.B., 2015. Validation of tractography: comparison with manganese tracing. *Hum. Brain Mapp.* 36 (10), 4116–4134.
- Lancaster, J.L., Woldorff, M.G., Parsons, L.M., Liotti, M., Freitas, C.S., Rainey, L., Kochunov, P.V., Nickerson, D., Milkiten, S.A., Fox, P.T., 2000. Automated Talairach Atlas labels for functional brain mapping. *Hum. Brain Mapp.* 10 (3), 120–131.
- Le Bihan, D., Poupon, C., Amadon, A., Lethimonnier, F., 2006. Artifacts and pitfalls in diffusion mri. *J. Magn. Reson. Imaging* 24 (3), 478–488.
- Liben-Nowell, D., Kleinberg, J., 2007. The link-prediction problem for social networks. *J. Am. Soc. Inf. Sci. Technol.* 58 (7), 1019–1031.
- Liptrot, M.G., Sidaros, K., Dyrby, T.B., 2014. Addressing the path-length-dependency confound in white matter tract segmentation. *PLoS One* 9 (5), e96247.
- Maier-Hein, K.H., Neher, P.F., Houde, J.-C., Côté, M.-A., Garyfallidis, E., Zhong, J., Chamberland, M., Yeh, F.-C., Lin, Y.-C., Ji, Q., et al., 2017. The challenge of mapping the human connectome based on diffusion tractography. *Nat. Commun.* 8 (1), 1–13.
- Messé, A., 2020. Parcellation influence on the connectivity-based structure function relationship in the human brain. *Hum. Brain Mapp.* 41 (5), 1167–1180. doi:10.1002/hbm.24866.
- Milchenko, M., Marcus, D., 2013. Obscuring surface anatomy in volumetric imaging data. *Neuroinformatics* 11 (1), 65–75.
- Miller, K., Jordan, M.I., Griffiths, T.L., 2009. Nonparametric latent feature models for link prediction. In: *Advances in Neural Information Processing Systems*, pp. 1276–1284.
- Moeller, S., Yacoub, E., Olman, C.A., Auerbach, E., Strupp, J., Harel, N., Ugurbil, K., 2010. Multiband multislice ge-epi at 7 tesla, with 16-fold acceleration using partial parallel imaging with application to high spatial and temporal whole-brain fmri. *Magn. Reson. Med.* 63 (5), 1144–1153.
- Morris, D.M., Embleton, K.V., Parker, G.J., 2008. Probabilistic fibre tracking: differentiation of connections from chance events. *Neuroimage* 42 (4), 1329–1339.
- Murphy, K., Birn, R.M., Bandettini, P.A., 2013. Resting-state fmri confounds and cleanup. *Neuroimage* 80, 349–359. doi:10.1016/j.neuroimage.2013.04.001. Mapping the Connectome
- Murre, J., Sturdy, D., 1995. The connectivity of the brain: multi-level quantitative analysis. *Biol. Cybern.* 73 (6), 529–545.
- Nieuwenhuys, R., 2013. The myeloarchitectonic studies on the human cerebral cortex of the vogt-vogt school, and their significance for the interpretation of functional neuroimaging data. In: *Microstructural Parcellation of the Human Cerebral Cortex*. Springer, pp. 55–125.
- Nieuwenhuys, R., Broere, C.A.J., Cerliani, L., 2015. A new myeloarchitectonic map of the human neocortex based on data from the Vogt–Vogt school. *Brain Struct. Funct.* 220 (5), 2551–2573.
- Nowicki, K., Snijders, T.A.B., 2001. Estimation and prediction for stochastic blockstructures. *J. Am. Stat. Assoc.* 96 (455), 1077–1087.
- Palla, K., Knowles, D., Ghahramani, Z., 2014. Relational learning and network modelling using infinite latent attribute models. *IEEE transactions on pattern analysis and machine intelligence* 37 (2), 462–474.
- Pariset, S., Arslan, S., Passerat-Palmbach, J., Wells, W.M., Rueckert, D., 2016. Group-wise parcellation of the cortex through multi-scale spectral clustering. *NeuroImage*.
- Pariset, S., Glocker, B., Ktena, S.I., Arslan, S., Schirmer, M.D., Rueckert, D., 2017. A flexible graphical model for multi-modal parcellation of the cortex. *NeuroImage* 162, 226–248.
- Passingham, R.E., Stephan, K.E., Kötter, R., 2002. The anatomical basis of functional localization in the cortex. *Nat. Rev. Neurosci.* 3 (8), 606–616.
- Power, J.D., Cohen, A.L., Nelson, S.M., Wig, G.S., Barnes, K.A., Church, J.A., Vogel, A.C., Laumann, T.O., Miezin, F.M., Schlaggar, B.L., et al., 2011. Functional network organization of the human brain. *Neuron* 72 (4), 665–678.
- Qi, S., Meesters, S., Nicolay, K., ter Haar Romeny, B.M., Ossenkop, P., 2015. The influence of construction methodology on structural brain network measures: a review. *J. Neurosci. Methods* 253, 170–182.
- Reveley, C., Seth, A.K., Pierpaoli, C., Silva, A.C., Yu, D., Saunders, R.C., Leopold, D.A., Frank, Q.Y., 2015. Superficial white matter fiber systems impede detection of long-range cortical connections in diffusion mr tractography. *Proc. Natl. Acad. Sci.* 112 (21), E2820–E2828.
- Robinson, E.C., Jbabdi, S., Glasser, M.F., Andersson, J., Burgess, G.C., Harms, M.P., Smith, S.M., Van Essen, D.C., Jenkinson, M., 2014. MSM: A new flexible framework for Multimodal Surface Matching. *NeuroImage* 100, 414–426.
- Røge, R., Ambrosen, K.S., Albers, K.J., Eriksen, C.T., Liptrot, M.G., Schmidt, M.N., Madsen, K.H., Mørup, M., 2017. Whole brain functional connectivity predicted by indirect structural connections. In: 2017 International Workshop on Pattern Recognition in Neuroimaging (PRNI). IEEE, pp. 1–4.
- Røge, R.E., Madsen, K.H., Schmidt, M.N., Mørup, M., 2017. Infinite von Mises–Fisher mixture modeling of whole brain fmri data. *Neural Comput.* 29 (10), 2712–2741.
- Roy, D.M., Kemp, C., Mansinghka, V.K., Tenenbaum, J.B., 2007. Learning annotated hierarchies from relational data. In: *Advances in Neural Information Processing Systems*, pp. 1185–1192.
- Salimi-Khorshidi, G., Douaud, G., Beckmann, C.F., Glasser, M.F., Griffanti, L., Smith, S.M., 2014. Automatic denoising of functional mri data: combining independent component analysis and hierarchical fusion of classifiers. *Neuroimage* 90, 449–468.
- Schmidt, M.N., Mørup, M., 2013. Nonparametric bayesian modeling of complex networks: An introduction. *Signal Process. Mag., IEEE* 30 (3), 110–128.
- Setsompop, K., Gagoski, B.A., Polimeni, J.R., Witzel, T., Wedeen, V.J., Wald, L.L., 2012. Blipped-controlled aliasing in parallel imaging for simultaneous multislice echo planar imaging with reduced g-factor penalty. *Magn. Reson. Med.* 67 (5), 1210–1224.
- Shen, X., Tokoglu, F., Papademetris, X., Constable, R.T., 2013. Groupwise whole-brain parcellation from resting-state fmri data for network node identification. *Neuroimage* 82, 403–415.
- Smith, S., 2013. Introduction to the neuroimage special issue mapping the connectome. *NeuroImage* 80 (Complete).
- Smith, S.M., Beckmann, C.F., Andersson, J., Auerbach, E.J., Bijsterbosch, J., Douaud, G., Duff, E., Feinberg, D.A., Griffanti, L., Harms, M.P., et al., 2013. Resting-state fmri in the human connectome project. *Neuroimage* 80, 144–168.
- Snijders, T.A., Nowicki, K., 1997. Estimation and prediction for stochastic blockmodels for graphs with latent block structure. *J. Classif.* 14 (1), 75–100.
- Sotiropoulos, S., Moeller, S., Jbabdi, S., Xu, J., Andersson, J., Auerbach, E., Yacoub, E., Feinberg, D., Setsompop, K., Wald, L., et al., 2013. Effects of image reconstruction on fiber orientation mapping from multichannel diffusion mri: reducing the noise floor using sense. *Magn. Reson. Med.* 70 (6), 1682–1689.
- Sporns, O., 2012. *Discovering the Human Connectome*. MIT press.
- Sporns, O., Tononi, G., Kötter, R., 2005. The human connectome: a structural description of the human brain. *PLoS Comput. Biol.* 1 (4), e42.
- Suárez, L.E., Markello, R.D., Betzel, R.F., Mísic, B., 2020. Linking structure and function in macroscale brain networks. *Trends Cognit. Sci.* 24 (4), 302–315. doi:10.1016/j.tics.2020.01.008.
- Thirion, B., Varoquaux, G., Dohmatob, E., Poline, J.-B., 2014. Which fmri clustering gives good brain parcellations? *Front. Neurosci.* 8 (167), 13.
- Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., Joliot, M., 2002. Automated anatomical labeling of activations in sm using a macroscopic anatomical parcellation of the mni mri single-subject brain. *Neuroimage* 15 (1), 273–289.
- Van Essen, D.C., Anderson, C.H., Felleman, D.J., 1992. Information processing in the primate visual system: an integrated systems perspective. *Science* 255 (5043), 419–423.
- Van Essen, D.C., Jbabdi, S., Sotiropoulos, S.N., Chen, C., Dikranian, K., Coalson, T., Harwell, J., Behrens, T.E., Glasser, M.F., 2013. Mapping connections in humans and non-human primates: aspirations and challenges for diffusion imaging. *Diffusion MRI, 2nd edition* (eds. Johansen-Berg, H. & Behrens, T.E.J.) 337–358.
- Van Essen, D.C., Smith, S.M., Barch, D.M., Behrens, T.E., Yacoub, E., Ugurbil, K., 2013. The wu-minn human connectome project: an overview. *Neuroimage* 80, 62–79.
- Van Essen, D.C., Ugurbil, K., Auerbach, E., Barch, D., Behrens, T., Bucholz, R., Chang, A., Chen, L., Corbetta, M., Curtiss, S.W., et al., 2012. The human connectome project: a data acquisition perspective. *Neuroimage* 62 (4), 2222–2231.
- Wang, C., Ng, B., Garbi, R., 2018. Multimodal brain parcellation based on functional and anatomical connectivity. *Brain Connect.* 8 (10), 604–617.
- Wang, J., Yang, Y., Fan, L., Xu, J., Li, C., Liu, Y., Fox, P.T., Eickhoff, S.B., Yu, C., Jiang, T., 2015. Convergent functional architecture of the superior parietal lobule unraveled with multimodal neuroimaging approaches. *Hum. Brain Mapp.* 36 (1), 238–257.
- Wu, Z., Xu, D., Potter, T., Zhang, Y., 2019. Effects of brain parcellation on the characterization of topological deterioration in alzheimers disease. *Front. Aging Neurosci.* 11, 113.
- Xu, J., Moeller, S., Strupp, J., Auerbach, E., Chen, L., Feinberg, D., Ugurbil, K., Yacoub, E., 2012. Highly accelerated whole brain imaging using aligned-blipped-controlled-aliasing multiband epi. In: *Proceedings of the 20th Annual Meeting of ISMRM*, Vol. 2306.

Yeo, B.T., Krienen, F.M., Sepulcre, J., Sabuncu, M.R., Lashkari, D., Hollinshead, M., Roffman, J.L., Smoller, J.W., Zöllei, L., Polimeni, J.R., et al., 2011. The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *J. Neurophysiol.* 106 (3), 1125–1165.

Zalesky, A., Fornito, A., Cocchi, L., Gollo, L.L., van den Heuvel, M.P., Breakspear, M., 2016. Connectome sensitivity or specificity: which is more important? *NeuroImage*.

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

Zhu, M., Lu, A.Y., 2004. The counter-intuitive non-informative prior for the bernoulli family. *J. Stat. Educ.* 12 (2).