Automated Iterative Reclustering Framework for Determining Hierarchical Functional Networks in Resting State fMRI

Seyed-Mohammad Shams,[1,2] **Babak Afshin-Pour,**[2]
Hamid Soltanian-Zadeh,[1,3,4]* **Gholam-Ali Hossein-Zadeh,**[1,3] and
Stephen C. Strother[2,5]

[1]*Control and Intelligent Processing Center of Excellence (CIPCE), School of Electrical and Computer Engineering, College of Engineering, University of Tehran, Tehran, Iran*
[2]*Rotman Research Institute, Baycrest, Toronto, Ontario, Canada*
[3]*School of Cognitive Sciences, Institute for Research in Fundamental Sciences (IPM), Tehran, Iran*
[4]*Department of Radiology, Image Analysis Laboratory, Henry Ford Hospital, Detroit, Michigan*
[5]*Department of Medical Biophysics, University of Toronto, Toronto, Ontario, Canada*

Abstract: To spatially cluster resting state-functional magnetic resonance imaging (rs-fMRI) data into potential networks, there are only a few general approaches that determine the number of networks/clusters, despite a wide variety of techniques proposed for clustering. For individual subjects, extraction of a large number of spatially disjoint clusters results in multiple small networks that are spatio-temporally homogeneous but irreproducible across subjects. Alternatively, extraction of a small number of clusters creates spatially large networks that are temporally heterogeneous but spatially reproducible across subjects. We propose a fully automatic, iterative reclustering framework in which a small number of spatially large, heterogeneous networks are initially extracted to maximize spatial reproducibility. Subsequently, the large networks are iteratively subdivided to create spatially reproducible subnetworks until the overall within-network homogeneity does not increase substantially. The proposed approach discovers a rich network hierarchy in the brain while simultaneously optimizing spatial reproducibility of networks across subjects and individual network homogeneity. We also propose a novel metric to measure the connectivity of brain regions, and in a simulation study show that our connectivity metric and framework perform well in the face of low signal to noise and initial segmentation errors. Experimental results generated using real fMRI data show that the proposed metric improves stability of network clusters across subjects, and generates a meaningful pattern for spatially hierarchical structure of the brain. *Hum Brain Mapp 36:3303–3322, 2015.* © 2015 Wiley Periodicals, Inc.

INTRODUCTION

The brain can be viewed as a complex network of functionally and structurally interconnected regions, which can be seen as hierarchical networks. In this hierarchical structure, each network may consist of subnetworks and each subnetwork is further composed of sub–subnetworks [Ferrarini et al., 2009; Meunier et al., 2010]. Bassett et al. [2010] presented clear evidence of a hierarchical organization in brain networks, and several studies in the field of neuroscience have focused on extracting properties of this organization. These include the number of networks, subnetworks, and sub–subnetworks, and the relationships among networks and subnetworks. fMRI has been extensively used to explore these networks and to extract their patterns of regional brain interactions in terms of changes in the blood oxygenation level dependent (BOLD) signals either during task or rest conditions.

Rs-fMRI is commonly used for investigating neural mechanisms in the brain when a subject is performing no externally directed task [Greicius et al., 2003; Beckman et.al. 2005; Fox and Raichle, 2007]. Rs-fMRI studies have focused on the mapping of the patterns of functional connectivity through measuring the level of synchronization between neurovascular events in anatomically separated brain regions [Friston et al., 1993; Fox et al., 2005; Li et al., 2009]. While the initial rs-fMRI study [Biswal et.al., 1995] illustrated the temporal correlations between a few predefined regions of interest (ROIs), many methods have since been proposed to extract whole-brain resting state networks (RSNs) formed by discrete sets of functionally connected regions. The terms "component," "network," and "cluster" are used interchangeably in the literature to refer to a RSN.

Analysis of the rs-fMRI data provides an efficient approach to probe effects of brain disease (e.g., schizophrenia, depression, and Alzheimer's) on brain networks [Grigg and Grady, 2010; Greicius, 2008; Schwindt et al., 2013]. Although traditional model-driven methods (e.g., seed-based univariate [Biswal et al., 1995], and multivariate [Grigg et al., 2010] [Afshin-Pour et al., 2014] techniques) are easy to implement and interpret, they suffer from the need to define proper seeds to generate data appropriate for specific research questions and hypotheses. Hence, many of the recent rs-fMRI studies identify RSNs based on functional connectivity among all brain regions [Cole et al., 2010]. Without defining seed regions, model-free methods (e.g., independent component analysis (ICA) [Beckmann et al., 2005; Calhoun et al., 2009], principal component analysis (PCA), [Afshin-Pour et al., 2014], and clustering [Bellec et al., 2010]) examine the overall patterns of connectivity among all regions in the brain instead of examining connectivity with a single seed region or among a small number of ROIs [van den Heuvel and Hulshoff Pol, 2010; Li et al., 2009]. All these approaches help to explore organization of the brain's functional connectivity and their relative utility is an important topic of study in the neuroimaging community [Afshin-Pour et al., 2014; Cole et al., 2010; Yan et al., 2011].

Clustering methods may be considered as model-free techniques that do not assume an explicit distribution for temporal vectors, that is, time series. In fMRI, clustering methods parcel the temporal vectors into several sets such that the time-series within each set have a high level of pair-wise similarity. The regions associated with the time-series of each set are functionally connected, and form a brain network through their similar, for example, correlated, time series. Hence, the clustering techniques have similarities with seed-based methods and their results are relatively easy to interpret [Margulies et al., 2010; van den Heuvel and Hulshoff Pol, 2010]. Moreover, in contrast to the more common use of ICA, clustering techniques do not require visual inspection and selection of relevant components, thus avoiding subjective human errors. Therefore, a variety of clustering algorithms have been applied to rs-fMRI data to partition the brain into clusters of voxels or regions that are functionally connected (i.e., RSNs). These algorithms implement hierarchical [Cohen et al., 2008; Salvador et al., 2005; Cordes et al., 2002], K-means [Mezer et al., 2009], and spectral clustering [van den Heuvel et al., 2008] approaches. In addition to the clustering algorithms, two other factors may generally affect the clustering results: the similarity measure and number of clusters. Quantifying the synchronization between regional time series, as a pair-wise similarity measure, plays an important role in functional connectivity analysis of RSNs. Therefore, a variety of metrics has been proposed to estimate such spatial coactivations [Cao and Worsley, 1999; Sun et al., 2004].

In spite of successful application of clustering methods in extracting RSNs, the number of network clusters strongly affects the results. In cluster analyses, typically 7–11 RSNs are extracted [Bellec et al., 2010; van den Heuvel et al., 2008]. For individual subjects, a large number of disjoint clusters may be used to cover all gray matter results in many small spatial networks that are spatio-temporally homogeneous but spatially irreproducible across subjects while providing a fine-detailed segmentation of each brain [Ma et al., 2011]. Alternatively, using a small number of clusters creates spatially large networks that are temporally heterogeneous but more spatially reproducible across runs, sessions, and subjects.

In addition, validation of RSNs is not trivial due to lack of "ground truth" [Bellec et al., 2010]. Hence any clustering strategy should deal with spatial reproducibility and within network temporal homogeneity of the derived RSNs as data-driven quality measures; and a clustering framework has to carefully evaluate these characteristics of the identified networks. Almost all of the clustering methods attempt to parcel the data into a predefined number of disjoint clusters so that these clusters have high intracluster similarity and low intercluster similarity. However, achieving this depends on the number of clusters chosen to avoid intracluster heterogeneity where two clouds of time series vectors are assigned to a single cluster while their similarity is relatively low.

Several works used the stability of clusters across different sessions to determine the number of clusters in fMRI data sets [Bellec, 2013; Bellec et al., 2010; Cai et al., 2014; Kelly et al., 2010, 2012; Thirion et al., 2014; Yeo et al., 2011]. However, they may fail in determining the correct number of clusters on datasets with hierarchical network structure [Bellec, 2013]. Although the exact nature of the structure of brain networks derived from rs-fMRI is still a matter of debate, the notion of a hierarchical structure of the brain is accepted and incorporated into some models [Ferrarini et al., 2009; Meunier et al., 2010]. Extraction of such a hierarchical structure is considered an important issue in the identification of brain networks.

In this study, we propose an automated, iterative reclustering framework to extract hierarchically organized RSNs and subnetworks from rs-fMRI data of a group of subjects. We will show that modeling of the brain networks as a hierarchal structure helps to reliably estimate the brain networks. This is due to the fact that for each iteration, the proposed algorithm estimates only a few networks from a relatively small similarity matrix. This means scaling down the dimensionality of the problem in an iteration, which results in a more reliable estimation of the brain networks.

The proposed framework includes two novel contributions to address the aforementioned issues: (1) new temporal similarity measures that quantify the functional connectivity, that is, energy ratio (ER) and weighted energy ratio (WER) and (2) a new iterative reclustering approach that discovers the network hierarchy in the brain by iteratively subdividing large heterogeneous networks to create new homogeneous subnetworks. To determine the number of subnetworks at each iteration, we maximized a spatial reproducibility metric as a cluster stability measure.

The proposed pair-wise, temporal similarity measure is defined based on an orthogonal projection. We show that in comparison to conventional measures, this similarity measure provides more spatially reproducible experimental results for real fMRI data sets. Since we implement the proposed framework as a region-based approach, we start by segmenting the brain into connected regions based on region growing [Bellec et al., 2010]. Then, we extract the time series associated with each connected region using PCA. In this approach, the number of time series is reduced with minimum loss of information. The similarity between two regions is then computed based on the ratio of signal subspace energy to the residual energy resulting from orthogonal projection of one region's time series onto the signal subspace spanned by another region's time series. The concept of ER has been used before in [Hossein-Zadeh et al., 2003] in activation detection in fMRI. It may also be considered as a multivariate regression signal to noise ratio measure. The ER approach can be easily replaced by other temporal similarity metrics that measure similarity between two regional subspaces. Hence, we briefly explain some important metrics below and discuss their merits by comparing experimental results obtained from real fMRI data sets as a function of the spatial reproducibility of the network results across subjects.

In the second step, we propose a spatial reproducibility metric that is constructed on the basis of split-half resampling and Jaccard similarity coefficients. Split-half resampling has been used successfully in NPAIRS to obtain prediction and reproducibility metrics for optimization of functional neuroimage processing and analysis [Strother et al., 2002, 2004]. Such subsampling has strong theoretical support for stabilizing resampling results in difficult high dimensional variable selection and regularization problems such as clustering [Meinshausen and Bühlmann, 2010]. We introduce the measurement of spatial reproducibility and illustrate the application of the proposed metric to: optimize the number of networks in the clustering approach in terms of maximizing the reproducibility of RSNs, and evaluate the performance of different similarity measures to provide the most reproducible networks that can be inferred from real fMRI data sets in group studies. Hence, the reproducibility measurement can be used as a criterion to evaluate and compare the results of different clustering approaches.

Subsequently, a temporal homogeneity measure is calculated for each reproducible network. The network's regions are iteratively subdivided until the overall homogeneity does not change substantially. The number of successive iterations creating subnetworks is determined in the same way as that of the first coarse network clustering scale, based on the reproducibility criterion. This iterative reclustering framework finally defines potential network clusters that are both spatially reproducible across subjects and temporally homogeneous within each network.

In the next section, we state the proposed clustering algorithm in detail, and introduce different possible temporal similarity measures including the ER and WER. To assess the performance of the proposed clustering algorithm and compare it with other approaches, we define a homogeneity measure for the clustering results. Also, we use the NPAIRS split-half resampling framework to estimate the reproducibility of the results. Using these two quality metrics, the proposed clustering algorithm is compared with a common spectral clustering in terms of the

homogeneity and reproducibility of the resulting clusters. Also, we compare different temporal similarity metrics and explore their effects on the final clusters.

MATERIALS AND METHODS

Real fMRI Data

In all subjects, a functional imaging sequence and a high-resolution anatomical sequence were acquired with a Siemens Trio 3T MRI magnet. In the fMRI experiment, T2* functional images were acquired using an EPI pulse sequence (TE = 30 ms, TR = 2,000 ms, flip angle = 70°, FOV = 200 mm). For each subject, a T1-weighted anatomical volume was also acquired using SPGR (TE = 2.6 ms, TR = 2,000 ms, FOV = 256 mm, slice thickness = 1 mm). During the acquisition of the resting-state fMRI data, the subjects were instructed to keep their eyes closed and the physiological data, that is, signals heart rate and respiratory, were also recorded during the scanning. The resting-state fMRI data of 31 healthy young adult subjects (20–30 years) were collected during a rest period of 340 s (i.e., 170 volumes). For each subject, a T1-weighted anatomical volume was also acquired using SPGR (TE = 2.6 ms, TR = 2,000 ms, FOV = 256 mm, slice thickness = 1 mm) for coregistration with the functional images.

Preprocessing

Preprocessing procedures were chosen based on Van Dijk et al. [2010], and performed using routines in the AFNI software package (afni.nimh.nih.gov/afni, version 2011_12_21_1014), with the exception of steps for band-pass filtering, nuisance components' regression, and normalization by voxels' standard deviations, which were executed in MATLAB R2011b (http://www.mathworks.com/). AFNI-based preprocessing included slice-scan time correction with Fourier interpolation followed by physiological artifact removal using RETROICOR. Using the VOLREG routine, the data were motion corrected to minimize the effects of head motion on subsequent analyses. The 50th scan of each subjects' fMRI data was taken as the target scan, and the remaining scans of the subjects were individually registered using 7th order Lagrange polynomial interpolation to their target scan. To transfer functional data to MNI[1] space, for each subject two transforms were estimated: a transform that registers the fMRI target scan of a subject to its structural image, and a transform that registers the structural image to the MNI template. These transforms were combined to constitute one aggregated transform, and subsequently each subject's scans were transformed to MNI space using its individual aggregated transform. Afterwards, the registered data were

resliced to a voxel size of 4 mm^3. To improve signal detection, spatial smoothing was achieved by applying a 3D spatial Gaussian filter with full width half maximum = 8 mm on the data in MNI space. Subsequently, very low frequency fluctuations and high frequency noise were removed using a Butterworth band-pass filter with the pass band from 0.01 to 0.1 Hz. The maximum attenuation for the pass band (0.01–0.1 Hz) is 1 db, and the minimum attenuation for stop bands (0–0.005 Hz and 0.125–0.25 Hz) is 20 db. The first 20 scans of each subject were omitted to remove the transient response effect of the band-pass filter. To estimate nuisance regressors, white matter and ventricle masks were generated by creating a binary mask from the ICBM[2] white matter and CSF[3] probabilistic maps [Mazziotta et al., 2001], based on a threshold level of 0.8 (≥80% chance of containing white matter or CSF). Then the subsequent masks were eroded using a 3 × 3×3 square morphological structuring element. For each subject, the white matter and ventricles' mean time-series were regressed out from all time-series. Finally, the time-series of each subject were mean-removed and normalized to their standard deviation. The preprocessing reduced the dimensionally of the data significantly. Generally, for each subject's preprocessed data, 95% of the variance can be described by its first 48 eigenvectors.

To extract functional regions, we used an Automated Anatomical Labeling atlas [AAL; Tzourio-Mazoyer et al., 2002] as an initial segmentation of the brain into 116 regions. This initial parcellation guarantees that the subsequent brain regions will include the voxels only from one anatomical segment. Also, it limits the computational demand [Tzourio-Mazoyer et al., 2002]. The white matter region was discarded as it is not expected to display significant BOLD activations during resting-state. Each of the remaining anatomical regions was segmented into several regions based on a region growing algorithm. We started with a seed voxel within an AAL segment. Since the region growing is sensitive to the seed, we considered the seed as the voxel with the time-series that had the highest mean correlation with the other time-series in the AAL segment. Then, the adjacent voxels were connected to the seed if their correlations were higher than a certain threshold, that is, 0.7 in this article. These voxels and the seed voxel formed a region. We continued adding the adjacent voxels to the region if their correlations with the seed time-series were above the defined threshold. The algorithm stops when there are no more voxels to add with a correlation above the threshold. We excluded the formed region from the AAL segment and reapplied the region growing with a new seed from the remaining voxels within the AAL segment until each voxel was assigned to

---

[1]Montreal Neurological Institute (http://www.bic.mni.mcgill.ca/ServicesAtlases/HomePage).

[2]International Consortium for Brain Mapping.

[3]Cerebrospinal fluid.

a region. We continued performing this algorithm on all AAL segments.

In the region-growing algorithm, time courses of spatially similar voxels in different subjects were concatenated in the temporal domain. The region-growing algorithm uses the correlation coefficient between the concatenated time-series as the similarity metric between voxels. Using this framework, the initial AAL regions were split into 680 regions. These 680 regions were clustered into several networks using the proposed framework described below.

Simulated Data Sets

We used simulated data sets to evaluate the ability of the proposed framework in determining the true number of cluster networks as well as the accuracy of the extracted networks. Our simulated data sets consisted of 5 fixed, disjoint networks (K) comprising 8 separate fixed, disjoint regions (R) for each network distributed over N_v voxels per region in 150 volumes per subject. A common, random stimulation pattern was generated for regions within each network. These random temporal patterns were convolved with the hemodynamic response function (HRF) to produce the simulated BOLD signal of each cluster. The HRF was modeled according to the following Gamma function [Friston et al., 1998; Lange and Zeger, 1997]:

$$h(t; \tau, \sigma) = \begin{cases} \exp\left(-t/\sqrt{\sigma t}\right)\left(\dfrac{e.t}{\tau}\right)^{\sqrt{\tau/\sigma}} & t > 0 \\ 0 & t < 0 \end{cases} \quad (1)$$

To model the HRF variations in different regions within a network, the τ and σ HRF parameters were randomly selected in the ranges of [3, 7] and [0.05, 0.21], respectively. Then, by adding fractional Gaussian noise with the Hurst exponent equal to 0.8, and different variances to the simulated activation time series, voxel time series were generated (for more detail on fractional Gaussian noise see [Maxim et al., 2005]). The noise variances were chosen to acquire voxel-wise signal to noise ratios (SNRs) of -15, -10, -5 (dB). Here, SNR is defined as the 20 $\log_{10}(STD_{signal}/STD_{noise})$, where STD represents the standard deviation of noise and signal. Similar to the approach proposed in [Bellec et al., 2010], we prepared different sets of simulated data by adjusting the following free parameters: the number of subjects involved in the analysis (N_S), the number of voxels in each region (N_V), and the average SNR in a network.

To explore the effect of error in the initial segmentation of gray matter we considered $P\%$ of the voxels in each region of a network to be mis-segmented and set them to have the activation signal of randomly chosen networks. The parameter P is referred to as the nonoverlapping parameter throughout the article, which specifies the percentage of nonoverlap between the segmented and actual regions. To study the effect of "size of segmented regions"

in our connectivity measurement and framework, the number of voxels in a region was also varied. The free parameters were selected in the following ranges: $N_S = \{10, 16, 22\}$; $N_V = \{10, 20, 30\}$; SNR$= \{-15, -10, -5\}$; and $P = \{0, 10, 20, 30, 40\}$. Note that N_s is limited to the range of 10–22 to span relatively weak to moderate group-wise signal-to-noise ratios. Also, N_s is chosen as an even number to avoid unbalanced split-halves in the resampling procedure.

Proposed Clustering Framework

In this section, we present an overview of the iterative reclustering framework that automatically produces a hierarchical structure of functional brain networks (see Fig. 1). In the following subsections, each block of the approach will be presented in detail. In the first step, we apply PCA on the time series data of each segmented region and select the first few eigenvectors that express 70% of the variance as the subspace representing the activations in that region. In the Supporting Information Figure S2 shows that 70% of the variance provides the most reproducible networks of the brain. Then, we compute similarity matrices for each subject using the PCA subspaces to obtain a temporal similarity measure between all pairs of the 680 segmented regions. Thereafter, we apply the clustering algorithm (see the next subsection) on the segmented regions to estimate the optimal number of clustered networks, and the reproducibility of the adjacency matrix describing the estimated functional connections in the brain. Afterwards, the homogeneity of each network cluster is measured by averaging the pair-wise similarities between the subspaces of the spatially separated regions within that network cluster. The similarities may be assessed by different metrics, and we describe and evaluate some common similarity metrics later in the article (see Reproducibility Metric with Split-half Resampling subsection). The estimated networks are sorted from the least homogeneous to the most homogeneous. Then, these networks are reentered into the clustering framework for up to m separate, reclustering iterations. In each iteration, the least homogenous network is selected to split into several subnetworks using the clustering algorithm. Then these new networks and the remaining networks from previous iterations are again sorted based on their homogeneity, and once again the least homogenous network is fed into the clustering algorithm. The iterations continue until the average homogeneity of all networks does not increase substantially by performing a new iteration. For each iteration, our approach provides the number of networks (i.e., K_m), the estimated segmented regions in each network, the total reproducibility (R_m), and the average homogeneity (i.e., H_m).

Clustering algorithm

Figure 2 presents a flowchart of the clustering algorithm and our scheme for estimating the number of clusters. For the ith iteration denote the estimated cluster networks
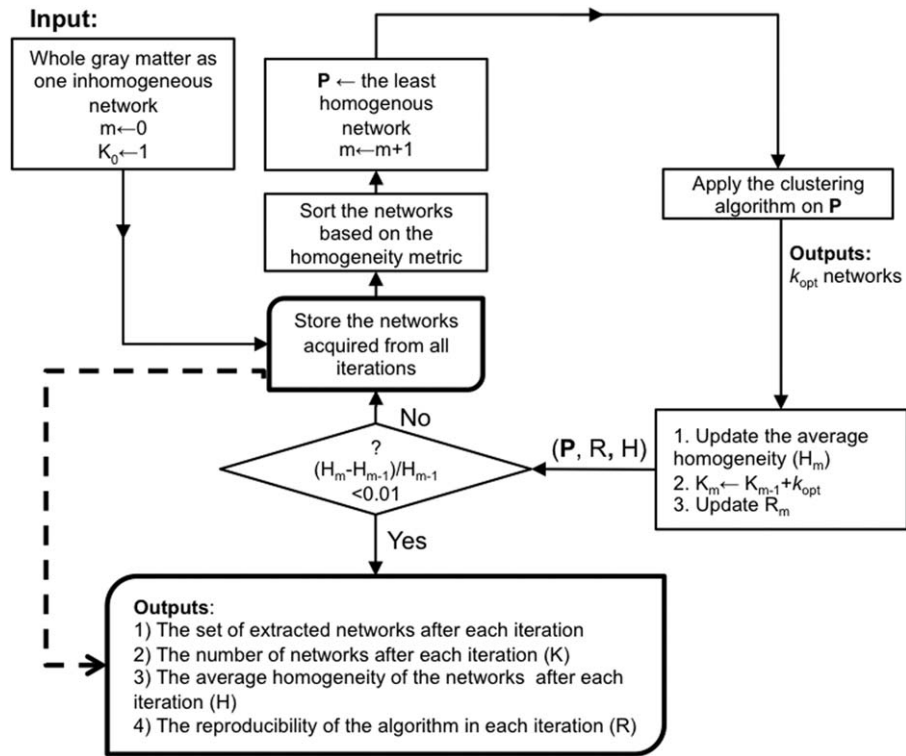
Figure 1.

The flowchart shows the proposed approach for the estimation of the brain networks. In each iteration, the least homogenous network is split into several subnetworks. The iteration continues until the percent change in homogeneity is less than 1%.

ordered by homogeneity as $RSN_1, ..., RSN_{Km-1}$, with RSN_1 the least homogenous network. In the next step, the algorithm clusters RSN_1 into k subnetworks. To determine the number of subcluster networks of RSN_1, we perform split-half resampling of the subjects 300 times and measure the reproducibility of the clustering results. For each split, this step: (1) produces split-similarity matrices by averaging the individual subjects' similarity matrices in each split; (2) extracts a binary adjacency matrix for each split by spectral clustering of the averaged, split-similarity matrix; and (3) computes spatial reproducibility as a function of the number of clusters (k) using a Jaccard similarity metric (see the next subsection). The above procedure is repeated for $k = 1, ..., k_{max}$ clusters and the "number of clusters" that maximizes the reproducibility of the clustering results (k_{opt}) is chosen as the optimal number, and two binary adjacency matrices were then added together to produce the split-half adjacency matrix for this splitting. Finally, we cluster the average of the 300 split-half adjacency matrices to extract the most reproducible clusters (i.e., network clusters) across subjects.

We have focused on clustering using the spectral clustering algorithm with "normalized cuts," out of the avail-

able variants of spectral clustering, because of its simple implementation and solution by known optimized linear algebra tools [Shi and Malik, 2000]. However, other clustering algorithms may also be used in the proposed framework.

Reproducibility metric with Split-half resampling

Cross-validation resampling schemes have been widely used to estimate model parameters. The basic idea of these procedures is to split the data into two independent sets (training and test) and use them to estimate model parameters and evaluate the accuracy of the estimated model [Stone, 1974]. Split-half resampling, which has been successfully used in the context of neuroimaging in the NPAIRS framework, splits the data into two independent, approximately equal-size sets and performs up to $^NC_{N/2}/2$ different data splits for N independent data sets, where $^NC_{N/2}$ represents the number of possible combinations of N objects taken $N/2$ objects at a time [Strother et al., 2002, 2004]. Similarly, we choose subjects as the basic resampling unit, and use the split-half resampling procedure to
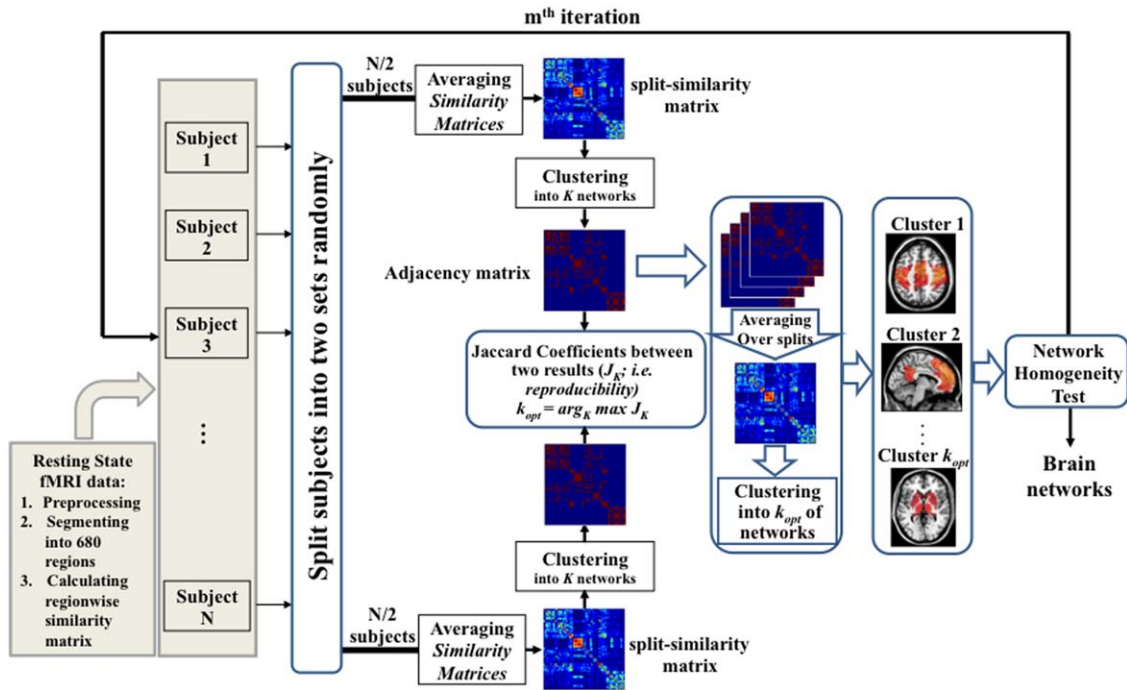
Figure 2.

The flowchart shows the split-half resampling based clustering algorithm in which the most heterogeneous network (i.e., cluster) is clustered into k_{opt} subnetworks. The algorithm splits the N subject into two sets randomly, and calculates two split-half similarity matrices. Note that only the regions within the input network are taken into account, and the similarity matrices describe the connections between the regions within the input network. Using the similarity matrices, the algorithm clusters the regions into k nonoverlapping clusters ($k = 1$, $k_{max} = 45$), and provides two adjacency matrices. The Jaccard coefficient between two adjacency matrices indicates the reproducibility of the algorithm. The split resampling framework is repeated several times. The final reproducibility is defined as the median of the Jaccard coefficients over different random splits. The k value that optimizes reproducibility is considered as the optimum number of clusters (k_{opt}). The adjacency matrices acquired for k_{opt} and for different split-half resampling are averaged Then, the clustering algorithm is applied on the average adjacency matrix, and the final k_{opt} clusters, that is, subnetworks are produced. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

nonparametrically calculate the reproducibility histograms of clustering as a function of "number of clusters."

We use B replications of split-half resampling. In each split, we average the subjects' similarity matrices of each half to yield two split-similarity matrices, which are considered the resampled observations. Clustering is only performed on the least homogeneous network (e.g., RSN$_1$), for which the averaged split-similarity matrices are clustered into $k=\{1,...,k_{max}\}$ clusters resulting in two different partitions: $P_k^1 = \{RSN_{1\text{-}1}^1, RSN_{1\text{-}2}^1... RSN_{1\text{-}k}^1\}$ and $P_k^2 = \{RSN_{1\text{-}1}^2, RSN_{1\text{-}2}^2 ... RSN_{1\text{-}k}^2\}$. To evaluate the reproducibility, we measure the similarity between partitioning results computing the maximum intercluster similarity for each RSN_{1-i}^1 across all RSN_{1-j}^2 by, $Jaccard\left(RSN_{1\text{-}i}^1, RSN_{1\text{-}j}^2\right)$. We define the reproducibility of the networks $RSN_{1\text{-}i}^1$ as follows:

$$\text{Rep}\left(RSN_{1-i}^1\right) = \max_j \text{Jaccard}\left(RSN_{1-i}^1, RSN_{1-j}^2\right) \quad (2)$$

Then, for k clusters, the similarity of two partitioning results is measured by averaging the k maximum-Jaccard-coefficients for RSN_{1-i}^1 with all RSN_{1-j}^2. By this procedure, we measure the reproducibility of the clusters for one pair of split-half as a function of k (i.e., number of clusters) and b (i.e., the split number and $b \in \{1,...,B\}$), $\hat{J}\{b\}(k)$. This can be mathematically expressed as:

$$\hat{J}\{b\}(k) = \frac{1}{k}\sum_{i=1}^{k} \text{Rep}\left(RSN_{1-i}^1\right) \quad (3)$$

We use the J values from all data splits to generate a reproducibility histogram for each "number of clusters," which for each k is summarized using its median over the entire b, $\bar{J}(k) = median\left(\hat{J}\{b\}(k); 1 \le b \le B\right)$, to avoid sensitivity to outlier values. Consequently, k_{OPT} is chosen to maximize the reproducibility of the results as

$$k_{OPT} = \text{argmax}_k\left(\bar{J}(k)\right) \quad (4)$$

The final networks, RSN_{1-i}, output for homogeneity testing in Figure 2 are generated by averaging the 300 k_{OPT} adjacency matrices acquired from different split-half

resamplings, and reapplying the spectral clustering algorithm on the average adjacency matrix to detect the final k_{OPT} cluster networks. Another strategy to acquire a group clustering is to apply the clustering on the average similarity matrices. We compared these two strategies later in the results (See Fig. 8). The final clusters RSN_{1-i} from the average adjacency matrix are matched with the estimated split-half cluster networks to obtain their Jaccard coefficient. We consider the reproducibility of RSN_{1-i} as the average reproducibility of its matching clusters in the split-half resampling procedures. If we assume RSN_{1-i} is matched with RSN^1_{1-i} then its reproducibility is defined as:

$$\text{Rep}(RSN_{1-i}) = \text{Rep}(RSN^1_{1-i}) =$$
$$\text{max}_j\text{Jaccard}\left(RSN^1_{1-i}, RSN^2_{1-j}\right) \qquad (5)$$
$$-\text{max2}_j\text{Jaccard}\left(RSN^1_{1-i}, RSN^2_{1-j}\right); 1 \leq i \leq k_{OPT}$$

where max2 returns the second maximum element. Generally, in low numbers of clusters, there is a higher chance to estimate similar clusters across the two splits. The second term in equation (5) removes that bias (See the Supporting Information for the bias assessment of different measures). After performing the above clustering procedure, the RSN_1 will be replaced with the subnetwork $RSN_{1-1},\ldots,RSN_{1-kopt}$. Consequently, the number of the cluster networks at the current iteration is updated as follows:

$$K_m = (K_{m-1} - 1) + k_{opt} \qquad (6)$$

In addition, we calculate the global reproducibility of the approach in the current iteration (i.e., R_m) as follows:

$$R_m = \begin{cases} R_{m-1} - w_1\text{Rep}(RSN_1) + \sum_{i=1}^{k_{opt}} w_{1-i}\text{Rep}(RSN_{1-i}); & m>1 \\ \sum_{i=1}^{k_{opt}} w_i\text{Rep}(RSN_i); & m = 1 \end{cases}$$
$$(7)$$

where w_i, and w_1-_i are the relative number of voxels in the whole brain included in RSN_1, RSN_{1-i}.

Between-region temporal similarity

The cross correlation between the average time series of two regions has been widely used to estimate their functional connectivity. This region-based, pairwise similarity reduces the multivariate problem to a bivariate one, using the average over the voxels in each region. To avoid this reduction and its simplifying assumptions about the local covariance structure, multivariate similarity measures, like canonical correlation analysis and RV coefficients, can be used to measure the association between times series from two regions incorporating the local covariance [Robert and Escoufier, 1976]. These are descriptive measures of the correlation between two sets and can be easily combined

with region-based functional connectivity algorithms [Zhang et al., 2010].

In the appendix, we briefly review two conventional similarity measures (RV-coefficient and canonical correlation analysis, CCA) and propose two additional similarity measures (ER and WER). We test the utility of these 4 measures in our proposed clustering framework. In all cases, we first apply PCA on the time series data of each region and select the first few eigenvectors that express 70% of the variance as the subspace representing the time series of that region. Then, we apply pair-wise functional connectivity measures to these subspaces.

Within-network homogeneity

To quantify within-network (temporal) homogeneity (H_i, i.e., overall connectivity in the network), we average the pair-wise temporal similarity between the regions in each of the k cluster networks. Then, we record the distribution and median across the k networks to provide an overall homogeneity measure.

RESULTS

Results of Simulated Data

We applied the proposed framework on the simulated data sets to produce reproducibility histograms for 50 splits, and grouped the simulated regions into K clusters to obtain the number of reliable network clusters and regions per cluster in the data sets. We examined these two issues for 40 randomized simulations as a function of the nonoverlapping parameter (P), SNR, number of subjects (Ns), and number of voxels (Nv) and plot the results in Figure 3. Figure 3a shows the percentage of correct estimation of actual number of clusters, that is, 5 clusters as a function of P, SNR, Ns, and Nv. Each row indicates the results acquired for each value of $Ns = 10$, 16, and 22, and the columns represent the results for each value of Nv ($Nv=10$, 20, 30). Finally, the colors blue, green, and brown represent the results acquired for SNR$=-5$, -10, -15 dB, respectively.

Figure 3b represents the percentage overlap of the estimated networks with the true networks as a function of the nonoverlapping parameter (P), SNR, number of subjects (Ns) and number of voxels (Nv). The accuracy of the resultant network clusters is calculated by measuring the Jaccard metric between the true networks and the estimated networks. The rows and columns reflect the values of N_S (10, 16, 18) and N_V (10, 20, 30), respectively. Finally, the colors blue, green, and red represent the results acquired for SNR$=-5$, -10, -15 dB, respectively.

Figure 3a shows that the number of subjects does not significantly affect accuracy of the determined number of clusters and Figure 3b shows that it has only a small impact on the accuracy of the resultant network clusters for small regions and lower SNR values. In Figure 3a, when the SNR is large, that is, -5 dB (blue), or $P < 40\%$,
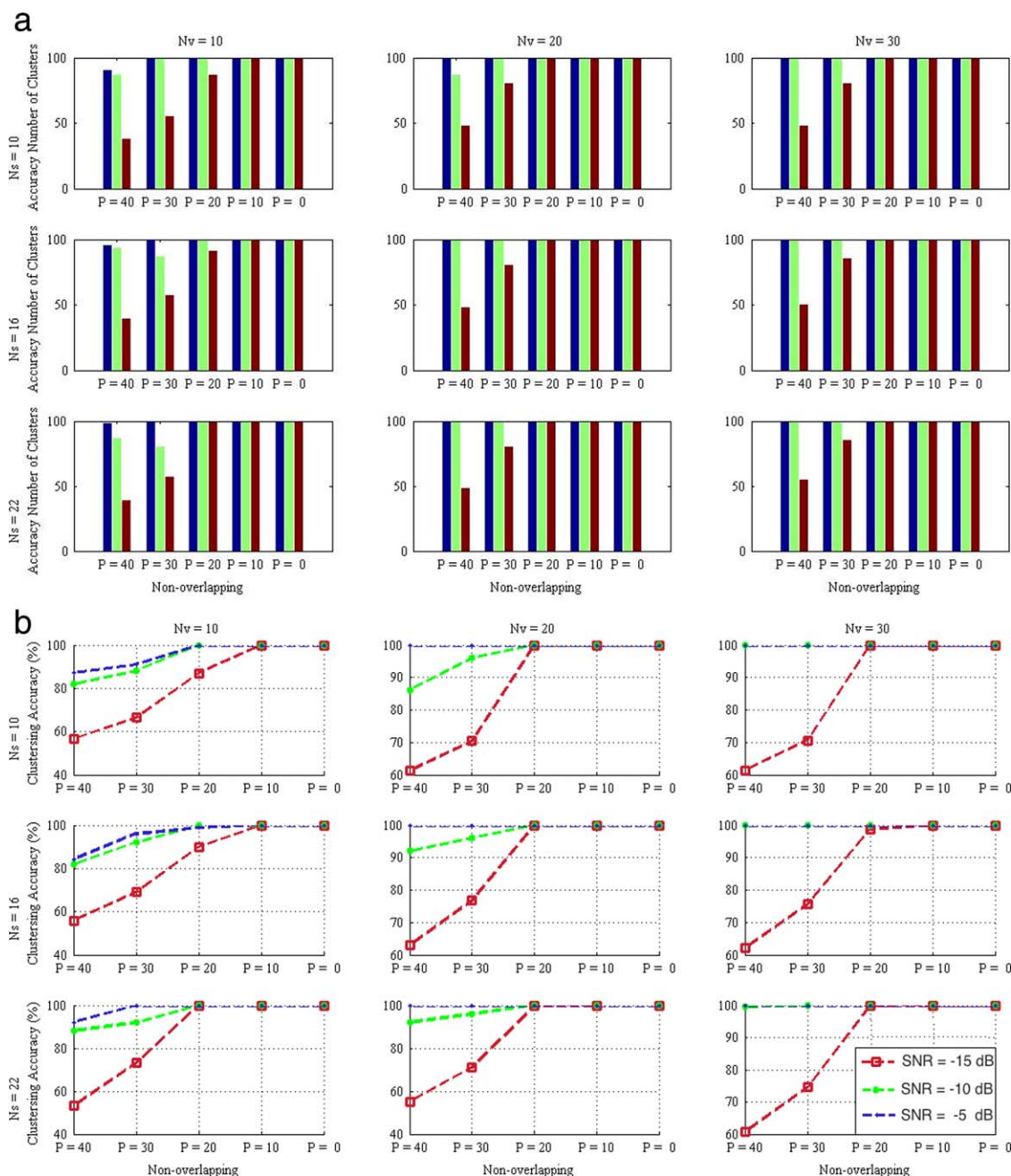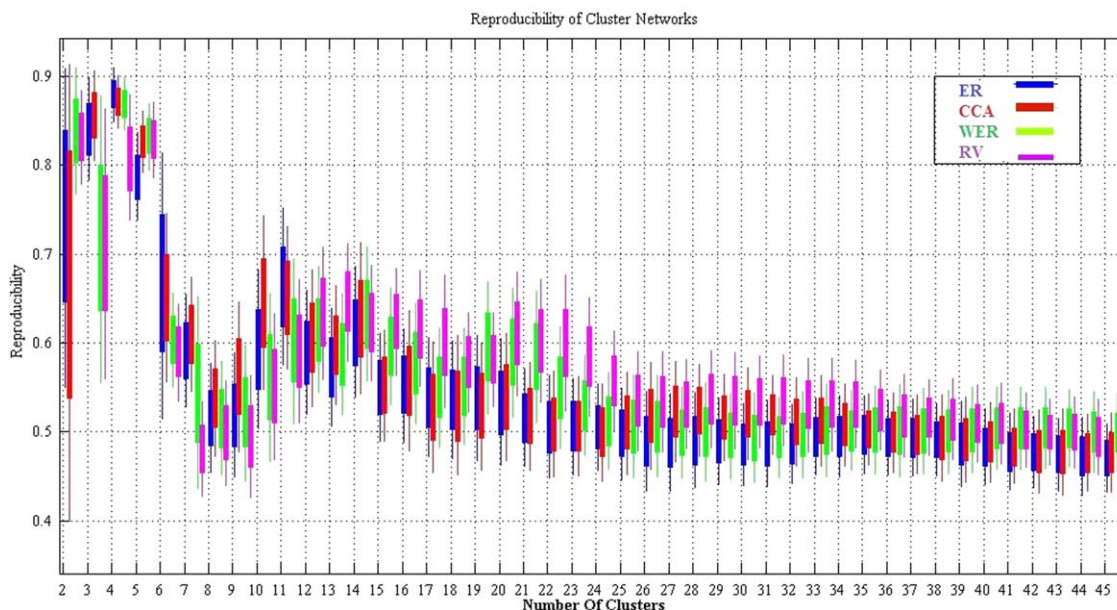
Figure 3.

The results acquired for the simulation dataset. (**a**) The percentage accuracy in estimating the correct number of clusters. (**b**) The percentage of the overlap between the estimated networks and the simulated networks, which is illustrated for different values of nonoverlapping parameter (*P*). Different colors represent the simulations with different SNRs. Rows represent the results for different number of subjects (*Ns*=10,16,22), and Columns indicate the results for different number of voxels (*Nv*=10,20,30). [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

the proposed framework successfully determines the correct number of clusters 100 percent of the realizations, regardless of *Nv* and *Ns*. Moreover, for *Nv* = 30, the

method only failed for the smallest SNR = −15 dB, and the largest lack of overlap between starting segmentations and true regions, *P* = 30 or 40%. Finally, Figure 3b shows

Figure 4.

Reproducibility acquired for the first iteration of the clustering algorithm versus the number of clusters ($k = 1$, $k_{max} = 45$). The clustering algorithm was applied on the real fMRI dataset with different similarity measures, that is, RV, CCA, ER, and WER. The boxplots represents the distribution of the reproducibility values acquired over 50 split-half resamples. $k = 4$ generates the most stable clustering in the first iteration. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

that the extracted networks are 100% accurate provided the regions are moderately sized (i.e., $N_V \geq 20$), and initial segmentation errors are not too large (i.e., $P \leq 20$). These simulated results indicate that the proposed method is quite accurate unless it encounters a combination of small cluster regions, with relatively low SNRs and large initial segmentation errors.

Results of Experimental fMRI Data

Evaluation

Four different similarity measures (RV coefficient, CCA, ER, and WER) were applied to the segmented fMRI data of each subject to construct four similarity matrices. Subjects were then randomly split into two independent, split-half sets, 300 times. For each split, the resultant average similarity matrices of each set were parceled into a range of $k = 2, \ldots, 45$ clusters using the normalized cut clustering technique. The Jaccard coefficients were calculated by Eq. (3) for each k. For the initial clustering step (i.e., $m = 0$) Figure 4 displays box-whisker plot distributions of the split-half reproducibility metrics, calculated using four different similarity measures, as a function of the number of clusters (k). It can be seen that all but RV of the similarity measures (ER, WER, and CCA) show their maximal median reproducibility (J) at $k = 4$ clusters while the RV measure shows the highest reproducibility at $k = 5$ clus-

ters. For ER, the greatest median reproducibility occurs at $k = 4$ clusters ($J_{ER}(4) = 0.88$), which is significantly larger than the second largest ER value at $k = 5$ ($P < 10^{-5}$, paired t test). Furthermore, for $k = 4$ ER is significantly larger than the second largest value of CCA ($P < 10^{-4}$). Hence, the optimal number of clusters is found to be four ($k_{OPT} = 4$). After $k = 4$ the reproducibility decreases sharply to a local minimum at 8 clusters ($J_{ER}(8) = 0.53$). Hence, this number of clusters produces the networks with the least reproducibility. After that local minimum, ER lifts slightly to a local maximum at 11 clusters ($J_{ER}(11) = 0.65$) which is significantly greater than $k = 10$ ($P < 10^{-5}$). After another local maximum at 14 clusters ($J_{ER}(14) = 0.62$), reproducibility decreases slowly to around 0.5 at 45 clusters.

Figure 5 shows the four network clusters formed according to the ER measure in the first step of our algorithm. Each of these four principal clusters is then reclustered to explore the hierarchy of its subclusters.

To evaluate the performance of the iterative reclustering framework under conditions similar to direct clustering (i.e., clustering in a one-step process with the same spectral clustering method using normalized cuts), we measured the cluster homogeneity and reproducibility metrics as a function of "number of clusters," that is, Figure 6a,b, respectively. Figure 6a illustrates the homogeneity measure of the resulting networks versus "number of clusters" ($k \in [1, \ldots, 26]$) for the ER method. The homogeneity for $k = 1$ indicates the initial homogeneity of the 680 segmented
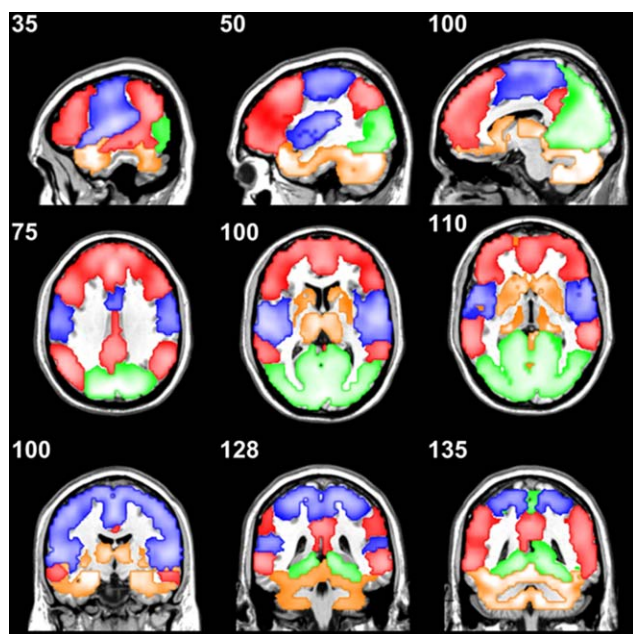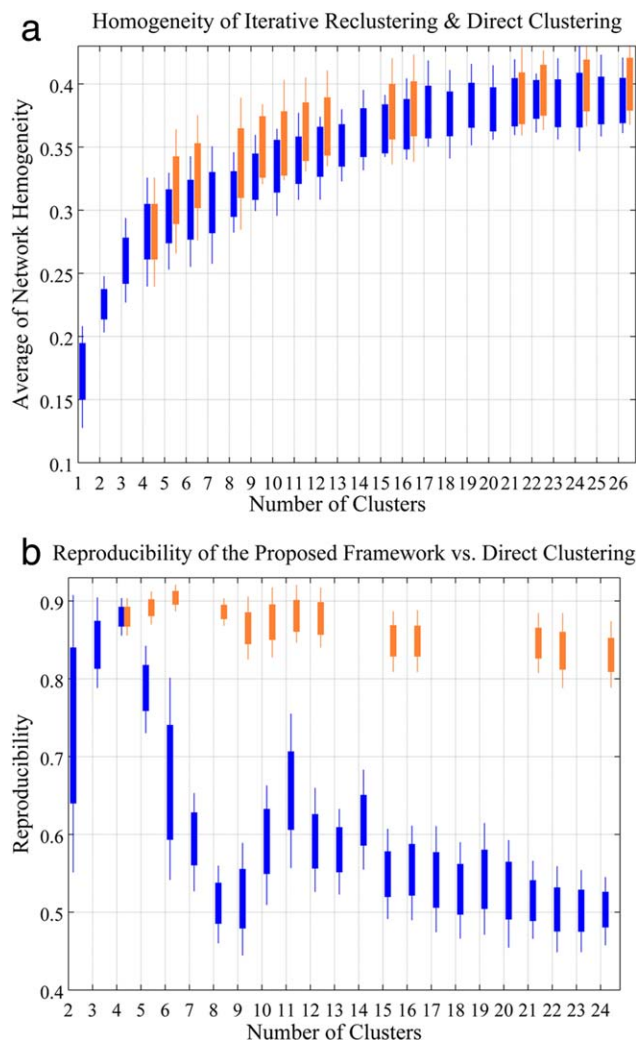
Figure 5.
The four large clusters obtained in the first iteration of the clustering algorithm with the ER similarity measure. Default mode, visual, somatosensory networks are represented in red, green, blue, respectively. The subcortial structures form a network, which is represented in beige. The color intensity of a voxel represents the cross correlation between that voxel's time series and the average time series over all voxel in the network. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

gray matter regions before applying the proposed clustering method. As Figure 6a shows, the overall network homogeneity for direct clustering of the results from the first iteration in Figure 4 (blue distributions) gradually increase while approaching an asymptotic value of 0.38, without obvious fluctuation in the range of 21–26 clusters. Subsequently, the network's regions are iteratively subdivided into k_{OPT} subnetworks that maximize subnetwork reproducibility until the homogeneity does not significantly change after a new iteration of the algorithm (H_{m+1} − H_m)/H_m <0.01 (orange distributions). In Figure 6a, the within-network temporal homogeneity of the iterative reclustering (orange) is always higher than that of regular clustering (blue) ($P<0.001$).

With iterative reclustering, the number of network clusters jumps discontinuously depending on the value of k_{OPT} for each iteration. In contrast, with regular clustering, the brain is partitioned into k networks without iteratively estimating subnetworks, so the number of network clusters increases continuously by 1. Figure 6b shows that for the proposed iterative subnetworks (orange distributions), spatial network reproducibility stays relatively high (>0.8) while it quickly drops for regular clustering to approximately 0.5 (blue distributions). Overall, while Figure 6a shows that iteratively reclustering subnetworks has a small

impact on network homogeneity, Figure 6b shows that it has a very large impact on network reproducibility. We should emphasis that our estimation of subnetworks' reproducibility at each iteration ignores the fact that their originating network is not perfectly reproducible. Therefore, the actual reproducibility of the subnetworks is

Figure 6.
(**a**) Within cluster homogeneity and (**b**) clustering reproducibility versus number of networks (i.e., clusters; k) obtained by the direct clustering approach (blue), and the proposed iterative reclustering framework (orange). The direct clustering approach clusters the initial 680 brain regions into k clusters. The iterative reclustering framework clusters the regions within a cluster that acquired using the previous iteration of the algorithm. In each iteration, the number of subnetworks is determined based on the reproducibility metric, and it may increase by more than 1. Therefore, the proposed clustering approach did not generate the results for some of the possible number of networks, for example, 1, 2, 3, 7, 13, 14, and so forth. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]
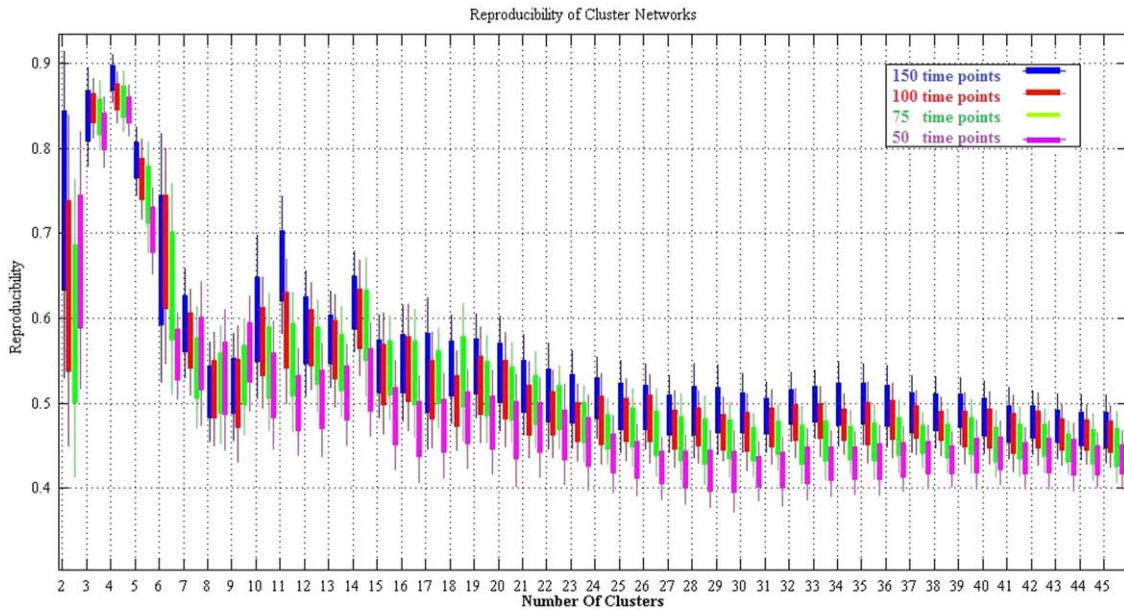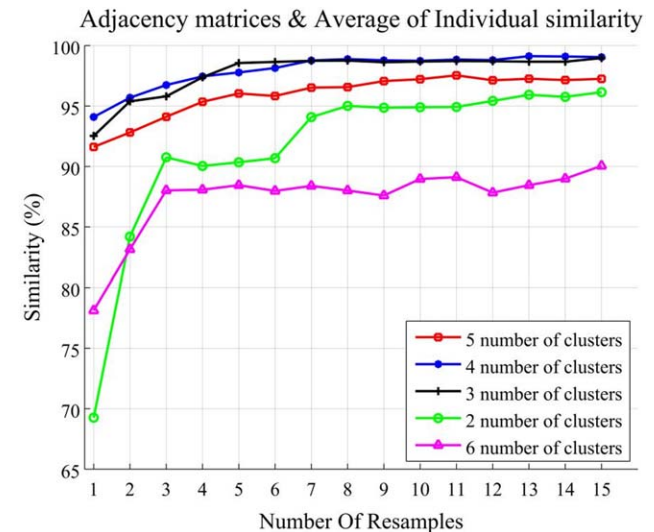
Figure 7.

The effect of time-series length on the reproducibility of the clustering algorithm, and the estimation of the optimum number of clusters. The figure shows the reproducibility of the first iteration of the clustering algorithm versus the number of clusters ($k = 1$, $k_{max} = 45$). The clustering algorithm was applied on the real fMRI dataset with ER as the similarity measure. The boxplots represent the distribution of the reproducibility values acquired over 50 split-half resamples. Different colors represent the results for the time-series length of 50, 75, 100, and 150. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

overestimated and is always less than that measured here. This is true for the subnetworks of all iterations except the first iteration, which we start with the whole gray matter as one cluster of the 680 segmented regions. This is an essential point when comparing the reproducibility of different techniques in Figure 6.

Figure 7 illustrates the reproducibility histograms of truncated time series in terms of the number of clusters. We applied the proposed framework on smaller fractions of the data (2/3, 1/2, and 1/3 of the temporal length of the data), not only to evaluate the efficiency of the proposed framework for determining the number of network clusters but also to examine consistency of reproducibility peaks from shorter time series. Note that the location of the global maximum of the reproducibility (i.e., $k_{OPT} = 4$) is not changed with the length of the time series used in the analysis from 300 to 100 s. The major effect of using truncated time series is a decrease in the overall reproducibility at each value of k. Moreover, when smaller fractions of the time series are used, the local peaks at 11 and 14 network clusters disappear but the trough of reproducibility (i.e., [7, 9]) does not significantly change.

The idea of constructing the group similarity matrix by averaging over the resampled adjacency results of clustering has been proposed in different forms in previous work [Salvador et al., 2005, van den Heuvel et al., 2008; Bellec et al., 2010]. Figure 8 shows the similarity of network

Figure 8.

The similarity values between the network clusters extracted from adjacency matrices and those extracted from average similarity matrix as a function of the number of split-half resamples. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]
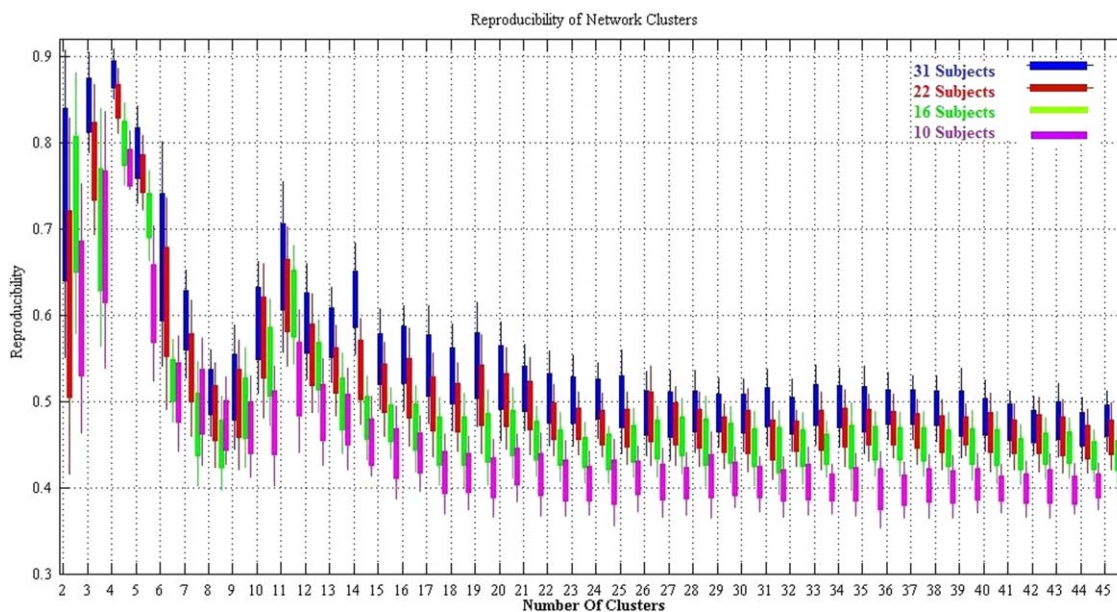
Figure 9.

The effect of group size on the reproducibility of the clustering algorithm, and the estimation of the optimum number of clusters. The figure shows the reproducibility of the first iteration of the clustering algorithm versus the number of clusters ($k = 1$, $k_{max} = 45$). The clustering algorithm was applied on the real fMRI dataset with ER as the similarity measure. The boxplots represent the distribution of the reproducibility values acquired over 50 split-half resamples. Different colors represent the results for group size of 10, 16, 22, and 31. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

results from the average of resampled adjacency matrices to the networks resulting from using the average of single subject similarity matrices. Similarity is measured as the Jaccard metric between two sets of results acquired from the two strategies. The similarity sharply increases while approaching an asymptotic value of 99% for $k_{OPT} = 4$ or 3 clusters. In contrast, the other numbers of clusters result in less similar results between the two different approaches. These results support our proposed use of k_{OPT} to extract conserved reproducible results from both average similarity and resampled adjacency matrices.

To evaluate the effect of the number of subjects involved in the analysis, we applied the proposed framework on smaller numbers of subjects (10, 16, and 22). Figure 9 illustrates the reproducibility histograms from smaller groups of subjects, as a function of the number of clusters. This figure shows that the reproducibility of the results decreases with decreasing numbers of subjects involved in the analysis, but the maximum spatial reproducibility of the resultant network clusters remains at four (i.e., $k_{OPT}=4$). Moreover, the framework shows consistent behavior in terms of a stable local peak occurring at 11 clusters in contrast to the local peak at 14 clusters, which dramatically drops off and disappears as the number of subjects decreases.

Functional brain hierarchy

The results of iterative reclustering of the four large network clusters with maximum ER are shown in Figure 10a–

d as four principal hierarchies that subcluster to a total of 24 minor network clusters, that is, (a) 3, (b) 4, (c) 5, and (d) 12. Each large network cluster at the first level is coded with a specific color. The first three main network clusters consist of the (a: RSN1) default mode networks, (b: RSN2) visual networks, and (c: RSN3) somatosensory networks, respectively. The fourth large network cluster consists of areas like hippocampus, rectus, amygdala, olfactory, cerebellum, caudate, putamen, thalamus, and a small part of supplementary-motor area.

In the first partitioning of Figure 10a, RSN1 is decomposed into the default mode network consisting of prefrontal, anterior cingulate, posterior cingulate, inferior temporal gyrus, superior parietal region [Damoiseaux et al., 2006], and RSN1-2 that includes some part of frontal-mid (L/R), cingulum-mid, frontal-sup-medial (R), and large sections of inferior frontal (pars triangularis) (L/R), super marginal (R/L), and Temporal mid (R/L). In this branch, the last reclustering iteration parcels the RSN1-2 into right and left hemisphere regions (RSN1-2-1/2).

The visual network of Figure 10b, the most homogeneous network cluster of the first 4, is decomposed into four extra homogeneous parts in one reclustering stage: occipital-superior, calcarine (R/L), and cuneus (R/L) are placed in RSN2-1, Lingual and fusiform are placed in RSN2-2, occipital-mid and occipital-inf are placed in RSN2-3, and part of precuneus (R/L) is placed in RSN2-4.
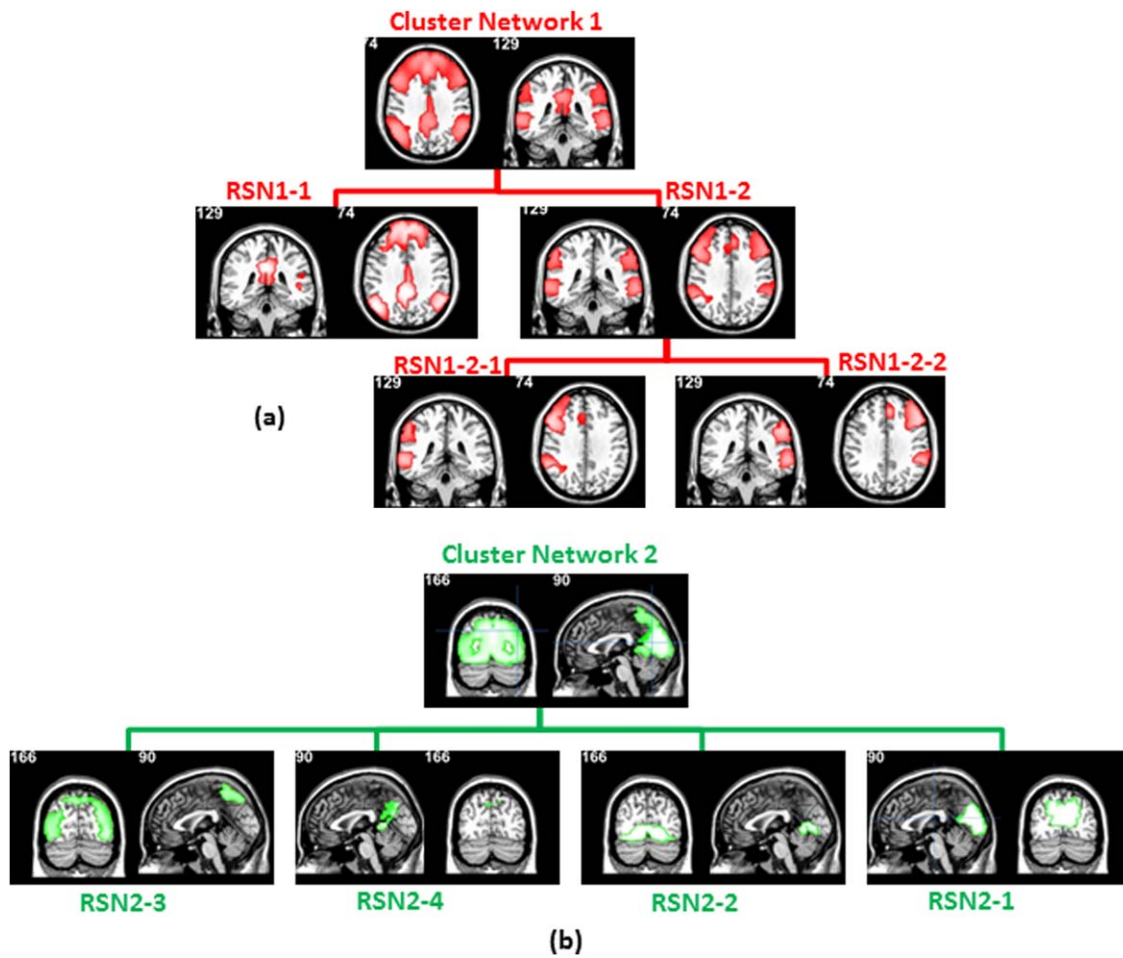
Figure 10.

The networks and their hierarchal structure extracted using the proposed clustering algorithm are represented. The first iteration of algorithm generates 4 networks including: (**a**) Default mode (RSN1), (**b**) Visual (RSN2), (**c**) Somatosentory (RSN3), and (**d**) subcortical structures (RSN4). (a): RSN1 is declared as a heterogeneous network, and in the second iteration, it is partitioned into two subnetworks (i.e., RSN1-1 and RSN1-2). RSN1-1 is declared as homogenous while RSN1-2 was declared as heterogeneous. Then, in the third iteration, RSN1-2 is partitioned into two subnetworks (i.e., RSN1-2-1 and RSN1-2-2), which are considered as homogenous subnetworks. (b): RSN2 is declared as a heterogeneous network, and in the second iteration, it is partitioned into four subnetworks (i.e., RSN1-1, RSN1-2, RSN1-3, and RSN1-4). All four subnetworks did not pass the heterogeneity criterion, and therefore, they did not partitioned. (c): RSN3 is declared as heter-ogeneous networks. Then it is splitted into 2 subnetworks (i.e., RSN3-1 and RSN3-2). RSN3-2 is heterogeneous, and then is partitioned into 2 subnetworks (i.e., RSN3-2-1 and RSN3-2-2). RSN3-2-1 is declared as heterogeneous, and in the fourth iteration is partitioned into 3 subnetworks (i.e., RSN3-2-1-1, RSN3-2-1-2, RSN3-2-1-3, and RSN3-2-1-4). (d): RSN4 is also known as heterogeneous, and it is then partitioned into two subnetworks (i.e., RSN4-1 and RSN4-2). Both RSN4-1 and RSN4-2 are declared as heterogeneous, and in the second iteration, they partitioned into 2 and 3 subnetworks, respectively. RSN4-2-1 and RSN4-2-2 are declared as heterogeneous, and they are partitioned into 6 and 2 subnetworks in the third iteration. RSN4-2-2-1 is declared as het-erogeneous, and it is splitted into 2 subnetworks in the fourth iteration. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

For RSN3 of Figure 10c, the first subclustering results in two subnetworks: RSN3-1 consisting of supp-motor-area, postcentral, paracentral-labule, and RSN3-2 including insula, rolandic-oper, cingulum-mid, heschl, temporal-sup, and parts of precentral. In the next reclustering iteration of RSN3-2, rolandic-oper, insula, and temporal-sup are placed in RSN3-2-2 and the rest of the regions, cingulum-mid, heschl and part of precentral, constitute RSN3-2-1. This then subclusters into three small subnetworks, plac-ing precentral, postcentral, and supramarginal in different

Cluster Network 3

RSN3-2   RSN3-1

RSN3-2-1   RSN3-2-2

RSN3-2-1-3   RSN3-2-1-2   RSN3-2-1-1

(c)

Cluster Network 4

RSN4-1   RSN4-2

RSN4-1-2   RSN4-1-1   RSN4-2-1   RSN4-2-2   RSN4-2-3

RSN4-2-2-1   RSN4-2-2-2

RSN4-2-1-(1-6)
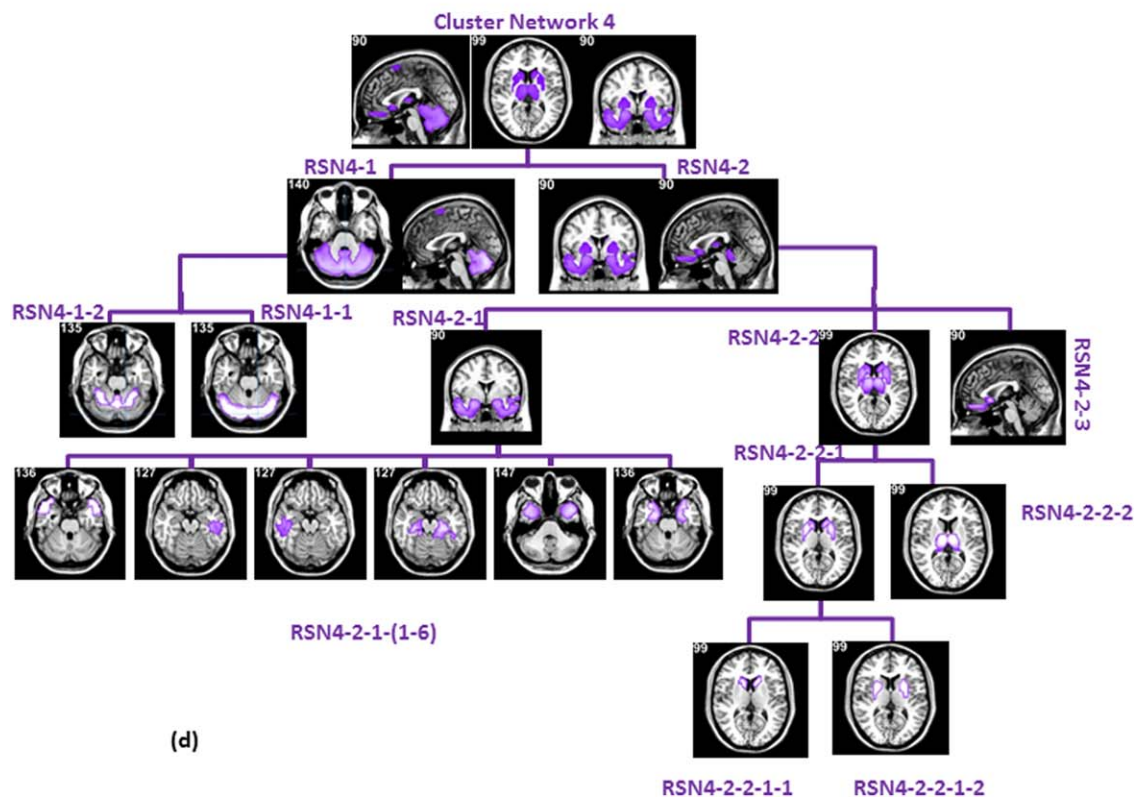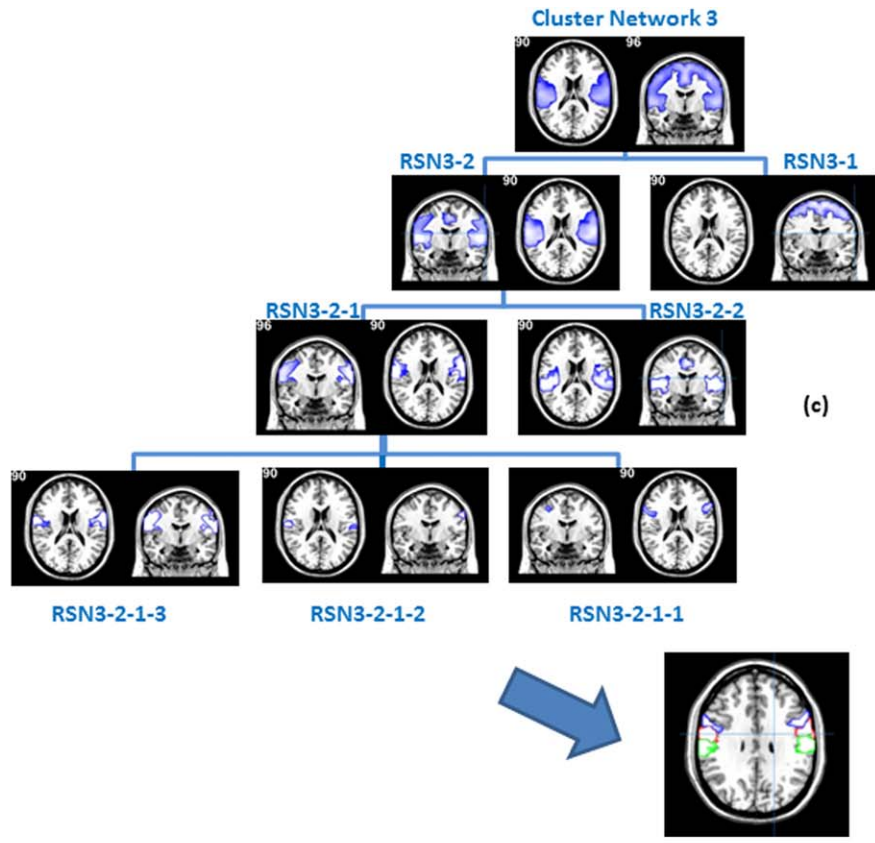
RSN4-2-2-1-1   RSN4-2-2-1-2

(d)

Figure 10.
Continued.

subnetworks as shown by the inset slice indicated by the blue arrow.

The fourth main network cluster, RSN4 of Figure 10d, has the least homogeneity, and is partitioned into two reproducible parts: the whole cerebellum along with a small part of supp-motor-area (RSN4-1) and the functional network consisting of subcortical, orbitofrontal cortex and medial temporal parts (RSN4-2). Continuing the iterative reclustering, RSN4-1 is decomposed into two sub–subnetworks consisting of cerebellum-crus (I/II) (RSN4-1-1) and cerebellum (RSN4-1-2) while a small part of supp-motor-area is linked to the cerebellum crus subnetwork. In the other branch, the algorithm finds the following three reproducible subnetworks: temporal lobe subnetwork: temporal-lobe-superior, temporal-middle, temporal pole mid, temporal inferior and hippocampus, and para-hippocampus; thalamus, putamen, and caudate; rectus, lateral superior frontal gyrus (Frontal_sup), and inferior orbit. Finally, the temporal lobe subnetworks (RSN-4-2-1) are divided into six further subnetworks satisfying the H_{Thres} constraint: hippocampus and part of amygdala, para-hippocampus, and part of fusiform, temporal-pole-sup (R/L), temporal-pole-midd (R/L), temporal-inf (L), and temporal-inf (R). In the other sub-branch, putamen and caudate (RSN4-2-2-1) separate from thalamus (RSN4-2-2-1) subnetworks and then, putamen and caudate move to two different subnetworks.

DISCUSSION

In this study, we have proposed a hierarchical framework for rs-fMRI data called iterative reclustering to automatically identify reproducible RSNs and subnetworks, which may characterize the hierarchical structure of the brain. Most of the previous studies have used clustering methods with a predefined number of clusters and have not discovered the hierarchy of the brain's functional networks. We have derived a reproducibility measure based on split-half resampling and a network homogeneity measure based on pair-wise temporal similarity, which tests the spatial reliability of the resultant networks, and optimizes the number of clusters for a particular homogeneity threshold, respectively.

One of the contributions of this study is to demonstrate the performance of an optimized reproducibility criterion to determine the number of network clusters in the data. The results of experimental data shown in Figures 4, 7, and 9 demonstrate that the reproducibility of the extracted networks fluctuate as a function of the number of clusters, subjects and length of time series per subject especially for less than 15 clusters but always has a clear peak at 4 network clusters followed by a large dip in the 7–9 cluster range.

Ignoring the reproducibility of extracted clusters may alter the results significantly. For instance, choosing a number of clusters that falls in a trough in the reproduci-

bility curve results in unstable networks across different groups of subjects. In addition, Figures 4, 7, and 9 all show a similar trough for all four of the similarity measures tested reflecting the fact that attempts to directly extract 7–10 network clusters are likely to produce relatively unstable networks.

Figure 6b shows that iterative reclustering is able to completely avoid this unstable trough, with improved temporal homogeneity, but only by hierarchically subdividing the 4 cluster networks shown in Figure 5.

Similar to subspace estimation techniques, the proposed clustering algorithm can potentially detect more detailed substructures of resting brain networks by increasing the number of iterations. Recently, some authors have suggested extracting 70 networks or more to acquire a more detailed picture of brain connectivity structures [Abou-Elseoud et al., 2010]. However, to assess between-network connectivity structure in subspace estimation techniques, for example, ICA, the correlation coefficients between the temporal responses of the estimated network have to be assessed [Jafri et al., 2008]. This may represent a problem for such approaches as some common subspace estimation techniques including ICA may not be stable in the temporal domain [Afshin-Pour et al., 2014]. In contrast, for each iteration of our framework, as each network is decomposed to its constituent subnetworks, their homogeneity and reproducibility is assessed, and the stable, hierarchical relationships between them are estimated.

We assessed the sensitivity of our approach to the choice of temporal similarity measure. As the results in Figure 4 show the ER and CCA measures produce generally similar curves and both peak at $k = 4$ clusters. Also, the WER and RV measures provide generally similar curves with the clear exception of $k = 4$, where RV produces clusters with significantly lower reproducibility. Despite such local differences, all the similarity measures produce the same general features of the reproducibility measure curve versus the number of clusters. For example, the local peaks seen in Figure 4 for k between 10 and 15 are in reasonable agreement with other resting-state studies in which 10–14 large-scale RSNs are detected from fMRI data of 10–30 subjects [Beckmann et al., 2005; Damoiseaux et al., 2006; De Luca et al., 2006; Laird et al., 2011; Afshin-Pour et al., 2014]. In addition, this finding is in agreement with related studies using the BrainMap database [Toro et al., 2008; Smith et al., 2009].

Using $N = 31$ subjects, these analysis techniques reveal only general large-scale networks. Studies with larger groups of subjects indicate that estimation of functional networks may require more than subspaces of 10–14 detectable components. For example, Kiviniemi et al. [2009] processed the resting-state fMRI dataset of $N = 55$ subjects, and detected 42 reliable independent components. Also, Abou-Elseoud et al. [2010] argue that the number of components ≤20 provides a general picture of large-scale brain networks, but model order 70 ± 10 offers a more detailed evaluation of RSNs in a group ICA

setting. The proposed technique generates the brain networks with the highest possible spatial resolution. This spatial resolution (i.e., the number of iterations in the proposed clustering algorithm) is bounded by the contrast to noise ratio (CNR) and heterogeneity of the group under study. Having a higher CNR helps the proposed algorithm to better identify the subnetworks within a network, and split them into the functional modules.

We have assessed the sensitivity of our approach in estimating the number of clusters in Figures 7 and 9; our clustering techniques can reliably estimate the number of clusters even with 50 time-points per subject or with 10 subjects. This robustness may be induced using a spatial criterion, that is, reproducibility, for primary estimation of the number of clusters. Having a large sample size of voxels in the spatial domain may alleviate instability which occurs from having a limited number of subjects or time-points. In fact, many fMRI resting-state techniques use spatial domain criterion to provide a robust estimation of the brain networks [e.g., Afshin-Pour et al., 2012].

Our cluster results in Figure 10 from experimental fMRI data are consistent with previous studies of the hierarchical organization of the human fMRI networks. For example, the first three large network clusters extracted in this study (RSN1-3) are fairly similar to the large modules reported previously in the Meunier et al., studies as Parieto-frontal, occipital, and central (2009, 2010).

The main advantage of our iterative reclustering framework over other methods is that it allows us to quantitatively assess the number of clusters to extract at any iterative stage while maximizing spatial reproducibility and temporal homogeneity. Then using these metrics we can investigate as large a hierarchy of networks and subnetworks as we wish to compute while preserving their reproducibility and temporal homogeneity.

However, our iterative reclustering framework suffers from several interacting limitations: the number of clusters in each scale is determined by split-half resampling, which can be a time consuming task without high performance computing resources, and the initial anatomical segmentation may affect the results. To overcome the latter limitation, our proposed framework could be applied on a voxel-wise rather than a region-wise basis provided sufficient computing resources are available.

Two sources of bias may exist in the results provided in this work: (1) as we are determining the number of clusters using the maximization of reproducibly, the estimated reproducibility for the proposed technique may be an overestimation of the actual reproducibility of the technique. Ideally, another dataset, that is, validation dataset, has to be used to measure the reproducibility of the proposed technique with the estimated optimum number of clusters from the original dataset. This will provide an unbiased measure of the reproducibility metric. (2) In the results in Figure 9, the effect of using different numbers of subjects on the reproducibility is provided. However, the subject subsamples are not independent and enforce a smooth transition toward the results of the larger sample, which may not reproduce well in a larger, independent resting-state fMRI dataset.

CONCLUSION

We developed an automatic framework called iterative reclustering to identify the RSNs in rs-fMRI data. Our framework contains three major contributions. First, to build a similarity matrix, two new similarity measures between pairs of regions are derived based on PCA and a subspace ER. Secondly, to determine the number of RSN clusters in the brain, a reproducibility measure is suggested, which is based on split-half resampling and the Jaccard similarity coefficient. Finally, to achieve reliable RSNs, an iterative reclustering approach is performed to enforce spatiotemporal homogeneity of the resulting networks, and maximize their reproducibility across subjects. Our iterative reclustering framework reveals a hierarchical organization of the brain, and defines potential network clusters that are both spatially reproducible across subjects and temporally homogeneous within each network.

We evaluated the proposed framework with real fMRI datasets against a direct clustering method. The evaluation results indicate that our method achieves consistent and substantial improvements over direct clustering in terms of reproducibility and network homogeneity.

APPENDIX
RV-COEFFICIENT

The RV coefficient, as a multivariate generalization of correlation, measures the closeness of two data matrices. As Escoufier defined in [Escoufier, 1973], RV is a similarity measure between positive semidefinite matrices (e.g., covariance matrices). Also, Robert and Escoufier showed that several linear multivariate statistical analyses (e.g., CCA and PCA) could be defined by maximizing the RV-coefficient subject to specific constraints [1976].

Consider two sets of eigenvectors obtained by applying PCA to the sets of time series from two brain regions. We put the eigenvectors of one region in the rows of a data matrix X ($p \times n$) and that of the other region in the rows of another data matrix Y ($q \times n$) where n refers to number of time points and p and q refer to the number of associated eigenvectors of each region. The RV coefficient is defined as:

$$\mathrm{RV}(X, Y) = \frac{\mathrm{trace}(S_{XY}.S_{YX})}{\left(\mathrm{trace}(S_X^2).\mathrm{trace}(S_Y^2)\right)^{1/2}} \quad (8)$$

where $S_X = X^T X$ and $S_Y = Y^T Y$ are the within set covariance matrices and $S_{XY} = X^T Y$ and $S_{YX} = Y^T X = S_{XY}^T$ are the cross covariance matrices of the vectors in the X and Y data matrices, if $1/n$ is used as a normalizing factor for all of them [Robert and Escoufier, 1976]. According to the definition, the RV coefficient particularly measures the closeness between two data sets' configurations even if the data points' dimensions in the two sets are different. Assuming

that all variables are zero-mean, RV(X,Y) is independent of data rotation, translation, and global changes of scale in each data sets. Moreover, its values fall in the closed interval [0,1].

Canonical Correlation Analysis

Proposed by Hotelling [1936], canonical correlation analysis (CCA) measures the overall linear correlation between two sets of multidimensional variables. With X and Y data matrices defined as above, consider two canonical vectors w_x and w_y of size $1{\times}p$ and $1{\times}q$. Canonical correlation analysis can be defined as the problem of finding w_x and w_y such that the correlation between the linear combinations $x=w_xX$ and $y=w_yY$ is maximized, that is,

$$(w_x, w_y) = \text{argmax}\,\text{corr}(x, y)$$
$$= \text{argmax}\left(\frac{w_xX.Y^Tw_y^T}{|w_xX|\,|w_yY|} = \frac{w_xS_{XY}w_y^T}{\sqrt{w_x^TS_Xw_x}\sqrt{w_y^TS_Yw_y}}\right)\tag{9}$$

As rescaling canonical vectors has no effect on the fraction, the above optimization is equivalent to the maximization of $w_xS_{XY}w_y^T$ subject to constraints $w_x^TS_Xw_x = w_y^TS_Yw_y = 1$. It has been shown that the canonical vectors, w_x and w_y, that maximize the above expression subject to the constraints are the eigenvectors corresponding to the largest eigenvalues of the matrices $S_X^{-1}S_{XY}S_Y^{-1}S_{YX}$ and $S_Y^{-1}S_{YX}S_X^{-1}S_{XY}$, respectively [Mardia, 1979]. The largest eigenvalues of these matrices are equal to the squared canonical correlation, $\max corr(x, y)$. In this article, the sum of the eigenvalues is used as a measure of similarity between two sets of data extracted from two regions.

ENERGY RATIO AND WEIGHTED ENERGY RATIO

As above, X and Y are two sets of multidimensional variables having associated time series x_i and y_i as columns, respectively. Hence, each x_i and y_i is one of the principal components obtained for each region. For calculating the similarity between each vector in region x (i.e., x_i) and time series in region y, $\{y_1, y_2, ..., y_q\}$, we used the orthogonal projection of x_i onto the signal subspace spanned by the y_i's (i.e., Y-subspace). The orthogonal projection is given by $P_Y.x_i$, where P_Y is the projector matrix defined as

$$P_Y = Y(Y^TY)^{-1}Y^T\tag{10}$$

P_Y is an idempotent matrix and thus, the energy of the projection of x_i onto the Y-subspace and the energy of the residual can be written as $x_i^TP_Yx_i$ and $x_i^T(I-P_Y)x_i$, respectively. Similarity between x_i and Y-subspace can be defined by the ratio of the energy of the signal in the Y-subspace to the energy of the residual

$$\text{Sim}(x_i, Y) = {x_i^TP_Yx_i}\big/{x_i^T(I-P_Y)x_i}\tag{11}$$

Consequently, the similarity between data matrices X and Y can be calculated as

$$\frac{1}{p}\sum_{i=1}^p \text{Sim}(x_i, Y) + \frac{1}{q}\sum_{j=1}^q \text{Sim}(y_j, X)\tag{12}$$

In the PCA process, the eigenvalues, $[\lambda_{x1}, \lambda_{x2}, . \lambda_{xp}]$ and $[\lambda_{y1}, \lambda_{y2}, . \lambda_{yq}]$ represent the variances of the variables on each principal component axis. As these eigenvalues take into account the contribution of each eigenvector to the total variation of time series in each region, we use the eigenvalues as the weights of the ERs to define the following similarity measure

$$\text{WER}(X, Y) = \frac{1}{\Lambda_X}\sum_{i=1}^p \lambda_{xi}\text{Sim}(x_i, Y) + \frac{1}{\Lambda_Y}\sum_{j=1}^q \lambda_{yj}\text{Sim}(y_j, X)\tag{13}$$

where $\Lambda_X = \sum_{i=1}^p \lambda_{xi}$ and $\Lambda_Y = \sum_{i=1}^q \lambda_{yi}$. In the rest of the article, we call these similarity measures, the ER and WER, respectively.

REFERENCES

Abou-Elseoud A, Starck T, Remes J, Nikkinen J, Tervonen O, Kiviniemi V (2010): The effect of model order selection in group PICA. Hum. Brain Mapp 31:1207–1216.

Afshin-Pour B, Hossein-Zadeh GA, Strother SC, Soltanian-Zadeh H (2012): Enhancing reproducibility of fMRI statistical maps using generalized canonical correlation analysis in NPAIRS framework. NeuroImage 60:1970–1981.

Afshin-Pour B, Grady C, Strother S (2014): Evaluation of spatio-temporal decomposition techniques for group analysis of fMRI resting state data sets. NeuroImage 87:363−382.

Bassett DS, Greenfield DL, Meyer-Lindenberg A, Weinberger DR, Moore SW, Bullmore E (2010): Efficient physical embedding of topologically complex information processing networks in brains and computer circuits. PLoS Comput Biol 6:1000748

Beckmann CF, DeLuca M, Devlin JT, Smith SM (2005): Investigations into resting-state connectivity using independent component analysis. Philos Trans R Soc Lond B Biol Sci 360:1001–1013.

Bellec, P (2013): Mining the Hierarchy of Resting-State Brain Networks: Selection of Representative Clusters in a Multiscale Structure, International Workshop on Pattern Recognition in Neuroimaging (PRNI), Philadelphia, PA, USA, pp 54–57.

Bellec P, Rosa-Neto P, Lyttelton OC, Benali H, Evans AC (2010): Multi-level bootstrap analysis of stable clusters in resting- state fmri. Neuroimage 51:1126–1139.

Biswal B, Yetkin FZ, Haughton VM, Hyde JS (1995): Functional connectivity in the motor cortex of resting human brain using echo-planar MRI. Magn Reson Med 34:537–541.

Cai W, Ryali S, Chen T, Li CR, Menon V (2014): Dissociable roles of right inferior frontal cortex and anterior insula in inhibitory control: evidence from intrinsic and Task-related functional parcellation, connectivity, and response profile analyses across multiple datasets. J Neurosci 34:14652–14667.

Calhoun VD, Liu J, Adali T (2009): A review of group ICA for fMRI data and ICA for joint inference of imaging, genetic, and ERP data. Neuroimage 45:S163–S172.

Cao J, Worsley KJ (1999): The geometry of correlation fields, with an application to functional connectivity of the brain. Ann Appl Probab 9:1021–57.

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

Cohen AL, Fair DA, Dosenbach NU, Miezin FM, Dierker D, Van Essen DC, Schlaggar BL, Petersen SE (2008): Defining functional areas in individual human brains using resting functional connectivity MRI. Neuroimage 41:45–57.

Cole DM, Smith SM, Beckmann CF (2010): Advances and pitfalls in the analysis and interpretation of resting-state FMRI data. Front Syst Neurosci 4:18

Damoiseaux JS, Rombouts SA, Barkhof F, Scheltens P, Stam CJ, Smith SM, Beckmann CF (2006): Consistent resting-state networks across healthy subjects. Proc Natl Acad Sci USA 103: 13848–13853.

De Luca M, Beckmann CF, De Stefano N, Matthews PM, Smith SM (2006): fMRI resting state networks define distinct modes of long-distance interactions in the human brain. Neuroimage 29:1359–1367.

Escoufier Y (1973): Le traitement des variables vectorielles. Biometrics 29:751–760.

Ferrarini L, Veer IM, Baerends E, van Tol MJ, Renken RJ, van der Wee NJ, Veltman DJ, Aleman A, Zitman FG, Penninx BW, van Buchem MA, Reiber JH, Rombouts SA, Milles J (2009): Hierarchical functional modularity in the resting-state human brain. Hum. Brain Mapp 30:2220–2231.

Friston KJ, Frith CD, Liddle PF, Frackowiak RS (1993): Functional connectivity: The principal-component analysis of large (PET) data sets. J Cereb Blood Flow Metab 13:5–14.

Friston KJ, Josephs O, Rees G, Turner R (1998): Nonlinear event-related responses in fMRI. Magn Reson Med 39:41–52. doi: 10.1002/mrm.1910390109

Fox MD, Raichle ME (2007): Spontaneous fluctuations in brain activity observed with functional magnetic resonance imaging. Nat Rev Neurosci 8:700−711.

Grady CL, et al (2010): A multivariate analysis of age-related differences in default mode and task-positive networks across multiple cognitive domains. Cerebral Cortex 20:1432–1447.

Greicius M (2008): Resting-state functional connectivity in neuropsychiatric disorders. Curr Opin Neurol 24:424–430.

Greicius MD, Krasnow B, Reiss AL, Menon V (2003): Functional connectivity in the resting brain: A network analysis of the default mode hypothesis. Proc Natl Acad Sci USA 100:253–258.

Grigg O, Grady CL (2010): Task-related effects on the temporal and spatial dynamics of resting-state functional connectivity in the default network. PLoS One 5:e13311

Hossein-Zadeh GA, Ardekani BA, Soltanian-Zadeh H (2003): Activation detection in fMRI using a maximum energy ratio statistic obtained by adaptive spatial filtering, IEEE Trans Med Imaging 22:795–805.

Hotelling H (1936): Relations between two sets of variates. Biometrika 28:312–377.

Jafri MJ, Pearlson GD, Stevens M, Calhoun VD (2008): A method for functional network connectivity among spatially independent resting-state components in schizophrenia. Neuroimage 39: 1666–1681.

Kelly C, Uddin LQ, Shehzad Z, Margulies DS, Castellanos FX, Milham MP, Petrides M (2010): Broca's region: linking human brain functional connectivity data and non-human primate tracing anatomy studies. Eur J Neurosci 32:383–398.

Kelly C, Torob R, Martinoa AD, CoxaCL, Bellece P, Castellanosa FX, Milham MP (2012): A convergent functional architecture of the insula emerges across imaging modalities. NeuroImage 61: 1129–1142

Laird AR, Fox PM, Eickhoff SB, Turner JA, Ray KL, McKay DR, Glahn DC, Beckmann CF, Smith SM, Fox PT (2011): Behavioral interpretations of Keller et al. Neural Correlates of BOLD Functional Connectivity J. Neurosci., April 10, 2013, 33(15): 6333–6342, 6341 intrinsic connectivity networks. J Cogn Neurosci 23:4022–4037.

Lange N, Zeger SL (1997): Non-linear fourier time series analysis for human brain mapping by functional magnetic resonance imaging. J R Stat Soc: Ser C (Appl Stat) 46:1–29. doi: 10.1111/1467-9876.00046

Li K, Guo L, Nie J, Li G, Liu T (2009): Review of methods for functional brain connectivity detection using fMRI. Comput Med Imaging Graph 33:131–139.

Ma S, Correa NM, Li X, Eichele T, Calhoun VD, Adalı T (2011): Automatic identification of functional clusters in fMRI data using spatial dependence. IEEE Trans Biomed Eng 58:3406–3417.

Margulies DS, Böttger J, Long X, Lv Y, Kelly C, Schäfer A, Goldhahn D, Abbushi A, Milham MP, Lohmann G, Villringer A (2010): Resting developments: A review of fMRI post-processing methodologies for spontaneous brain activity. Magn Reson Mater Phys 23:289–307.

Maxim V, Sendur L, Fadili J, Suckling J, Gould R, Howard R, Bullmore E (2005): Fractional gaussian noise, functional MRI and alzheimer's disease. Neuroimage 25:141–158.

Mazziotta JC, Toga A, Evans A, Fox P, Lancaster JL, Zilles K, Woods R, Paus T, Simpson G, Pike B, Holmes C, Collins L, Thompson P, MacDonald D, Iacoboni M, Schormann T, Amunts K, Palomero-Gallagher N, Geyer S, Parsons L, Narr K, Kabani N, Le Goualher G, Boomsma D, Cannon T, Kawashima R, Mazoyer B (2001): A probabilistic atlas and reference system for the human brain: International Consortium for Brain Mapping (ICBM). Philos Trans R Soc Lond B Biol Sci 356:1293–1322.

Meinshausen N, Bühlmann P (2010): Stability selection. J R Stat Soc B 72:417–473.

Meunier D, Lambiotte R, Bullmore ET (2010): Modular and hierarchically modular organization of brain networks. Front Neurosci 4:200

Mezer A, Yovel Y, Pasternak O, Gorfine T, Assaf Y (2009): Cluster analysis of resting-state fMRI time series. Neuroimage 45:1117–1125.

Fox MD, Snyder AZ, Vincent JL, Corbetta M, Van Essen DC, Raichle ME (2005): The human brain is intrinsically organized into dynamic, anticorrelated functional networks. Proc Natl Acad Sci USA 102:9673–9678.

Robert P, Escoufier Y (1976): A unifying tool for linear multivariate statistical methods: The RV-coefficient. Appl Stat 25:257–265.

Salvador R, Suckling J, Coleman MR, Pickard JD, Menon D, Bullmore E (2005): Neurophysiological architecture of functional magnetic resonance images of human brain. Cereb Cortex 15:1332–1342.

Smith SM, Fox PT, Miller KL, Glahn DC, Fox PM, Mackay CE, Filippini N, Watkins KE, Toro R, Laird AR, Beckmann CF (2009): Correspondence of the brain's functional architecture during activation and rest. Proc Natl Acad Sci U S A, 106: 13040–13045.

Strother SC, Anderson J, Hansen LK, Kjems U, Kustra R, Sidtis J, Frutinger S, Muley S, LaConte S, Rottenberg D (2002): The quantitative evaluation of functional neuroimaging experiments: The NPAIRS data analysis framework. Neuroimage, 15: 747–771.

Strother S, La Conte S, Kai Hansen L, Anderson J, Zhang J, Pulapura S, Rottenberg D (2004): Optimizing the fMRI data-processing pipeline using prediction and reproducibility performance metrics: I. A preliminary group analysis. Neuroimage 23:S196–S207.

Schwindt GC, Chaudhary S, Crane D, Ganda A, Masellis M, Grady CL, Stefanovic B, Black SE (2013): Modulation of the default-mode network between rest and task in alzheimer's disease. Cerebral Cortex 23:1685−1694.

Shi J, Malik J (2000): Normalized cuts and image segmentation. IEEE Trans Pattern Anal Machine Intell 22:888–905.

Stone M (1974): Cross-validatory choice and assessment of statistical predictions. J R Stat Soc. Ser B (Methodological) 36:111–147.

Sun FT, Miller LM, D'Esposito M (2004): Measuring interregional functional connectivity using coherence and partial coherence analyses of fMRI data. NeuroImage 21:647−658.

Thirion B, Varoquaux G, Dohmatob E, Poline JB (2014): Which fMRI clustering gives good brain parcellations? Front Neurosci 8:167

Toro R, Fox PT, Paus T (2008): Functional coactivation map of the human brain. Cereb Cortex 18:2553–2559.

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

van den Heuvel MP, Hulshoff Pol HE (2010): Exploring the brain network: A review on resting-state fMRI functional connectivity. Eur Neuropsychopharmacol 20:519–534.

van den Heuvel M, Mandl R, Hulshoff Pol H (2008): Normalized cut group clustering of Resting-state fMRI data. PLoS One 3:e2001

Van Dijk KRA, Hedden T, Venkataraman A, Evans KC, Lazar SW, Buckner RL (2010): Intrinsic Functional Connectivity As a Tool For Human Connectomics: Theory, Properties, and Optimization. J Neurophysiol 103: 297–321.

Yan X, Kelley S, Goldberg M, Biswal BB (2011): Detecting overlapped functional clusters in resting state fMRI with connected iterative scan: A graph theory based clustering algorithm. J Neurosci Methods 199:108–118.

Yeo BTT, Krienen FM, Sepulcre J, Sabuncu MR, Lashkari D, Hollinshead M, Roffman JL, Smoller JW, Zöllei L, Polimeni JR, Fischl B, Liu H, Buckner RL (2011): The organization of the human cerebral cortex estimated by intrinsic functional connectivity. J Neurophysiol 106:1125–1165.

Zhang H, Zhang X, Sun Y, Liu J, Li W, Tian J (2010): A weighted-RV method to detect fine-scale functional connectivity during resting state. NeuroImage 54:2885–2898.