Medical Image Analysis (2020)

Contents lists available at ScienceDirect

Medical Image Analysis

journal homepage: www.elsevier.com/locate/media

Improving spectral clustering method for accurate detection of brain resting-state networks

Jason Barrett[a], Haomiao Meng[b], Zongpai Zhang[a], Song Chen[a], Li Zhao[c], David C. Alsop[d], Xingye Qiao[b], Weiying Dai[a,*]

[a]*Computer Science, State University of New York at Binghamton, Binghamton, NY 13902 USA*
[b]*Mathematical Sciences, State University of New York at Binghamton, Binghamton, NY 13902 USA*
[c]*Diagnostic Imaging and Radiology, Children's National Medical Center, Washington, DC 20010 USA*
[d]*Radiology, Beth Israel Deaconess Medical Center and Harvard Medical School, Boston, MA 02215 USA*

ARTICLE INFO

Article history:
Received 1 May 2013
Received in final form 10 May 2013
Accepted 13 May 2013
Available online 15 May 2013

Communicated by S. Sarkar

ABSTRACT

This paper proposes a data-driven analysis method to accurately partition large-scale resting-state functional brain networks from fMRI data. The method is based on a spectral clustering algorithm and combines eigenvector direction selection with Pearson correlation clustering in the spectral space. The method is robust at different noise levels and is capable of maintaining the true locations of active brain networks, even at the noise level of real fMRI data.

Keywords: spectral clustering method, brain resting-state networks, functional magnetic resonance imaging (fMRI)

1. Introduction

The brain, even during rest, exhibits spontaneous neural activity. This intrinsic activity manifests as temporally correlated signal fluctuations across different brain regions, frequently referred to as resting-state brain networks, in functional magnetic resonance imaging (fMRI) and neuroscience studies. The coherent neural activity in brain networks reflects the brain's functional organization.

Seed-based correlation analysis (SBA) is the most frequently used approach for extracting the brain networks (Biswal et al., 1995). It detects these brain networks by identifying the voxels whose time courses are correlated with the time course of a selected seed voxel or seed region of interest (ROI). However, SBA requires prior knowledge of the brain functional anatomy or task fMRI methods to select the location for a reliable seed because the location of a seed ROI affects the quality of detected brain networks.

Independent component analysis (ICA) is a data-driven approach for separating the brain networks (Beckmann and Smith, 2004). It decomposes the fMRI image time series into non-Gaussian and statistically independent components (ICs), which are either sources of interest (brain networks) or artifactual noise. Due to the lack of criteria for prioritizing the components, it is difficult to determine which ICs correspond to true brain networks.

Another data-driven approach that can potentially prioritize the fluctuating signals of brain functional images is the spectral clustering algorithm (SCA) (Luxburg, 2007). SCA was originally used to identify groups of "similar behavior". It has recently been applied to isolate brain networks (including default mode network, visual network, motor network, and dorsal attention network) and has demonstrated very promising results mit. SCA has also been combined with spatial constraints to partition brain into homogeneous functional ROIs (Craddock et al., 2011). However, it is not clear how the identified brain networks or functional ROIs correspond to the true brain network locations, especially when a large number of partitions is needed to partition the fluctuations from fMRI images.

Here, we propose an SCA-based algorithm that is capable of robustly identifying large-scale brain networks at different levels of noise. Compared to the classical method that implements

*Corresponding author: Tel.: +1-607-777-4859; fax: +1-607-777-4729; email:wdai@binghamton.edu

SCA with Euclidean distance as a k-means clustering metric in the spectral space (SCA-ED), the proposed method uses Pearson correlation as the k-means clustering metric (SCA-PC). In the spectral space, we enhance the performance of brain network identification by selecting the directions of eigenvectors to avoid the randomness of their signs. The paper is laid out as follows. First, we review the theory of the SCA method, the Nyström method, a computationally efficient approximation to the SCA method for a large number of brain voxels, and the normalized SCA method for improved separation of brain networks. Second, we demonstrate the theoretical benefit of SCA-PC over SCA-ED. Third, we introduce the eigenvector direction selection method for improving the performance of SCA-PC. Fourth, we use simulations to show that SCA-PC is more robust in retaining the true locations of brain networks than SCA-ED. Fifth, real fMRI data are used to assess the relative performance of SCA-PC and SCA-ED. The brain networks from SBA serve as references to compare with SCA-PC and SCA-ED.

2. Theory

2.1. Spectral Clustering Algorithm

We are interested in partitioning brain voxels into different groups such that brain voxels within each group are similar to each other according to their temporal correlation. A similarity graph $G = <V, E, W>$ can be used to represent the brain voxels and their temporal correlation. V is the set of nodes where each node represents a brain voxel, E is the set of edges between brain voxels v_i and v_j. W is an $n \times n$ weighted adjacency matrix, where n is the number of brain voxels and its component w_{ij} is the Pearson correlation coefficient of the signal time courses between brain voxels v_i and v_j. The brain network identification problem can be restated as finding a partition of the graph nodes into groups such that the edges between different groups have low weights and the edges within the group have high weights.

The problem of finding a partition from a similarity graph can be solved using the spectral clustering algorithm based on the spectral graph theory (Luxburg, 2007). When there exist k well separated connected components (clusters) in the graph G (no edge between the connected components), the graph Laplacian matrix has the eigenvalue 0 with multiplicity of k (Mohar, 1991, 1997). Without loss of generality, we assume that the nodes are ordered according to the connected components to which they belong. Therefore, our Laplacian matrix L is of block diagonal form and can be written as:

$$L = \begin{bmatrix} L_1 & 0 & \dots & 0 \\ 0 & L_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & L_k \end{bmatrix}$$

where each submatrix L_j is the Laplacian matrix of the j^{th} connected components. As a result, $\vec{v_j}$, defined as the j^{th} eigenvector for the eigenvalue 0, can be written as:

$$\vec{v_j} = (\vec{0}_{L_1}, ..., \vec{0}_{L_{j-1}}, \vec{e}_{L_j}, \vec{0}_{L_{j-1}}, ..., \vec{0}_{L_k})$$

where $\vec{0}_{L_i}$ is a $|L_i|$-dimensional row vector with all elements as zero, and $\vec{e}_{L_j}$ is a $|L_j|$-dimensional row vector with all elements as one. By this, the eigenvector $\vec{v_j}$ determines the nodes belonging to the j^{th} connected component, and consequently, after composing the $n \times k$ matrix with the $\vec{v_j}'s$ $(j = 1, \dots, k)$ as columns, each node in the graph can now be represented by a k-dimensional feature vector $(0, \dots, 0, 1, 0, \dots, 0)$, which contains a single element as 1 and the rest of elements as 0 and the position of 1 indicates the connected component to which this node belongs.

The similarity graph in real applications (e.g., the identification of brain networks) does not contain k well-separated clusters because each pair of two nodes has a nonzero correlation (weight). The eigenvalues of the graph Laplacian matrix for the similarity graph cannot be perfectly zero. In fact, it is even difficult to distinguish the eigenvalues close to zero from those which are not close to zero. Therefore, it remains a challenge how to determine the number of clusters in real applications, which is not the focus of the paper. Even if the number of clusters k is given, the k-dimensional feature vectors for the nodes are no longer ideal vectors like $(0, \dots, 0, 1, 0, \dots, 0)$ although one can expect the feature vector to be near $(0, \dots, 0, 1, 0, \dots, 0)$. Consequently, a k-means clustering algorithm is frequently used to identify the clusters and centroids (centers of the clusters).

2.2. The Nyström Approximation

The spectral clustering method faces serious computational challenges when applied to brain networks. The weighted adjacency matrix W grows as the square of the number of nodes. Even for a low-resolution fMRI image with a voxel resolution of 4mm × 4mm × 4mm, which has over 32,000 brain voxels, it is infeasible to fit such a large matrix W in memory and compute its eigenvector decomposition. The Lanzos method (Lanczos, 1950) and Nyström method (Nyström, 1930; Fowlkes et al., 2004) are two methods for approximating the eigenvalue decomposition for a large matrix. The Lanzos method uses the tridiagonal approximation of the matrix and then calculates the eigenvalues, while the Nyström method samples a random portion of the data to approximate the eigenvalues of the matrix. The Nyström method has been applied in earlier work for the identification of brain networks (Venkataraman et al., 2009). It is fast and achieves accurate eigenvalue decomposition and therefore is adopted in the paper.

2.3. Normalized Spectral Clustering

Several spectral clustering methods have been proposed, including unnormalized versions and normalized versions Ng et al. (2002); Luxburg (2007); Fowlkes et al. (2004). The main difference is that they use different graph Laplacian matrices. We adopt the normalized spectral clustering algorithm from Ng et al. (2002) because the normalized Laplacian matrix can minimize the dissimilarity between clusters as well as maximize the similarity within clusters, while unnormalized Laplacian matrix only implements the first objective (Luxburg, 2007). We use the row normalized Laplacian matrix, which were formally justified in their perturbation theory section (Ng et al., 2002).

In summary, for the similarity graph of real applications, the k-dimensional feature vectors contain a large value (corresponding to the member cluster) and smaller nonzero values. Using matrix perturbation, the row normalization proposed by Ng et al. (2002) attempts to correct for the non-zero values in the row vectors that do not correspond to the optimal cluster. By normalizing the rows of the eigenvector matrix, the values corresponding to nonmember clusters are sufficiently small, allowing us to better separate the eigenvectors with large elements from those close to zeros. This serves as an approximation of k-dimensional feature space in k well-separated connected components.

With row normalization supported by matrix perturbation theory, the k-dimensional feature vectors (points) cluster around k-dimensional orthonormal basis vectors (centroids). However, it is possible for points to be equidistant from several centroids in real-life noisy data. The findings of Luxburg (2007) acknowledge that such points may be considered as outliers, and we may not care about which cluster such a point is assigned to. However, since the voxels from the real brain fMRI image data are noisy because of the physiological noise and motion artifacts, we may have many of these points. When the noisy voxels lie around the border of two brain networks, assigning the voxels to either network does not affect the central locations of the two networks. However, with a large number of such points, after applying the traditional k-means, the central locations of brain networks may not be retained. Instead, we may observe a large false cluster which does not correspond to any brain network, multiple clusters which are split from one network, or one cluster which is combined from multiple separate networks. In this case, the identified locations of the clusters will be influenced by the noisy voxels and hence different from true locations of the clusters; In this paper, we aim to develop an algorithm that can retrieve the true locations of the clusters (i.e., centroids are close to the ground-truth centroids and are robust against those "noisy" nodes).

2.4. K-Means: Pearson Correlation vs Euclidean Distance

After performing the eigendecomposition and row normalization, with each node v_i now represented by k-dimensional feature vector y_i, the k-means clustering algorithm is used to identify the clusters. The most frequently used clustering metric for k-means clustering is Euclidean distance. However, when the noise distribution is unknown or a large number of nodes have quasi equidistance from the centroids, k-means clustering using Euclidean distance as clustering metric may distort the locations of the cluster centroids. Instead, k-means clustering using Pearson correlation as the clustering metric is more robust in dividing the noisy nodes towards the true locations of the clusters.

To demonstrate this, let us first consider the 3-dimensional (3D) feature space. $\vec{u} = (u_1, u_2, u_3)$ and $\vec{v} = (v_1, v_2, v_3)$ are points on a unit sphere (because all nodes have unit length after row normalization). The Euclidean distance of $\vec{u}$ and $\vec{v}$ can be represented as follows.

$$\begin{aligned} ED^2(\vec{u}, \vec{v}) &= \Sigma_{i=1}^{3}(u_i - v_i)^2 \\ &= \Sigma_{i=1}^{3}u_i^2 - 2\Sigma_{i=1}^{3}u_i v_i + \Sigma_{i=1}^{3}v_i^2 \\ &= 2 - 2\Sigma_{i=1}^{3}u_i v_i \\ &= 2 - 2 \cdot cos(\vec{u}, \vec{v}) \end{aligned} \tag{1}$$

In Equation 1, the Squared Euclidean distance between two points on a unit sphere is proportional to the cosine similarity between them, which is defined as the cosine of the angle between the two vectors. Pearson correlation between two points can also be expressed as cosine similarity.

$$\begin{aligned} corr(\vec{u}, \vec{v}) &= \frac{(\vec{u} - \bar{u}\mathbb{1})^\top(\vec{v} - \bar{v}\mathbb{1})}{\left\|(\vec{u} - \bar{u}\mathbb{1})\right\| \left\|(\vec{v} - \bar{v}\mathbb{1})\right\|} \\ &= cos(\vec{u} - \bar{u}\mathbb{1}, \vec{v} - \bar{v}\mathbb{1}) \end{aligned} \tag{2}$$

where $\bar{u}$ and $\bar{u}$ are the mean of $\vec{u}$ and $\vec{u}$, respectively, and $\mathbb{1}$ is the $n \times 1$ vector whose elements all equal 1. In this work, we propose to use the Pearson correlation $corr(\vec{u}, \vec{v})$ to replace $cos(\vec{u}, \vec{v})$ to represent the "distance" in k-mean method.

Let us show that for a vector $\vec{u}$, $\bar{u}\mathbb{1} = \vec{u_p}$, where $\vec{u_p}$ is the projection of $\vec{u}$ on the vector $(\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}})^\top$.

$$\bar{u} \cdot \mathbb{1} = \begin{pmatrix} \frac{1}{3}(u_1 + u_2 + u_3) \\ \frac{1}{3}(u_1 + u_2 + u_3) \\ \frac{1}{3}(u_1 + u_2 + u_3) \end{pmatrix} = \begin{pmatrix} \frac{1}{3}\vec{u} \cdot \mathbb{1} \\ \frac{1}{3}\vec{u} \cdot \mathbb{1} \\ \frac{1}{3}\vec{u} \cdot \mathbb{1} \end{pmatrix} \tag{3}$$

$$\begin{aligned} \frac{1}{3}\vec{u} \cdot \mathbb{1} &= \frac{1}{3}\left\|\vec{u}\right\| \cdot \|\mathbb{1}\| \cdot cos(\alpha) \\ &= \frac{1}{3}\left\|\vec{u}\right\| \cdot cos(\alpha) \cdot \|\mathbb{1}\| \\ &= \frac{1}{3}\left\|\vec{u_p}\right\| \cdot \|\mathbb{1}\| \\ &= \frac{\sqrt{3}}{3}\left\|\vec{u_p}\right\| \end{aligned} \tag{4}$$

where α is the angle between $\vec{u}$ and $\mathbb{1}$. Then we can have: $\bar{u}\mathbb{1} = \begin{pmatrix} \frac{\sqrt{3}}{3}\|\vec{u_p}\| \\ \frac{\sqrt{3}}{3}\|\vec{u_p}\| \\ \frac{\sqrt{3}}{3}\|\vec{u_p}\| \end{pmatrix}$, so $\|\bar{u}\mathbb{1}\| = \sqrt{(\frac{\sqrt{3}}{3}\left\|\vec{u_p}\right\|)^2 \cdot 3} = \left\|\vec{u_p}\right\|$. Since $\bar{u}\mathbb{1}$ and $\vec{u_p}$ are of the same length and direction, $\bar{u}\mathbb{1} = \vec{u_p}$.

Next, we know that:

$$\vec{u} - \bar{u} \cdot \mathbb{1} = \vec{u} - \vec{u_p} = \vec{u_\perp} \tag{5}$$

where $\vec{u_\perp}$ is the vertical component of $\vec{u}$.

Let us look at any semicircle lying on the unit sphere whose endpoints are $(\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}})$ and $(-\frac{1}{\sqrt{3}}, -\frac{1}{\sqrt{3}}, -\frac{1}{\sqrt{3}})$ like the one shown in Fig. 1a. For any two vectors $\vec{u}, \vec{v}$ on the semicircle, $corr(\vec{u}, \vec{v}) = cos(\vec{u_\perp}, \vec{v_\perp}) = 1$. This is because $\vec{u_\perp}$ and $\vec{v_\perp}$ are parallel. Therefore, for any two points on the semicircle, they have perfect correlation, meaning we can choose one point as a representative for all points on the semicircle.

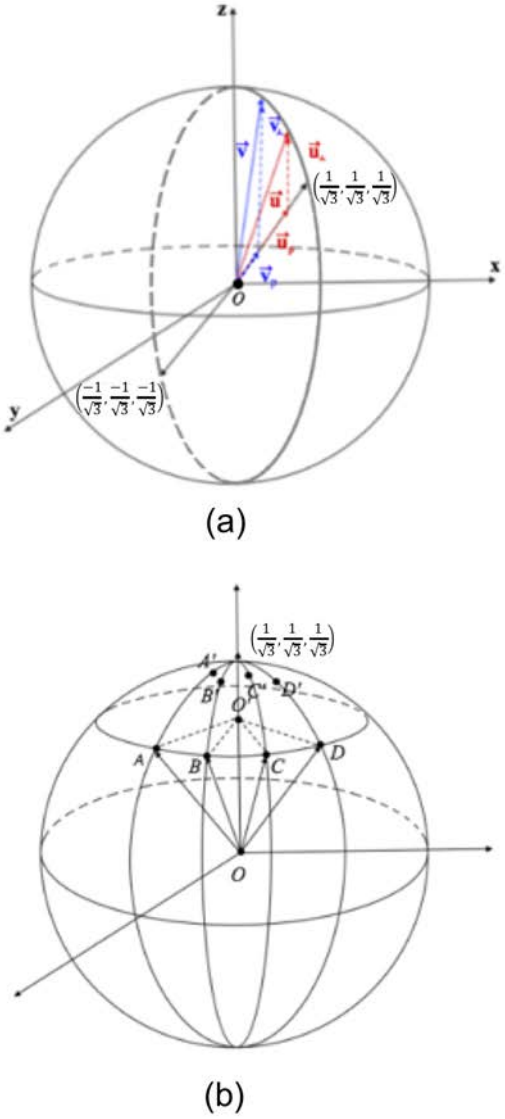


Fig. 1: Visualization of correlation on the 3D unit sphere. (a) the vectors $\vec{u}, \vec{v}$ on a semicircle that has endpoints $(\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}})$ and $(-\frac{1}{\sqrt{3}}, -\frac{1}{\sqrt{3}}, -\frac{1}{\sqrt{3}})$. $\vec{u_p}$ and $\vec{v_p}$ are the projection of $\vec{u}, \vec{v}$ onto the vector $(\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}})^\top$; the vectors $\vec{u_\perp}$ and $\vec{v_\perp}$ are the vectors corresponding to $\vec{u} - \bar{u}\mathbb{1}, \vec{v} - \bar{v}\mathbb{1}$. (b) a reorientation of our unit sphere, with the vector $(\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}})^\top$ in the direction of the z-axis and origin O. Here, four points A, B, C, D lie on 4 unique semicircles with endpoints $(\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}})$ and $(-\frac{1}{\sqrt{3}}, -\frac{1}{\sqrt{3}}, -\frac{1}{\sqrt{3}})$, with points A', B', C', D' lying on their respective semicircles and around the point $(\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}})$.

Without loss of generality, we can rotate the vector $(\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}})^\top$ to the z-axis so that all such semicircles can be better visualized (see Fig. 1b). For instance, we have four points A', B', C', D', which are very close to $(\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}})$ and from the semicircle clusters represented by A, B, C, D. ED-based k-means is to distinguish between the vectors OA', OB', OC', OD'. The four vectors are very close and therefore are difficult to separate due to the very small distances. Correlation-based k-means is to distinguish between the vectors $O'A, O'B, O'C, O'D$ where O' is the common projection on the vector $(\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}})$. In this case, it is clear that the differences between the four vectors are much larger and can be separated into the four clusters represented by A, B, C, D.

The points around $(\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}})$, mostly from artifacts, may affect the location of the chosen centroids if Euclidean-based k-means is used. $(\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}})$ is the point on the unit sphere with equal distance to the ideal centroids $(1, 0, 0), (0, 1, 0)$ and $(0, 0, 1)$, and it becomes difficult to assign nearby points to a centroid. As we show in Section IV.A, if the data is sufficiently noisy, k-means will classify such points as their own cluster since these points are relatively close on the sphere. But such a cluster is obviously not very informative since it is mostly composed of noise. However, with the correlation-based k-means algorithm, the points A', B', C', D' can be better separated into the circle passing through A, B, C, D, despite the fact that A', B', C', D' are very close in distance on the sphere. Therefore, the location of the clusters that we identify are in more close proximity to the ground truth clusters, and so the original centroids will be better preserved.

2.5. Eigenvector Direction Selection

The direction (or sign) of a eigenvector is arbitrary from the eigendecomposition algorithms: if the eigenvector v satisfies $Av = \lambda v$ for some eigenvalue λ, so does -v. However, if we use SCA-PC, the directions of these vectors influence our clustering results. Considering k-dimensional unit vectors, flipping the sign of any dimension will affect the correlation between two unit vectors. Therefore, we consider the orientation of the eigenvectors that can enhance the clustering performance.

The issue of arbitrary sign of eigenvectors has been studied before. The solution has been proposed and it showed improvement on a computer-vision segmentation problem (Bro et al., 2008). In short, the sign of an eigenvector (positive or negative) should be similar to the sign of the majority of vectors. We adopt the eigenvector direction selection idea into our algorithm. Specifically, we project the preprocessed data matrix M (each row stands for a data point) into the eigenvector direction. The signed projections (keep the sign of the projections but square their amplitudes) for each eigenvector v_i are added together. The sign is selected based on the major energy direction. The eigenvector direction selection algorithm is listed below.

For each vector v_i:

1. Compute $x_i = M \cdot v_i$
2. Calculate $s_i = \Sigma_j \, sign(x_{ij}) \cdot (x_{ij}^2)$
3. If $s_i < 0$, set $x_i = -x_i$

3. Methods

3.1. Simulations

To compare the performance of the SCA-PC and SCA-ED, we generate clusters in 3D space for clear visualization. Three clusters are centered around the three standard basis vectors (1,0,0), (0,1,0) and (0,0,1). We simulate each cluster with 1000 points on a unit sphere centered at a standard basis vector. The points on a sphere can be represented by (ρ, θ, ϕ) where ρ is

the distance from the origin, θ is the angle of the corresponding vector and z-axis, and ϕ is the angle on the x-y plane projection of the vector and the y-axis (Fig. 2). In our simulation, for the points centered around (0,0,1), ϕ is given randomly from a uniform distribution $\mathcal{U}(0, 2\pi)$ and θ from a normal distribution $\mathcal{N}(0, \sigma)$, where $\sigma = \frac{\pi}{4}, \frac{\pi}{2}, \pi$. For all points, $\rho = 1$. For the points centered around the other basis vectors, θ and ϕ are defined relative to the basis vectors as in the basic vector (0,0,1).

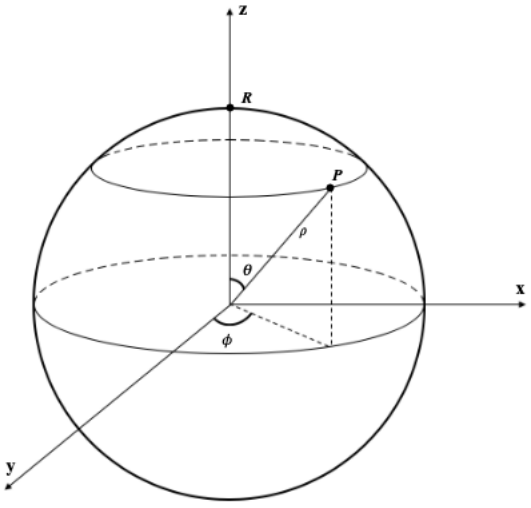


Fig. 2: Visualization of sphere population during simulation. With reference to the point R, we generate a point P on the sphere using randomly generated ϕ, θ.

To display our results from the simulation, we set up the plane tangent to the point $(\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}})^\top$ on the sphere equidistant from our three chosen standard vectors. For each point on the sphere, we project it onto this plane and compute the direction from that tangent point to the projected point. We then determine the length of that vector by computing the spherical distance from the tangent point to the point on the unit sphere, thereby yielding the location of that point on our 2-dimensional projection.

After performing the k-means clustering using the Pearson correlation and Euclidean distance metrics (k = 3), we obtain a partition of three clusters using Pearson correlation metrics P = P1, P2, P3 and using Euclidean distance metrics E = E1, E2, E3. We compare their clustering performance with the ground-truth partition G. Several popular similarity indexes between two partitions, including the Rand index (RI), the Variation of Information index (VII), the van Dongen index (VDI), the Centroid Similarity index (CSI), and Centroid-Matched Jaccard (CMJ) index (Fränti et al., 2014; Rezaei and Franti, 2016; Wu et al., 2009), are adopted for the comparisons. Normalized versions of the indexes are used to remove the similarity of two partitions which merely originates from chance (Vinh et al., 2009). The RI is the pair counting index, which is based on counting the pairs of points on which the two partitions agree or disagree. The VII is defined as the sum of conditional entropies of $H(P|G)$ and $H(G|P)$, where $H(P|G)$ stands for the number of bits needed to store the partition P given the ground-truth partition G. $H(P|G)$ is zero if P and G are the same partition because no bit is needed to store P if G is given. $H(P|G)$ is large if P and G are very different because knowing G does not provide much information for storing P. The VDI is based on matching each set in P with a set in G with the maximum intersection of nodes. The three indicies consider the average set matching between one set in P with all the sets in G. CSI and CMJ are the similarity measures based on one-on-one set matching between P and G. CSI and CMJ matches the sets by permutating the sets in P and choosing the permutation which maximizes the total number of nodes overlapping between G and P. CSI is based on the total number of nodes overlapping between G and P after the set matching. CMJ is the summed Jaccard similarity between paired sets after set matching. That is, for each pair of set G and P, the size of their intersection divided by the size of their union is calculated. CMJ is the summation of these derived values for all matched sets. CSI and CMJ reflect how well the cluster locations are preserved according to the ground-truth partition. Among the five index metrics, higher values of the RI, CSI, and CMJ indicate better clustering performance, while lower values of the VII and VDI, based on distance notation,indicate better clustering performance. 1-VII and 1-VDI are used in the clustering evaluation together with the other three measures to ensure the consistency among the metrics: the higher values, the better performance. We expect that the first three metrics will not show much difference between SCA-ED and SCA-PC. We compared the different similarity indexes to show the comprehensive perspective for how the two algorithms SCA-PC and SCA-ED are different.

3.2. Human Subjects, Scanning, and Image Analysis

We test the performance of SCA-PC and SCA-ED on a pre-existing resting-state blood oxygen level dependent (BOLD) fMRI dataset (Dai et al., 2015). Twenty healthy subjects (8 females/12 males, 22–38 years old 30.3±4.6) were imaged on a 3 Tesla HDxt (GE Healthcare, Waukesha, WI, USA) magnetic resonance scanner using an 8-channel head coil receive array and the body coil for transmission. For resting-state BOLD fMRI, 28 axial slices with a thickness of 4 mm, a gap of 1 mm, field-of-view of 24 cm and a matrix size of 64×64 covering the whole brain were obtained using gradient-echo EPI with TR=2 s and TE=25 ms. Three hundred BOLD image volumes were collected in 10 min and 4 s. Axial T1-weighted magnetization prepared rapid gradient echo (MPRAGE) images covering the whole brain were acquired in 3 min 58 s (96 slices with Matrix: 192 × 192; thickness: 2.0 mm, TE: min Full, TR: 25 ms, TI: 1100 ms, FOV: 24 cm, rBW: 31.25 kHz).

The BOLD fMRI images are motion corrected and registered to the standard MNI space using T1-weighted MPRAGE images as intermediate images. Each subject's BOLD fMRI time series is represented as 2D time × space matrix. Global signal fluctuations are removed by regressing out the global mean

signals over the entire brain. All subjects' data are concatenated along the time dimension by stacking all global-signal-regressed 2D matrices of every single subject on top of each other. Each voxel is normalized with mean of 0 and variance of 1. The correlation matrix is then reconstructed using pair-wise correlation of all brain voxels. A normalized Laplacian matrix is formed from the correlation matrix, and singular-value decomposition is performed on the Laplacian matrix. The k=20 smallest eigenvalues are selected and the corresponding 20 eigenvectors are calculated using Nyström algorithm. We normalize the 20-dimensional feature vectors to unit length.

The directions of the eigenvectors are corrected as in Section 2.5. SCA-PC is applied to the direction-corrected eigenvector space to produce 20 clusters. SCA-ED is directly applied to the eigenvector space for comparison with SCA-PC. The direction correction is not applied for the SCA-ED method because it does not affect the performance of SCA-ED. We then rank the clusters by their average intra-cluster correlations of member voxels obtained from the correlation matrix.

To evaluate the quality of our results, we compare our clusters to those obtained by traditional seed-based analysis. We acquire the MNI coordinates of the seed voxels for the brain networks: the Default Mode (Posterior Cingulate Cortex), Medial Visual (Primary Visual Cortex), Lateral Visual (Lateral Visual System), Left and Right Lateral (Left Dorsolateral Prefrontal Cortex, Right Dorsolateral Prefrontal Cortex), Sensory Motor (Primary Motor Cortex), and Dorsal attention (Left Posterior Intraparietal Sulcus) and Auditory attention networks (Primary Auditory Cortex) from Song et al. (2016), Adriaanse et al. (2014), and Campbell et al. (2015). We then produce 2-cm^3 seed ROIs and compute voxel-wise correlation maps for each subject. The correlation maps are Z-transformed to conform with the normal distribution. The voxels with Z scores significant different from zero (voxel-level $p < 0.001$ and multiple-comparison corrected cluster-level $p < 0.05$) are reported as significant clusters.

4. Results

4.1. Simulation Results

When the noise variance is set to be relatively low, $\sigma = \frac{\pi}{4}$, simulations do not show much difference in clustering performance between SCA-PC and SCA-ED (Fig. 3a-3b). This is consistent with the quantitative comparison, showing no statistically significant difference in all five clustering metrics from 1000 simulations (Fig. 4). When $\sigma = \frac{\pi}{2}$, SCA-PC exhibits slightly better performance compared to SCA-ED (Fig. 3c-3d). A few orange dots (appearing near the centroids of blue and yellow dots in the 2D projection plot) are close to equidistance on a 3-D sphere from the centroids and hence does not matter which clusters they are divided into. Only the Normalized VDI shows statistically better performance in SCA-PC ($p = 4.964 \times 10^{-7}$), (Fig. 4). Due to the large range of the VDI across different noise levels, the clustering performance difference using normalized VDI is not clearly visible in the figure. Supplementary Fig. S1a is shown to visualize the superior clustering performance of SCA-PC compared to SCA-ED using normalized VDI.

But when the variance is set to a higher value π, the clustering methods yield different results. SCA-PC produces clusters with centroids close to the ground truth, whereas SCA-ED does not

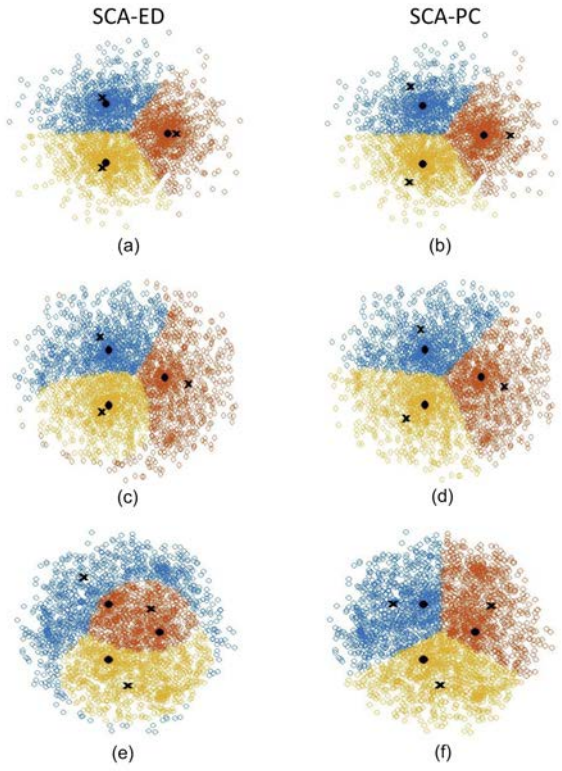


Fig. 3: Clustering results of spectral clustering algorithm with Euclidean distance (SCA-ED) and spectral clustering algorithm with Pearson correlation (SCA-PC) with different noise levels: $\sigma = \frac{\pi}{4}, \frac{\pi}{2}, \pi$. (a)-(b): clustering results from SCA-ED and SCA-PC with $\sigma = \frac{\pi}{4}$. (c)-(d): clustering results from SCA-ED and SCA-PC with $\sigma = \frac{\pi}{2}$. (e)-(f): clustering results from SCA-ED and SCA-PC with $\sigma = \pi$. Black dots and x's correspond to the ground-truth standard vectors in $\mathbb{R}_3$ and the derived centroids from either SCA-ED or SCA-PC.

produce corresponding centroids with the ground truth: a cluster (orange color) contains two of our ground-truth centroids, another cluster has no centroid (Fig. 3e-3f). All five cluster measures show a significant difference between the two methods (RI: $p = 0.00178$; VII: $p = 0.0308$; VDI: $p = 0.00930$; CSI: $p = 0.215 \times 10^{-231}$; CMJ: $p = 0.148 \times 10^{-258}$) (Fig. 4). The CSI and CMJ measures demonstrate a markedly better performance in SCA-PC, which is clearly visible in Fig. 4d-4e. The clustering performance differences using normalized RII, VII, and VDI are not clearly visible in Fig. 4a-4c. Supplementary Fig. S1b-S1d are shown to visualize the superior clustering performance of SCA-PC than SCA-ED using normalized RII, VII, and VDI. Although the clustering performance of SCA-PC is statistically better than SCA-ED using the normalized RII, VII, and VDI, there exists large overlap in their performance in the three measures. These are expected because the measures consider only the averaging matching between any two sets from different clustering methods. The markedly large improvement of SCA-PC using the CSI and CMJ indexes suggests that the SCA-PC can retain the ground-truth cluster locations even when analyzing the data with large noise.

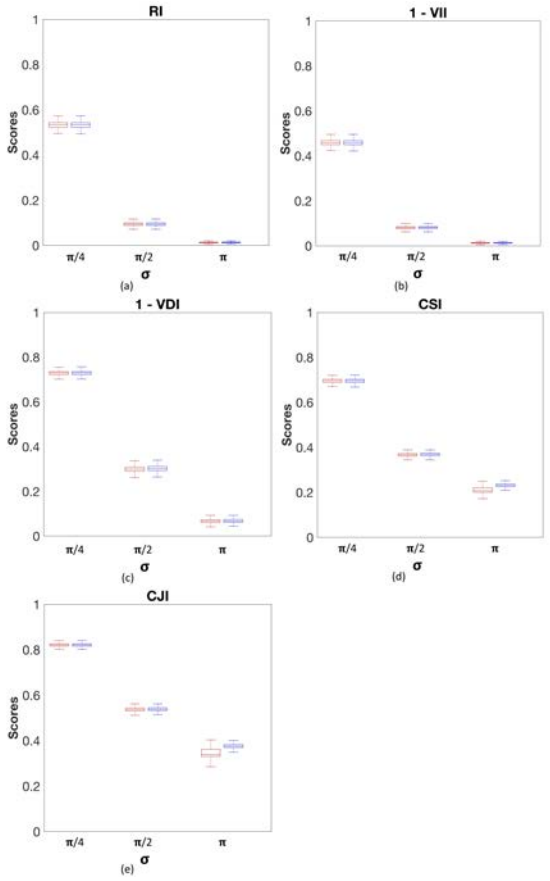


Fig. 4: Clustering performance of spectral clustering algorithm with Euclidean distance (SCA-ED) and spectral clustering algorithm with Pearson correlation (SCA-PC) on the simulation with difference noise levels: $\sigma = \frac{\pi}{4}, \frac{\pi}{2}, \pi$. Red and blue bars stand for scores of clustering results calculated from SCA-ED and SCA-PC, respectively. Clustering performance is evaluated using (a) normalized RI, (b) normalized VII, (c) normalized VDI, (d) CSI, and (e) CJI measures.

4.2. Human Subject Results

When the SBA method is applied to our in-vivo rs-fmri data, brain networks that are relatively consistent to the literature are found Fig. 5 (Damoiseaux et al., 2006; Rosazza and Minati, 2011; Zhu et al., 2013). However, the left lateral network and right lateral network appear to be more noisy than those frequently reported from the large-scale lateral networks (Damoiseaux et al., 2006; Rosazza and Minati, 2011; Zhu et al., 2013); the location of the dorsal attention network appears to be more posterior compared to its frequently reported location. Further investigation shows that the brain networks detected from SBA are sensitive to the seed locations (not shown). Therefore, the networks detected from SBA cannot be taken as ground-truth networks to be compared against the clustering performance of SCA-PC and SCA-ED.

Our proposed SCA-PC method yields brain networks that highly overlap with those frequently reported in the literature (Damoiseaux et al., 2006; Rosazza and Minati, 2011; Zhu et al., 2013), even compared with the exact anatomical locations in the corresponding brain networks (Fig. 5). This indicates that SCA-PC, as predicted by the simulation, is a robust method to maintain the true location of the cluster. In contrast, SCA-ED fails to include the anterior portion of the default mode network, whereas SCA-PC and SBA produce the network in one cluster. A further investigation shows that SCA-ED yields the anterior portion of the default mode network as a separate cluster. SCA-ED also fails to include the posterior portions of both the left lateral network and the right lateral network (Fig. 5). The missing portions of the two lateral networks have been mostly combined into the dorsal attention network using SCA-ED. SCA-ED clustering that splits one networks into two (e.g., default mode network) and combines several network portions into one cluster (e.g. posterior portions of both the left lateral network and the right lateral network, and dorsal attention network), occurs as predicted by the simulation when the images are sufficiently noisy. The appearance of this phenomenon also indicates that the real fMRI imaging data is noisy and the current state-of-art clustering algorithm may suffer from unreliable partitioning of large-scale brain networks. With SCA-PC, we are not able to completely separate the noise clusters from the brain networks using the mean intra-cluster correlation value as a rank score, although the noise clusters 14-20 tend to rank last (Fig. 6). Noise clusters 8 and 11 still mix with the brain networks. The SCA-PC is promising in separating noise clusters from clusters formed by brain networks, if the noise level of BOLD fMRI can be improved.

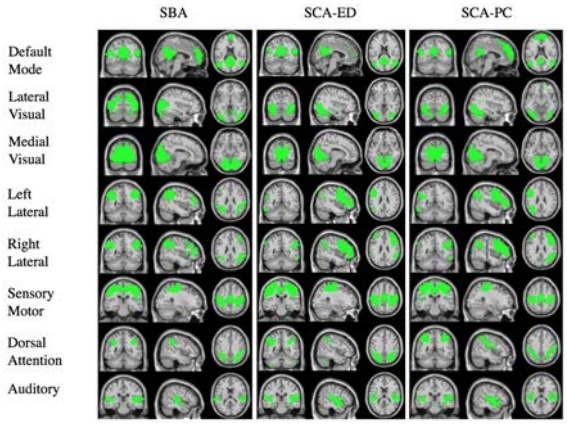


Fig. 5: Clusters derived from 20 BOLD fMRI volumes using seed-based analysis (SBA), spectral clustering algorithm with Euclidean distance (SCA-ED), and spectral clustering algorithm with Pearson correlation (SCA-PC). The Default Mode, Medial Visual, Lateral Visual, Left Lateral, Right Lateral, Sensory Motor, Dorsal Attention, and Auditory Attention networks were retrieved from our data.

5. Conclusion

We have proposed a SCA-PC method, which is based on the SCA method and using Pearson correlation as a k-means clustering metric in the spectral space. The SCA-PC method is compared with the current state-of-art SCA method, SCA-ED, using theory, simulation, and real fMRI data. We have demonstrated that SCA-PC is a robust method for the partition of brain networks using real fMRI data. It can maintain the true locations of brain networks with different levels of noise.

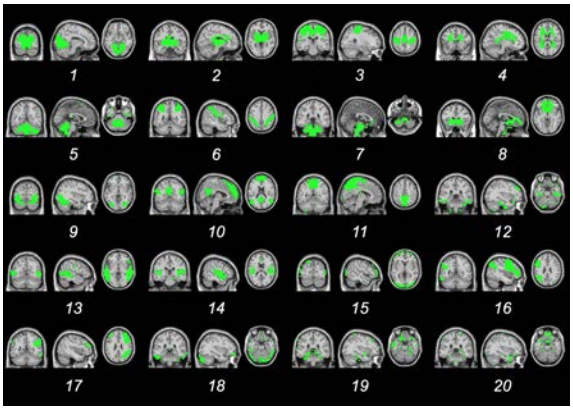


Fig. 6: Clusters derived from 20 BOLD fMRI volumes using spectral clustering algorithm with Pearson correlation (SCA-PC) ordered by mean intra-cluster correlation.

Acknowledgments

This work was supported by Transdisciplinary Areas of Excellence (TAE) seed grant at the State University of New York at Binghamton. J. Barrett was supported by the Clark Fellowship at the State University of New York at Binghamton. L. Zhao was supported by AARF-18-566- 347 from the Alzheimer's Association. D. C. Alsop was supported by R01-MH080729 from the National Institute of Mental Health (NIMH).

Appendix. Additional details on clustering performance

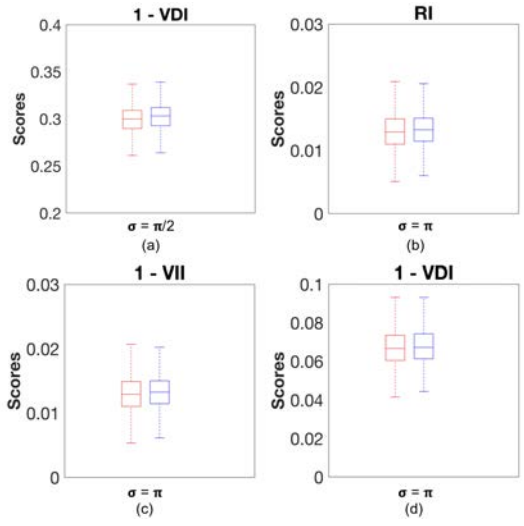


Fig. A.1: Clustering performance of spectral clustering algorithm with Euclidean distance (SCA-ED) and spectral clustering algorithm with Pearson correlation (SCA-PC) on the simulation with selected noise levels. Red and blue bars stand for scores of clustering results calculated from SCA-ED and SCA-PC, respectively. Clustering performance was evaluated using (a) 1 - normalized VDI at $\sigma = \pi/2$, (b) normalized RI at $\sigma = \pi$, (c) 1 - normalized VII at $\sigma = \pi$, (d) 1 - normalized VDI at $\sigma = \pi$.

References

Adriaanse, S., Binnewijzend, M., Ossenkoppele, R., Tijms, B., Flier, W.V.D., Wink, A.M., Scheltens, P., Berckel, B.V., Barkhof, F., 2014. Widespread disruption of functional brain organization in early-onset alzheimers disease patients. PLoS ONE e102995. doi:https://doi.org/10.1371/journal.pone.0102995.

Beckmann, C., Smith, S., 2004. Probabilistic independent component analysis for functional magnetic resonance imaging. IEEE Transactions on Medical Imaging 23, 137–152. doi:https://doi.org/10.1109/tmi.2003.822821.

Biswal, B., Yetkin, F.Z., Haughton, V.M., Hyde, J.S., 1995. Functional connectivity in the motor cortex of resting human brain using echo-planar mri. Magn Reson Med 34, 537–541.

Bro, R., Acar, E., Kolda, T.G., 2008. Resolving the sign ambiguity in the singular value decomposition. Journal of Chemometrics 22, 135–140. doi:https://doi.org/10.1002/cem.1122.

Campbell, M.C., Koller, J.M., Snyder, A.Z., Buddhala, C., Kotzbauer, P.T., Perlmutter, J.S., 2015. Csf proteins and resting-state functional connectivity in parkinson disease. Neurology 84, 2413–2421. doi:https://doi.org/10.1212/wnl.0000000000001681.

Craddock, R.C., James, G., Holtzheimer, P.E., Hu, X.P., Mayberg, H.S., 2011. A whole brain fmri atlas generated via spatially constrained spectral clustering. Human Brain Mapping 33, 1914–1928. doi:https://doi.org/10.1002/hbm.21333.

Dai, W., Varma, G., Scheidegger, R., Alsop, D.C., 2015. Quantifying fluctuations of resting state networks using arterial spin labeling perfusion mri. Journal of Cerebral Blood Flow & Metabolism 36, 463–473. doi:https://doi.org/10.1177/0271678x15615339.

Damoiseaux, J.S., Rombouts, S.A.R.B., Barkhof, F., Scheltens, P., Stam, C.J., Smith, S.M., Beckmann, C.F., 2006. Consistent resting-state networks across healthy subjects. Proceedings of the National Academy of Sciences 103, 13848–13853. doi:https://doi.org/10.1073/pnas.0601417103.

Fowlkes, C., Belongie, S., Chung, F., Malik, J., 2004. Spectral grouping using the nystrom method. IEEE Transactions on Pattern Analysis and Machine Intelligence 26, 214–225. doi:https://doi.org/10.1109/tpami.2004.1262185.

Fränti, P., Rezaei, M., Zhao, Q., 2014. Centroid index: Cluster level similarity measure. Pattern Recognition 47, 3034–3045. doi:https://doi.org/10.1016/j.patcog.2014.03.017.

Lanczos, C., 1950. An iteration method for the solution of the eigenvalue problem of linear differential and integral operators. Journal of Research of the National Bureau of Standards 45, 255. doi:https://doi.org/10.6028/jres.045.026.

Luxburg, U.V., 2007. A tutorial on spectral clustering. Statistics and Computing 17, 395–416. doi:https://doi.org/10.1007/s11222-007-9033-z.

Mohar, B., 1991. The Laplacian spectrum of graphs, in Graph theory, combinatorics, and applications. Wiley, New York. p. 871–898.

Mohar, B., 1997. Some applications of Laplace eigenvalues of graphs, in Graph Symmetry: Algebraic Methods and Applications. NATO ASI Series C 497, Kluwer. p. 225–275.

Ng, A., Jordan, M., Weiss, Y., 2002. On spectral clustering: Analysis and an algorithm. Adv. Neural Inf. Process. Syst 14.

Nyström, E.J., 1930. Über die praktische auflösung von integralgleichungen mit anwendungen auf randwertaufgaben. Acta Mathematica 54, 185–204. doi:https://doi.org/10.1007/bf02547521.

Rezaei, M., Franti, P., 2016. Set matching measures for external cluster validity. IEEE Transactions on Knowledge and Data Engineering 28, 2173–2186. doi:https://doi.org/10.1109/tkde.2016.2551240.

Rosazza, C., Minati, L., 2011. Resting-state brain networks: literature review and clinical applications. Neurological Sciences 32, 773–785. doi:https://doi.org/10.1007/s10072-011-0636-y.

Song, X., Panych, L.P., Chen, N.K., 2016. Data-driven and predefined roi-based quantification of long-term resting-state fmri reproducibility. Brain Connectivity 6, 136–151. doi:https://doi.org/10.1089/brain.2015.0349.

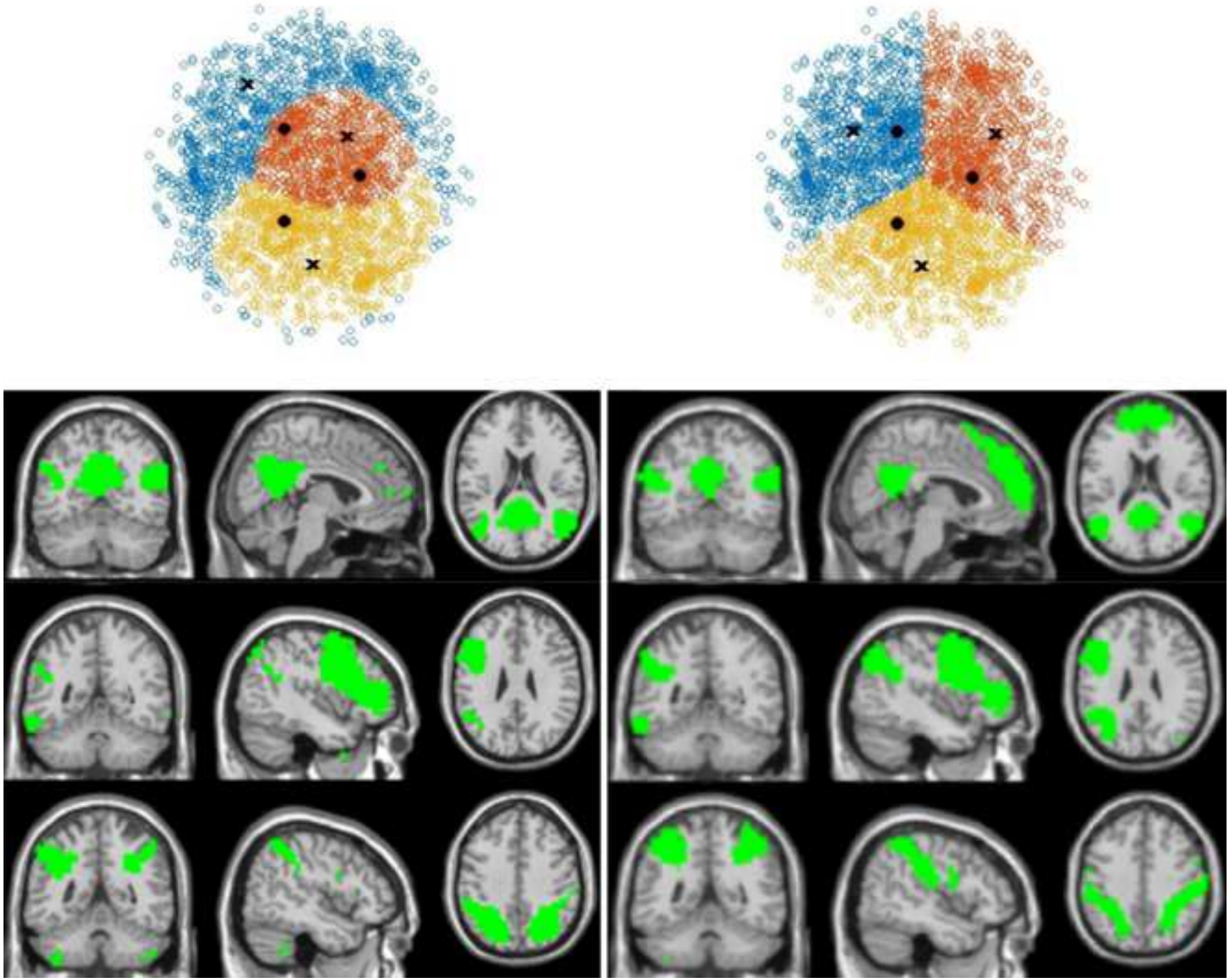
Venkataraman, A., Dijk, K.V., Buckner, R., Golland, P., 2009. Exploring functional connectivity in fmri via clustering. NeuroImage 47. doi:https://doi.org/10.1016/s1053-8119(09)70542-7.

Vinh, N.X., Epps, J., Bailey, J., 2009. Information theoretic measures for clusterings comparison. Proceedings of the 26th Annual International Conference on Machine Learning - ICML 09. doi:https://doi.org/10.1145/1553374.1553511.

Wu, J., Xiong, H., Chen, J., 2009. Adapting the right measures for k-means clustering. Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD 09. doi:https://doi.

org/10.1145/1557019.1557115.
Zhu, S., Fang, Z., Hu, S., Wang, Z., Rao, H., 2013. Resting state brain function analysis using concurrent bold in asl perfusion fmri. PLoS ONE 8, e65884. doi:https://doi.org/10.1371/journal.pone.0065884.

***Graphical Abstract**



***Highlights**

- We introduce a method for partitioning brain into different resting-state networks.
- The method is a data-driven analysis method without the need of prior knowledge of brain anatomy.
- The method shows improved accuracy compared to current methods.
- The method can retain the true locations of brain networks in real fMRI data.

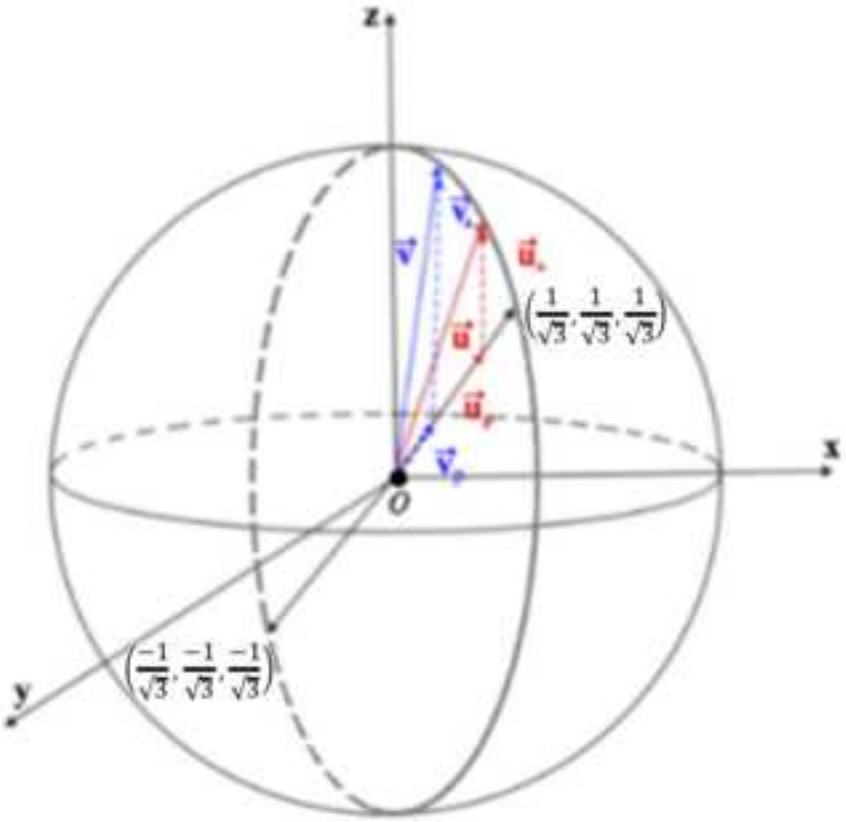
***Conflict of Interest Form**

Conflicts of Interest

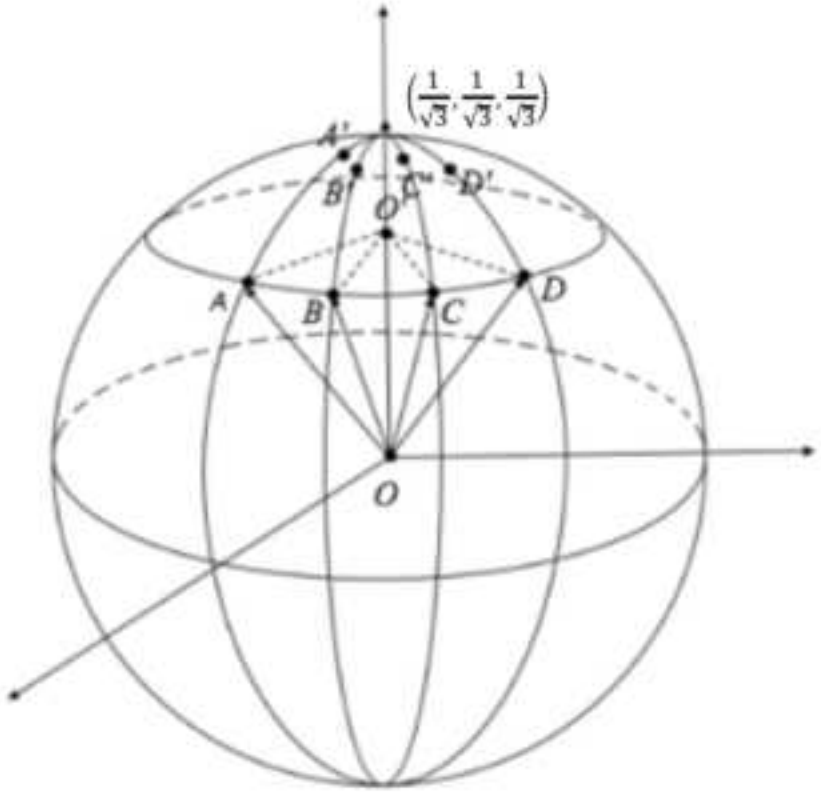
None

Figure 1.

Click here to download high resolution image



(a)



(b)

Fig. 2.

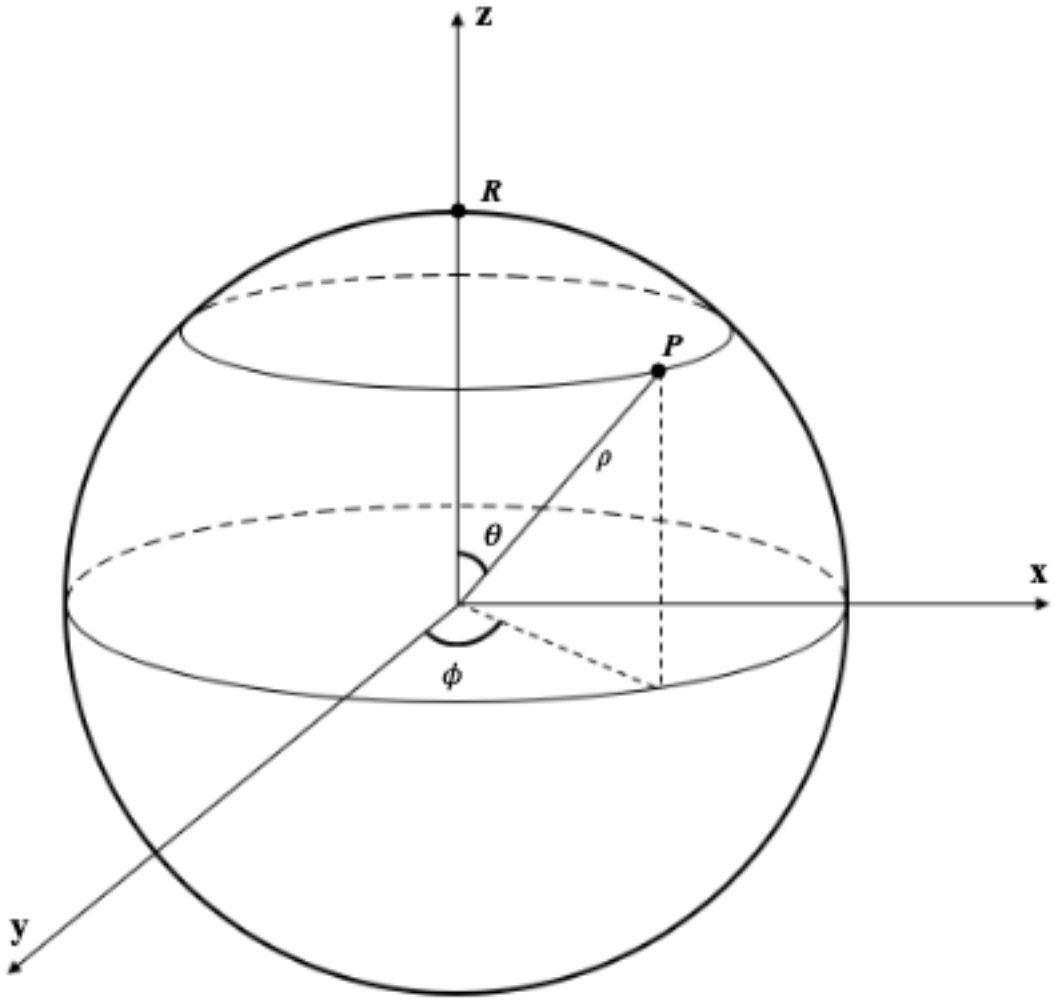


Fig. 3.

Click here to download high resolution image

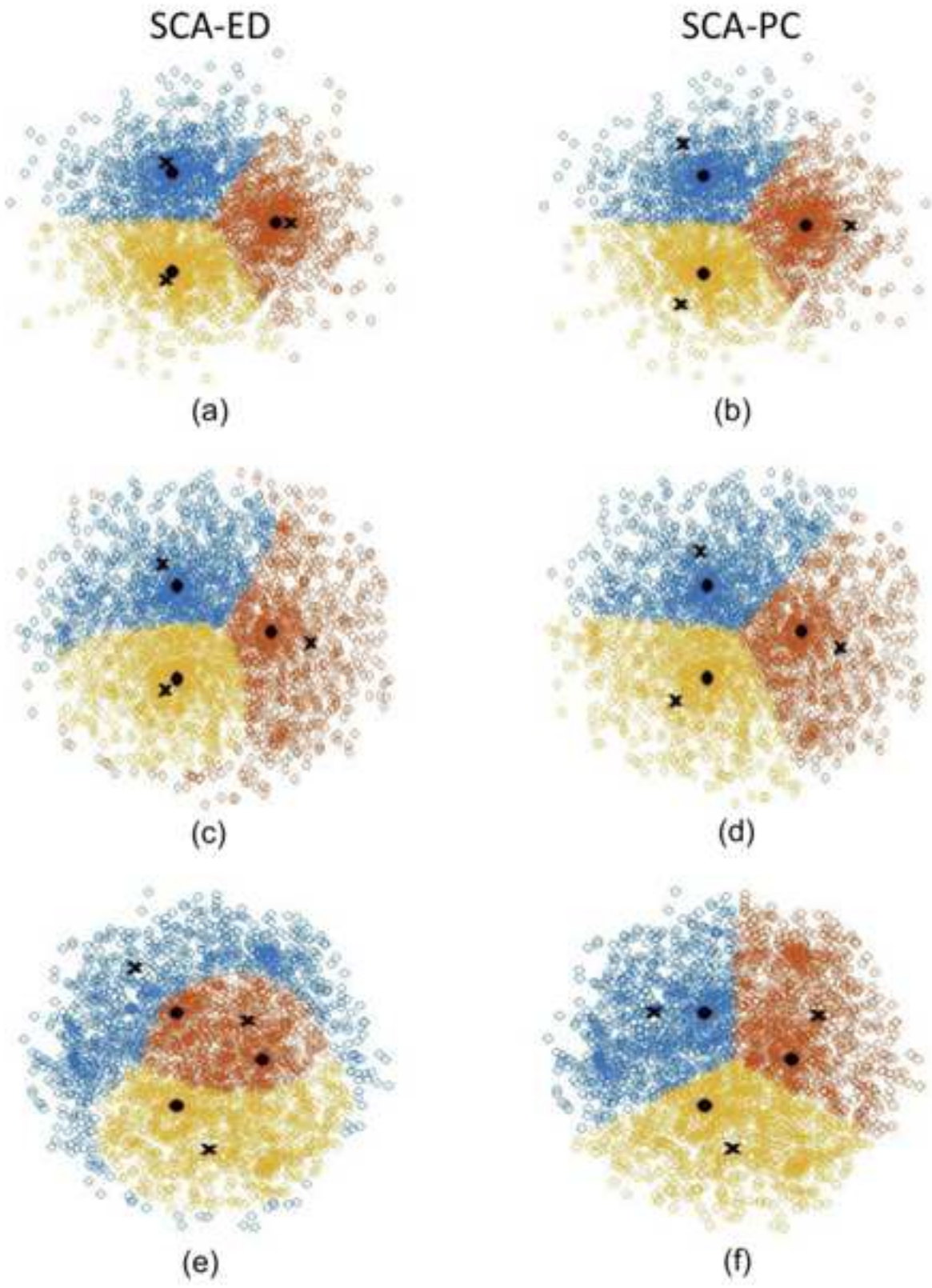


Fig. 4.

Click here to download high resolution image

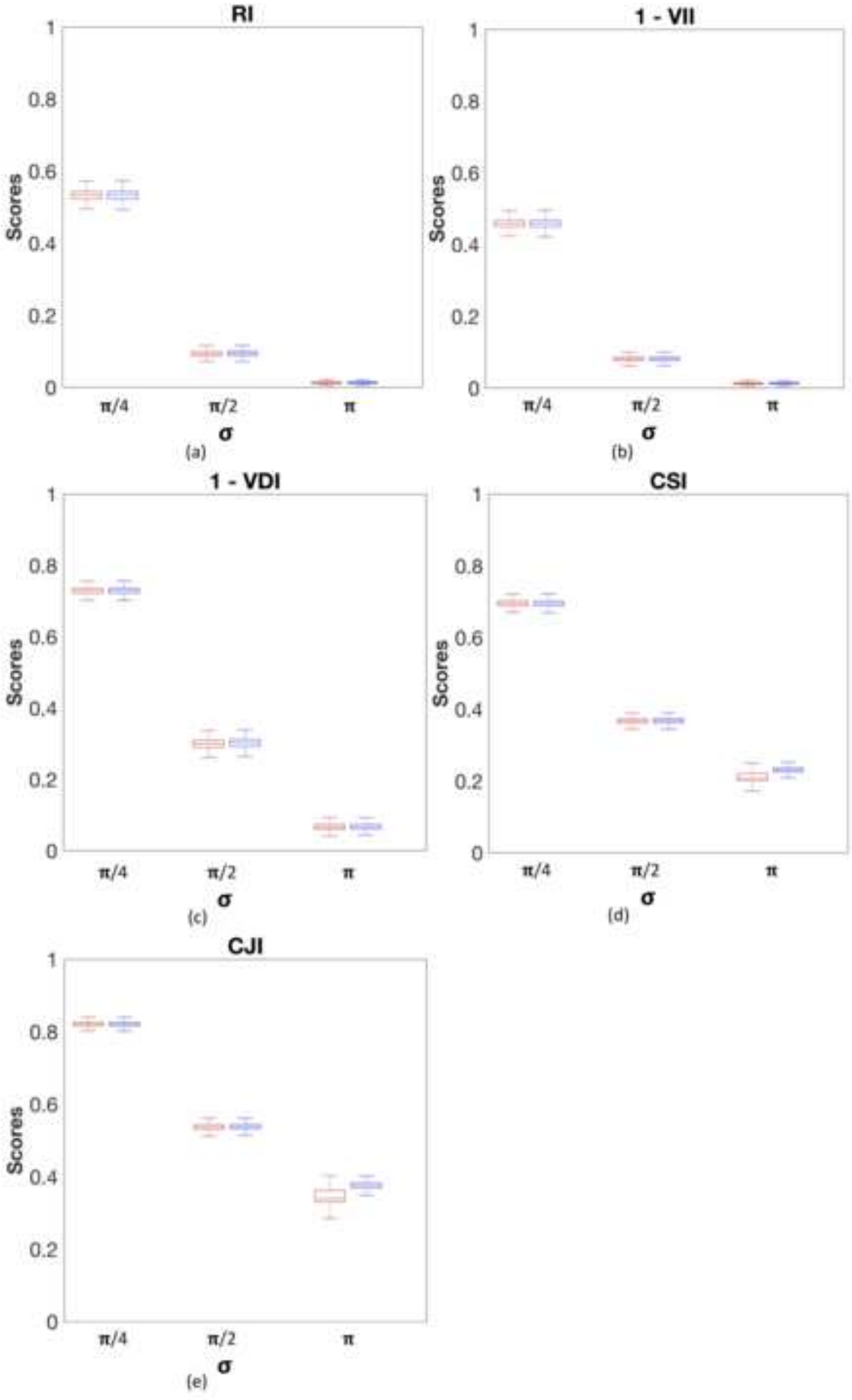


Fig. 5.

Click here to download high resolution image

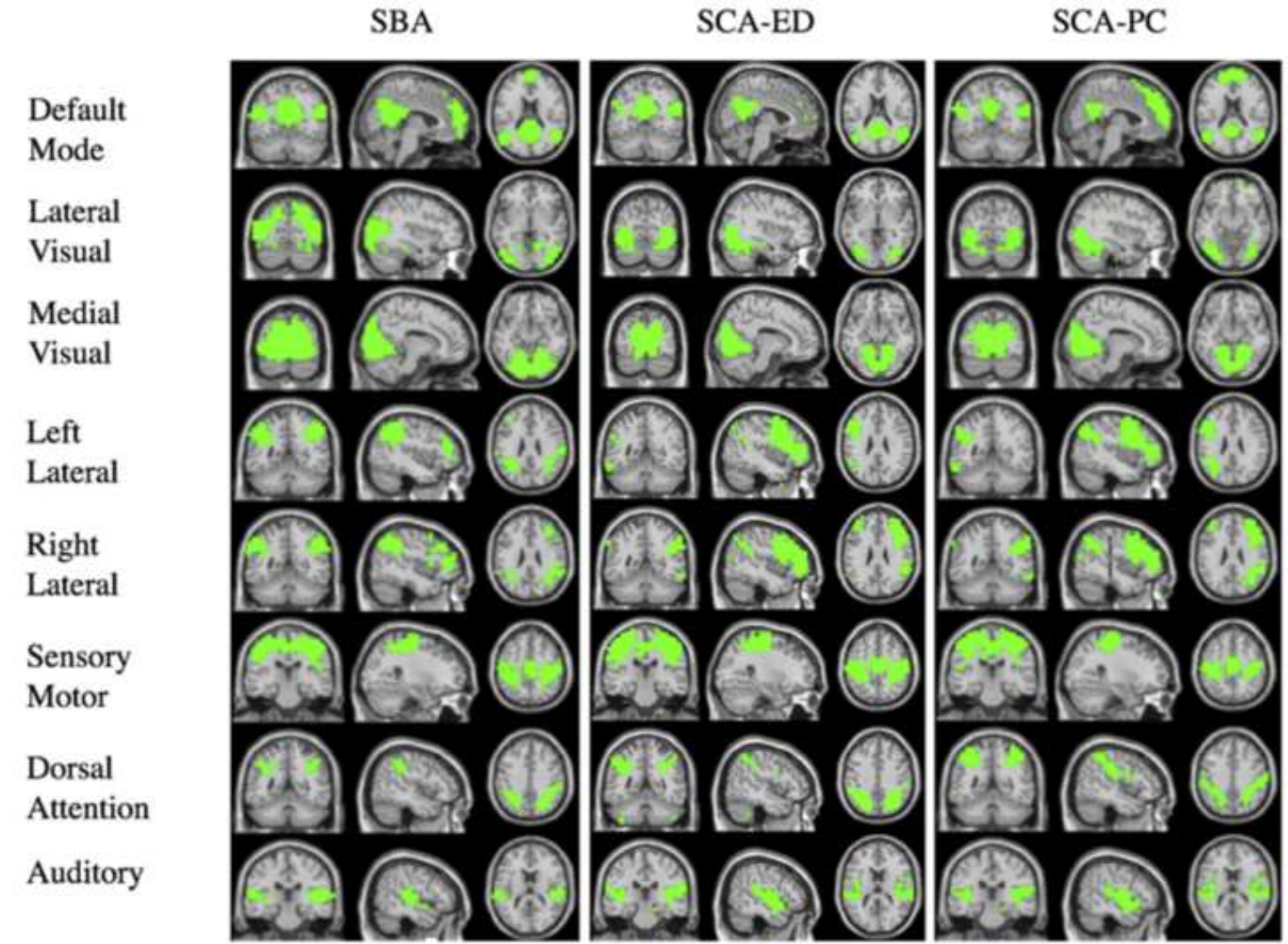


Fig. 6.

Click here to download high resolution image