Exploring functional connectivity in fMRI via clustering

EXPLORING FUNCTIONAL CONNECTIVITY IN FMRI VIA CLUSTERING

Archana Venkataraman[1], *Koene R.A. Van Dijk*[2], *Randy L. Buckner*[2], *Polina Golland*[1]

[1]MIT Computer Science and Artificial Intelligence Laboratory (CSAIL), Cambridge, MA
[2]Department of Psychology, Harvard University, Cambridge, MA
pega85@mit.edu, {*kvandijk,rbuckner*}*@wjh.harvard.edu*, *polina@csail.mit.edu*

ABSTRACT

In this paper we investigate the use of data driven clustering methods for functional connectivity analysis in fMRI. In particular, we consider the K-Means and Spectral Clustering algorithms as alternatives to the commonly used Seed-Based Analysis. To enable clustering of the entire brain volume, we use the Nyström Method to approximate the necessary spectral decompositions. We apply K-Means, Spectral Clustering and Seed-Based Analysis to resting-state fMRI data collected from 45 healthy young adults. Without placing any *a priori* constraints, both clustering methods yield partitions that are associated with brain systems previously identified via Seed-Based Analysis. Our empirical results suggest that clustering provides a valuable tool for functional connectivity analysis.

Index Terms— Magnetic Resonance Imaging, Clustering Methods, Biomedical Imaging, Brain Modeling

1. INTRODUCTION

Recent studies based on functional Magnetic Resonance Imaging (fMRI) reveal the presence of spontaneous, low-frequency (< 0.08 Hz) fluctuations in the brain. While independent of external stimuli, these signals are strongly correlated across brain structures. Functional connectivity analysis aims to detect and characterize these coherent patterns of activity as a means of identifying brain systems. Analysis is typically performed on resting-state fMRI data collected in the absence of any experimental tasks. See [1, 2] for an overview of this method.

Seed-Based correlation Analysis (SBA) [3] is the most common approach for functional connectivity analysis. It identifies the set of voxels correlated with the mean time course in a user-specified 'seed' region. SBA has been extremely useful in identifying brain systems reliably across subjects. However, SBA does have some inherent limitations. It requires *a priori* knowledge of the brain's functional organization. Furthermore, since the detected systems are dependent on the seed region locations, the additional challenge of consistent seed placement arises in group analysis. We propose clustering as a means to automatically identify candidate "seed time courses" based on the fMRI data.

Independent Component Analysis (ICA) [4, 5] is an alternative data-driven method, which has gained popularity for functional connectivity analysis. ICA isolates independent spatial sources to account for activity variation across the brain and can be used to delineate different functional networks. A limitation of using ICA is the need to select sources of interest from the often numerous spatial sources that emerge. There is currently no standard or robust solution for prioritizing the sources. Clustering provides a potentially powerful data-driven approach to solving the prioritization challenge while still allowing for data exploration with minimal *a priori* assumptions. Although this paper focuses on comparing clustering with the traditional SBA, exploring the connection between clustering and ICA is clearly an interesting direction for future work.

Here we investigate the application of clustering algorithms as a complementary approach to SBA. Clustering methods are entirely data-driven and thus do not require any prior information about the brain's spatial or functional organization. Although clustering has previously been used for fMRI analysis [6, 7, 8], we take the novel approach of applying two distinct algorithms to resting-state data collected over the entire brain volume. This allows us to partition the whole brain into an increasing number of clusters. Both K-Means and Spectral Clustering yield cluster patterns that correspond to known functional systems. This result supports the effectiveness of clustering for functional connectivity analysis.

2. K-MEANS CLUSTERING

The K-Means (KM) algorithm assumes that the time course $\mathbf{x}_i$ of voxel i is drawn from one of k multivariate Gaussian distributions with unique means $\{\mathbf{m}_j\}_{j=1}^k$ and a spherical covariance $\sigma^2\mathbf{I}$.

$$\mathbf{x}_i = \mathbf{m}_j + \mathbf{e}_i, \quad \forall i = 1, \ldots, N \tag{1}$$

where N is the number of voxels in the volume and $\{\mathbf{e}_i\}_{i=1}^N$ model $i.i.d.$ Gaussian noise.

K-Means uses the hard-assignment version of the Expectation-Maximization algorithm [9] to simultaneously determine the Gaussian means and to estimate which Gaussian is responsible for each time course. In each iteration the time courses are first assigned to the closest mean, as measured by the L^2 Euclidean distance:

$$d^2(\mathbf{x}_i, \mathbf{x}_j) = \sum_{n=1}^{T} (x_i[n] - x_j[n])^2 \tag{2}$$

where T is the length of the time courses. The mean signals of each cluster are then recomputed as the average of all time courses assigned to it. To ensure that KM properly explores the non-convex solution space, we perform multiple runs of the algorithm using different random initializations. We then select the solution that minimizes the overall sum of L^2 distances.

This naïve hard KM implementation has some attractive properties when applied to functional connectivity analysis. The algorithm

This work was supported in part by the National Alliance for Medical Image Analysis (NIH NIBIB NAMIC U54-EB005149), the Neuroimaging Analysis Center (NIH NCRR NAC P41-RR13218), the NSF CAREER Grant 0642971 and Howard Hughes Medical Institute. A. Venkataraman is supported by the National Defense Science and Engineering Graduate Fellowship (NDSEG). K. Van Dijk is supported by the Netherlands Organization for Scientific Research (NWO).

is based on minimizing the L^2 distance between the time courses and a template signal. For functional connectivity analysis, the time courses are typically normalized to have zero mean and unit variance. Therefore, $d^2(\mathbf{x}_i, \mathbf{m}_j) = 2 - 2\rho(\mathbf{x}_i, \mathbf{m}_j)$, where $\rho(\mathbf{x}_i, \mathbf{x}_j) = \sum_n x_i[n]x_j[n]$ represents the discrete-time correlation. Minimizing L^2 distance is therefore equivalent to maximizing correlation, and KM can be viewed as a natural data-driven extension of SBA.

KM can be implemented using a linear amount of memory with respect to N, the number of voxels in the brain volume. Moreover, KM typically converges in under 30 iterations, each of which terminates in linear time with respect to $k \cdot N$, where k is the number of clusters. Consequently, we have no parameters to tune, and no need for approximations as the amount of data increases.

3. SPECTRAL CLUSTERING

Spectral Clustering (SC) employs the eigen-decomposition of a pairwise affinity matrix constructed from the data points [10, 11]. The eigenvectors of this matrix induce a low-dimensional representation for the data, which is then clustered using the simple KM algorithm. SC often outperforms model-based approaches such as KM because it does not presume any parametric form for the data. Instead, it captures the natural structure of the dataset. This implies that SC can identify clusters with complex signal geometries [11].

3.1. The Spectral Clustering Algorithm

We model elements of the symmetric pairwise affinity matrix W as

$$W_{ij} = e^{-d^2(\mathbf{x}_i, \mathbf{x}_j)/2\sigma^2} \tag{3}$$

where $\mathbf{x}_i$ and $\mathbf{x}_j$ represent two voxel time courses, $d^2(\mathbf{x}_i, \mathbf{x}_j)$ is the distance defined in Equation (2), and σ^2 is the kernel width parameter. Equation (3) corresponds to the standard Gaussian kernel often used in Spectral Clustering [11]. The exponential is used to transform $d^2(\mathbf{x}_i, \mathbf{x}_j)$ into a consistent similarity measure.

Given the affinity matrix W, SC seeks a partitioning of the corresponding dataset based on a spectral decomposition of the data. In this work, we use the Normalized Cut (Ncut) variant of SC [12], which minimizes an objective based on the ratio of the sum of affinities W_{ij} between clusters to the sum of affinities within a cluster.

While the optimal solution requires a combinatorial search over all possible partitions, we can formulate a continuous relaxation of the Ncut optimization as the eigenvalue problem

$$D^{-1/2}WD^{-1/2}\mathbf{y} = \lambda\mathbf{y} \tag{4}$$

where D is a diagonal matrix such that $D_{ii} = \sum_j W_{ij}$. We define a vector of row sums $\mathbf{d}$, i.e., $d_i = D_{ii}$. The left and right multiplications by $D^{-1/2}$ in Equation (4) correspond to a symmetric normalization of W where each entry W_{ij} is divided by $\sqrt{d_i d_j}$.

The largest eigenvectors $\{\mathbf{y}_1, \ldots, \mathbf{y}_{k+1}\}$ of $D^{-1/2}WD^{-1/2}$ contain important information about the underlying geometry of the dataset $\{\mathbf{x}_i\}_{i=1}^N$. We exploit this information by clustering the rows of the matrix $Y = [D^{-1/2}\mathbf{y}_1 \ldots D^{-1/2}\mathbf{y}_{k+1}]$. Since the new representation Y tends to isolate voxel groups with small pairwise L^2 distances, the resulting clusters form the desired partition.

3.2. The Nyström Approximation

One downside of Spectral Clustering is that it relies on the eigen-decomposition of an $N \times N$ matrix, where N is the number of voxels in the whole brain. Since N is on the order of ∼200,000 voxels, it is infeasible to compute the full eigen-decomposition given realistic memory and time constraints. To solve this problem, we *approximate* the leading eigenvalues and eigenvectors of $D^{-1/2}WD^{-1/2}$ via the Nyström Method [13].

Given an N-sample dataset, we first select $N_s \ll N$ samples at random. The $N \times N$ affinity matrix W can be represented as

$$W = \begin{bmatrix} A & B \\ B^T & C \end{bmatrix} \tag{5}$$

where A is the $N_s \times N_s$ matrix of affinities between the randomly selected samples, and B is the $N_s \times (N - N_s)$ matrix of affinities between the random samples and the remaining data points. C is a large $(N - N_s) \times (N - N_s)$ matrix of remaining affinities that we want to avoid computing.

For Ncut SC we first normalize W by the matrix $D^{-1/2}$. As shown in [13], we can approximate the row sum vector $\mathbf{d}$ via

$$\hat{\mathbf{d}} = \begin{bmatrix} A\mathbf{1}_{N_s} + B\mathbf{1}_{N-N_s} \\ B^T\mathbf{1}_{N_s} + B^TA^{-1}B\mathbf{1}_{N-N_s} \end{bmatrix} \tag{6}$$

where $\mathbf{1}_M$ denotes an all-ones vector of length M. The normalized matrices $\tilde{A}$ and $\tilde{B}$ are given by

$$\tilde{A}_{ij} = \frac{A_{ij}}{\sqrt{\hat{d}_i \hat{d}_j}} \qquad \tilde{B}_{ij} = \frac{B_{ij}}{\sqrt{\hat{d}_i \hat{d}_{j+N_s}}}$$

The Nyström Method approximates the eigenvectors of $\tilde{W} = D^{-1/2}WD^{-1/2}$ using $\tilde{A}$ and $\tilde{B}$. Let $U\Lambda U^T$ denote the SVD of the $N_s \times N_s$ symmetric matrix $\tilde{A} + \tilde{A}^{-1/2}\tilde{B}\tilde{B}^T\tilde{A}^{-1/2}$. The N_s leading eigenvectors of $\tilde{W}$ are then computed as

$$V = \begin{bmatrix} \tilde{A} \\ \tilde{B}^T \end{bmatrix} \tilde{A}^{-1/2}U\Lambda^{-1/2} \tag{7}$$

Once we obtain V, the dataset is partitioned by clustering rows of the matrix $\hat{Y} = \hat{D}^{-1/2}V$.

4. MATCHING CLUSTER LABELS

Clustering algorithms arbitrarily assign label indices in each run. However, a correspondence between the labels assigned to each voxel across runs is required in order to perform a group-wise analysis of the cluster patterns. Ideally, indices would be assigned to maximize the number of voxels consistently labeled across cluster patterns. Unfortunately, a naïve approach reduces to a combinatorial search over all possible labeling combinations.

In this work, we employ a pairwise greedy algorithm to approximate the above result. Given two clusterings, we first pick one of them to be our "template" image. We then compute a $k \times k$ histogram matrix H, which contains the number of corresponding voxels between labels of the template image (down the rows of H) and of the test image (across the columns of H). During each iteration of the algorithm, we (1) select the largest element of the histogram H_{ij}, (2) reassign voxels labeled j in the test image to have the cluster index i, and (3) set all values in the i^{th} row and in the j^{th} column of H to be -1. The algorithm terminates when all entries of H are negative, meaning that the label indices of the test image have been matched to those of the template.

When aligning more than two trials, we match each of the images to a selected template using the above procedure. While this approach may not yield the globally optimal alignment across runs given arbitrary data, the cluster patterns in our application are often similar enough for this method to accurately match corresponding systems across subjects and/or trials.

	Median	Maximum	% Minimum L^2 Pattern
Two Clusters	1	4	90
Three Clusters	2	5	80
Four Clusters	2	5	67
Five Clusters	2	7	70

Table 1. Statistics on the number of different KM cluster patterns across participants for 10 random initializations. The third column shows how often KM found the clustering that corresponds to the minimum L^2 distance.

5. EXPERIMENTAL SETUP

We study the performance of the algorithms on resting-state fMRI data obtained from 45 healthy young adults (mean age 21.5, 26 female) [14]. The structural (MPRAGE) and functional (EPI-BOLD) rest-state images for each participant were collected using a Siemens Vision 3T scanner. Four $2mm$ isotropic functional runs were acquired from each participant. Each scan lasted for $6m20s$ with $TR = 5s$. The first 4 time points in each run were discarded, yielding 72 time samples per run.

We performed standard preprocessing on each of the four runs using FSL [15]. This included motion correction by rigid body alignment of the volumes, slice timing correction and registration to the MNI atlas space. The data was spatially smoothed with a $6mm$ 3D Gaussian filter, temporally low-pass filtered using a $0.08Hz$ cutoff, and motion corrected via linear regression. Next, we estimated and removed contributions from the white matter, ventricle and whole brain regions (assuming a linear signal model). We masked the data to include only brain voxels and normalized the time courses to have zero mean and unit variance. Finally, we concatenated the four runs into a single time course for analysis.

The main goal of this work is to compare the performance of SC and KM to that of SBA, which provides a good baseline due to the close relationship between correlation (the core of SBA) and Euclidean L^2 distance (the basis for SC/KM). We consider SC and KM partitions for $k = 2, \ldots, 5$ clusters.

We select the SC Gaussian variance σ^2 independently for each participant based on a qualitative measure of clustering integrity. Empirically, we observe that the resulting cluster patterns are robust for values of σ^2 in the range $100 - 250$. For most participants σ^2 is set to 100 or 150. We use these values in all experiments, regardless of the number of Nyström samples N_s or the number of clusters k.

For SBA, we select five seeds, corresponding to the motor and visual cortices, the ventral anterior cingulate cortex (vACC), the posterior cingulate cortex (PCC), and the intraparietal sulcus (IPS).

6. RESULTS

We first examine the variation in KM clustering results and the robustness of the Nyström approximation for SC.

For each value of k, we perform KM ten times on each participant using different random initializations. We then compute the number of distinct clustering patterns for each participant. We define two clustering results to be equivalent if they agree on at least 97% of the voxels. Equivalent clusterings are then grouped into distinct patterns. Table 1 summarizes the statistics of these different clustering patterns. In many cases, the KM trials converge to different local minima, especially as k increases. Thus, it is important to iterate KM several times to ensure a good quality clustering.

Next, we study the robustness of Spectral Clustering to the number of random samples. In this experiment, we start with a 4,000-sample Nyström set, which is the computational limit of our machine. We then iteratively remove 500 samples and examine the effect on clustering performance. After matching the resulting clusters to those estimated with 4,000 samples, we compute the percentage of mismatched voxels between each trial and the 4,000-sample template. This procedure is repeated twice for each participant.

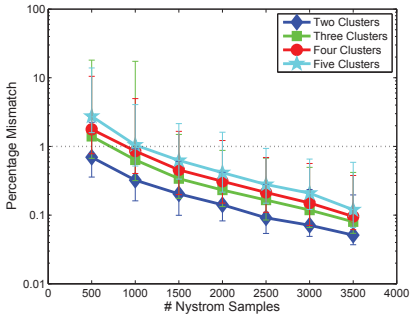

Fig. 1. Median clustering difference when varying the number of Nyström samples. Values represent the percentage of mismatched voxels w.r.t. the 4,000-sample template. Error bars delineate the $10^{th}-90^{th}$ percentile region.

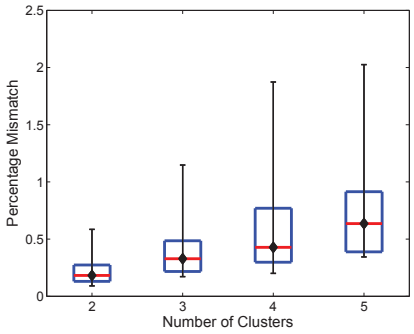

Fig. 2. Nyström clustering consistency for $N_s = 2,000$. The red lines indicate median values, the box corresponds to the upper and lower quartiles, and error bars denote the 10^{th} and 90^{th} percentiles.

Figure 1 shows the median mismatch between the estimated clusterings over all 90 (45×2) trials, with error bars corresponding to the 10^{th} and 90^{th} percentiles. The median clustering difference is less than 1% for 1,000 or more Nyström samples, and the 90^{th} percentile difference is less than 1% for 1,500 or more samples. This experiment suggests that Nyström-based SC converges to a stable clustering pattern as the number of samples increases. Based on these results, we chose to use 2,000 Nyström samples for the remainder of this work. At this sample size, less than 5% of the runs for 2,4,5 clusters and approximately 8% of the runs for 3 clusters differed by more than 5% from the 4,000-sample template.

The box plot in Figure 2 summarizes the consistency of Nyström-based Spectral Clustering across different random samplings. Here, we perform SC 10 times on each participant using 2,000 Nyström samples. We then align the cluster maps and compute the percentage of mismatched voxels between each unique pair of runs. This yields a total of 45 comparisons per participant. Figure 2 reports the statistics over all 2025 (45×45) pairwise comparisons. In all cases, the median clustering difference is less than 1%, and the 90^{th} percentile value is less than 2.1%. Empirically, we find that Nyström SC predicably converges to a second or third cluster pattern in only a handful of participants. This experiment suggests that we can obtain consistent clusterings with only 2,000 Nyström samples.

In the second series of experiments, we compare the regions

identified via Spectral Clustering, K-Means Clustering and standard SBA, as illustrated in Figure 3. We use SC and KM to partition the brain into five clusters. Only voxels assigned to the respective cluster in at least half of the participants are shown. For SBA we compute correlations over the entire brain using each seed. The third column in Figure 3 shows voxels where the correlation coefficient with the seed region was 0.2 or higher in at least 50% of the participants.

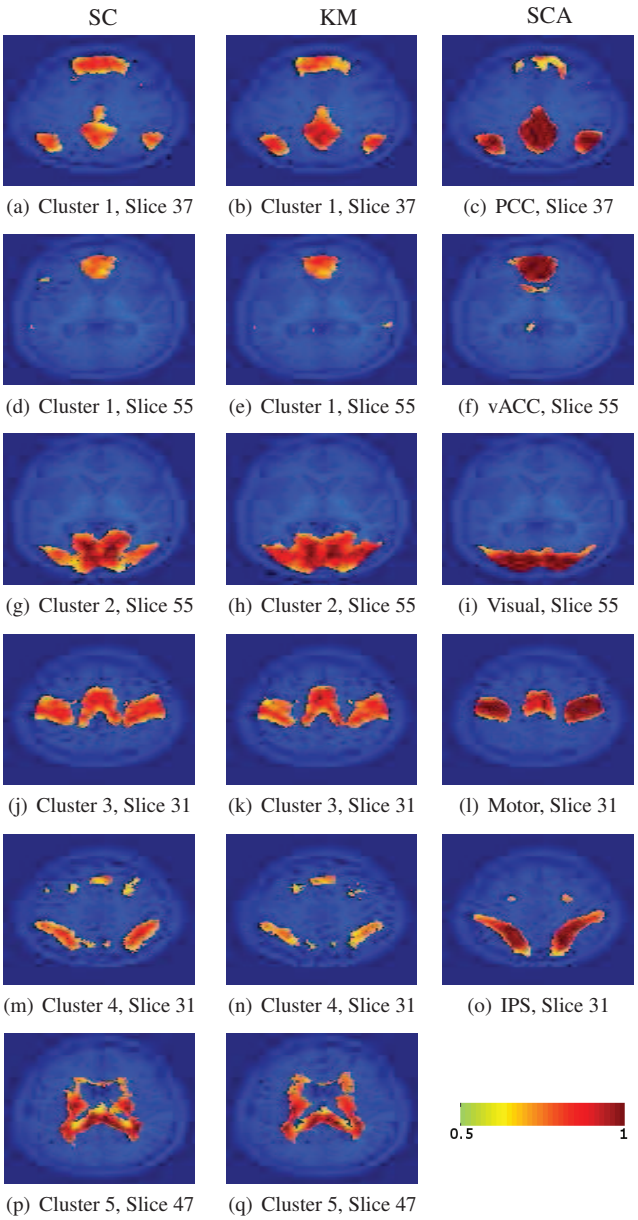

Fig. 3. Clustering results across participants. The brain is partitioned into 5 clusters using SC/KM, and various seed are selected for SCA. The color indicates the proportion of participants for whom the voxel was included in the detected system.

Figure 3 shows clearly that both SC and KM can identify well-known structures such as the default network (a-f), the visual cortex (g-i), the motor cortex (j-l), and the dorsal attention system (m-o). SC and KM also identified white matter (p-q). In general, one would not attempt to delineate this region using SBA. Since we regress out the white matter signal during the preprocessing, we would not expect SBA to isolate white matter as a system. In our experiments SC and KM achieve similar clustering results across participants. Furthermore, both methods identify the same functional systems as SBA without requiring *a priori* knowledge about the brain and without significant computation. Thus, clustering algorithms offer a viable alternative to standard functional connectivity analysis techniques.

7. DISCUSSION

The results presented in Section 6 indicate that clustering methods provide a promising approach for functional connectivity analysis. Although SC and KM make different assumptions about the data, both algorithms successfully identify well-characterized functional systems of the brain. Neither method requires *a priori* knowledge about the location of seed regions to the extent that SBA does, and both algorithms provide informative partitions with just five clusters.

The fact that K-Means produces fairly consistent clusterings across participants suggests that a simple multivariate Gaussian may be a realistic model for the pre-processed resting state fMRI data. Nonetheless, Spectral Clustering may still be useful for this application. Figures 1 and 2 show that the Nyström Method is robust while allowing for drastic decreases in memory and runtime. Furthermore, it enables exploration of alterative similarity measures.

In the future, we will use KM and SC to explore a potential hierarchical organization of the brain's functional systems. We also plan to partition the brain into a larger number of clusters, aiming to identify new and robust structures across the population. Finally, in future analysis we will compare data-driven clustering with ICA.

8. REFERENCES

[1] M. D. Fox and M. E. Raichle, "Spontaneous fluctuations in brain activity observed with functional magnetic resonance imaging," *Nature*, vol. 8, pp. 700–711, 2007.

[2] R. L. Buckner and J. L. Vincent, "Unrest at rest: Default activity and spontaneous network correlations," *NeuroImage*, vol. 37, pp. 1091–1096, 2007.

[3] B. Biswal et. al., "Functional connectivity in the motor cortex of resting human brain using echo-planar MRI," *MRM*, vol. 34, pp. 537–541, 1995.

[4] A. J. Bell and T. J. Sejnowski, "An information-maximization approach to blind separation and blind deconvolution," *Neural Comp.*, vol. 7, pp. 1129–1159, 1995.

[5] M. J. McKeown et. al., "Analysis of fmri data by blind separation into spatial independent components," *HBM*, vol. 6, pp. 160–188, 1998.

[6] P. Filzmoser, R. Baumgartner, and E. Moser, "A hierarchical clustering method for analyzing functional MR images," *MRM*, vol. 17, pp. 817–826, 1999.

[7] D. Cordes et. al., "Hierarchical clustering to measure connectivity in fMRI resting-state data," *MRM*, vol. 20, pp. 305–317, 2002.

[8] P. Golland, Y. Golland, and R. Malach, "Detection of spatial activation patterns as unsupervised segmentation of fMRI data," in *MICCAI*, pp. 110–118.

[9] A. P. Dempster et. al., "Maximum likelihood from incomplete data via the EM algorithm," *JRSS: Series B (Methodological)*, vol. 39, pp. 1–38, 1977.

[10] Ulrike von Luxburg, "A tutorial on spectral clustering," Tech. Rep. TR-149, Max Planck Institute for Biological Cybernetics, 2006.

[11] B. Nadler et. al., "Diffusion maps, spectral clustering and reaction coordinates of dynamical systems," *ACHA*, vol. 21, pp. 113–127, 2006.

[12] J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE PAMI*, vol. 22, pp. 888–905, 2000.

[13] C. Fowlkes et. al., "Spectral grouping using the Nystrom method," *IEEE PAMI*, vol. 26, pp. 214–225, 2004.

[14] I. Kahn et. al., "Distinct cortical anatomy linked to subregions of the medial temporal lobe revealed by intrinsic functional connectivity," *J. of Neurophysiology*, vol. 100, pp. 129–139, 2008.

[15] S. M. Smith et. al., "Advances in functional and structural MR image analysis and implementation as FSL," *NeuroImage*, vol. 23, pp. 208–219, 2004.