1. Identification of Sparse Connectivity Patterns

Given N subject-level correlation matrices $\Sigma_1, \Sigma_2, \ldots \Sigma_N \in \mathbf{S}_+^P$, the sparse learning based dimensionality reduction method (Eavani et al., 2015) finds Sparse Connectivity Patterns (SCPs) $\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_K \in \mathbf{R}^P$ and associated SCP coefficients $\mathbf{c}_1, \mathbf{c}_2, \ldots \mathbf{c}_N \in \mathbf{R}_+^K$ by solving the following optimization problem:

$$
\begin{aligned}
& \underset{\mathbf{B},\mathbf{C}}{\text{minimize}} \sum_{n=1}^{N} \left|\left|\mathbf{\Sigma}_n - \mathbf{B}\ \text{diag}(\mathbf{c}_n)\ \mathbf{B}^T\right|\right|_F^2 \\
& \text{subject to} \\
& \qquad ||\mathbf{b}_k||_1 \leq \lambda, \quad k = 1, \ldots, K, \\
& \qquad -1 \leq \mathbf{b}_k(i) \leq 1, \quad \max_i |\mathbf{b}_k(i)| = 1, \quad i = 1, \ldots, P \\
& \qquad \mathbf{c}_n \geq 0, \qquad n = 1, \ldots, N
\end{aligned}
\tag{1}
$$

Each one of the K SCPs $\mathbf{b}_k$ consists of a small number of regions, controlled by the parameter λ. The average within-SCP connectivity within each subject is summarized in its associated scalar SCP coefficient $\mathbf{c}_n$. Thus, this method performs dimensionality reduction based on the covariance of correlation values across subjects. Each SCP consists of those regions whose pair-wise correlation value covaries across subjects.

The user-defined parameters, number of SCPs K and the sparsity controlling parameter λ can be found using split-sample validation. In this strategy, the dataset is repeatedly split into halves. For each set of parameters (K, λ), the method is independently applied to each half. The resulting SCPs are matched one-to-one using the Hungarian algorithm (Munkres, 1957). Then are compared between the two halves using inner-product, which measures the reproducibility of the SCP. In addition the approximation error is estimated by computing the fit of SCPs from the first half to the data from the second half (and vice versa). Values of (K, λ) for which reproducibility is high, and error is low is chosen for final analysis. The variation of the reproducibility and error for varying (K, λ) parameters is shown in Figures 1 and 2.

Hierarchical decomposition of the data can be obtained by reapplying the method on re-weighted correlation matrices. Foe each of the K primary SCPs, one can define the re-weighted data as the Hadamard (element-wise) product of the correlation matrices and the SCP approximation: $\mathbf{\Sigma}_n \circ \left(\mathbf{b}_k * \mathbf{b}_k^T\right)$.

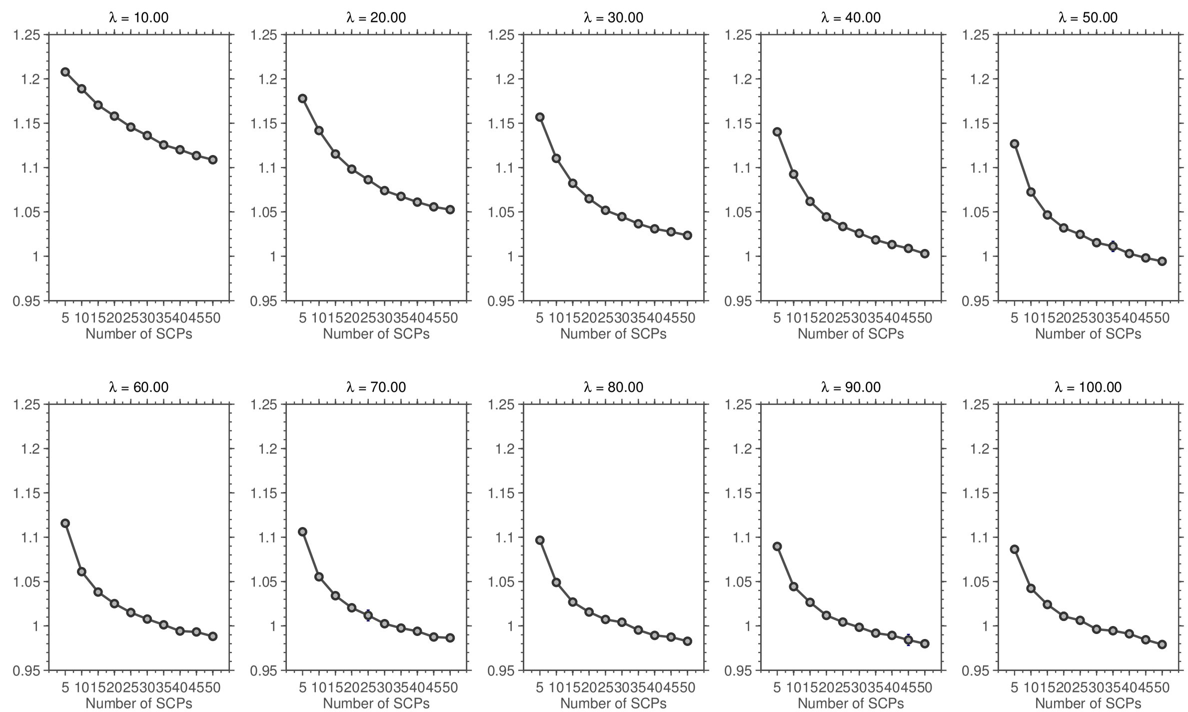


Figure 1: Variation of split-sample error with K, λ

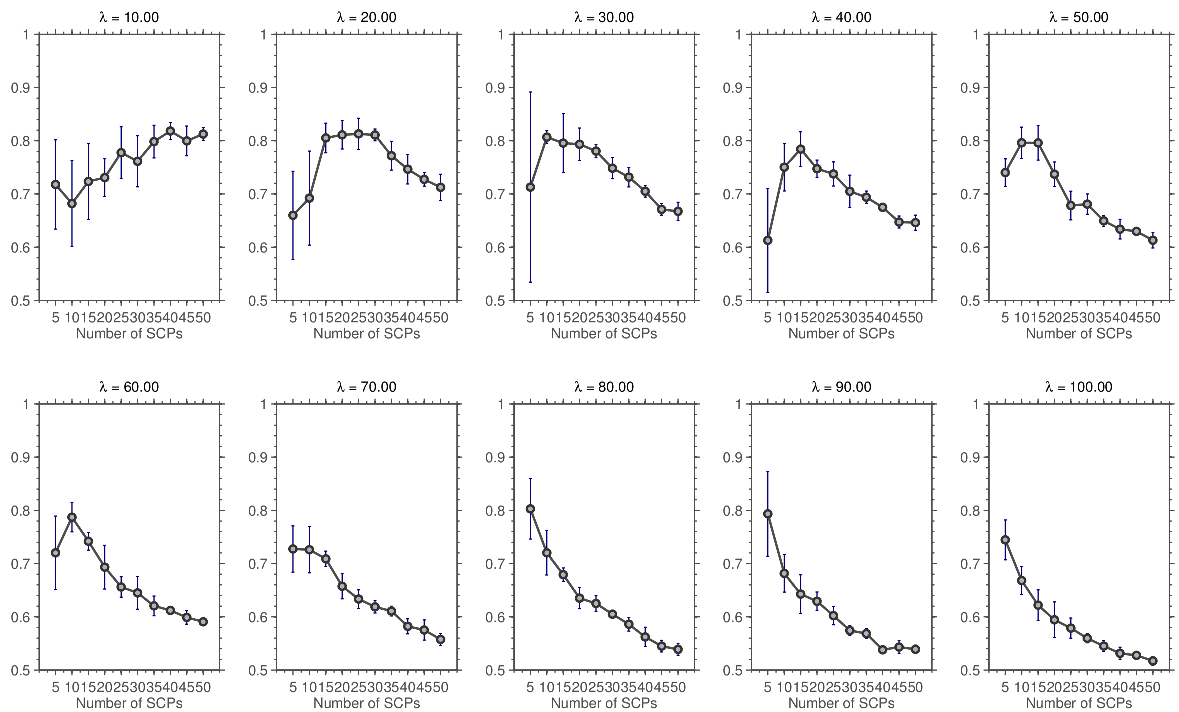


Figure 2: Variation of split-sample reproducibility with K, λ

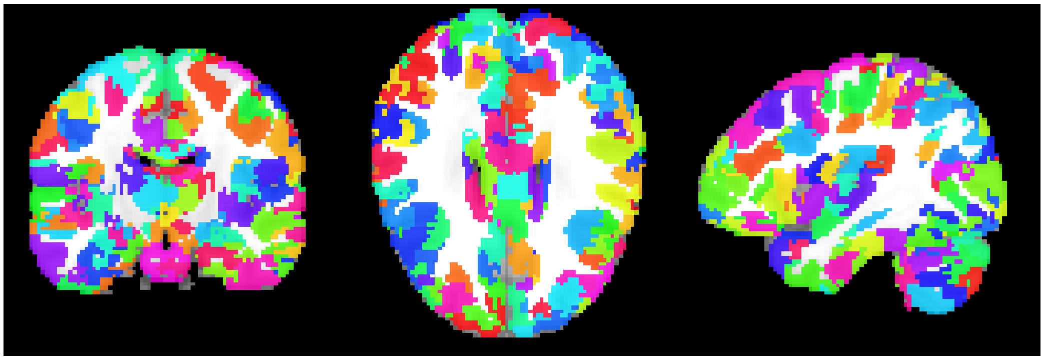


Figure 3: Common group-parcellation of BLSA data obtained using GraSP (Honnorat et al., 2015)

2. Data-driven rsfMRI-based parcellation of grey matter using GraSP

We obtained a common group parcellation of the grey matter volume in MNI space using GraSP (Honnorat et al., 2015). GraSP is a graph-based parcellation method that identifies spatially localized functionally coherent parcels that partition the entire grey matter volume. It is a data-driven method that relies solely on local functional connectivity to delineate parcels. It has a "label cost" parameters that controls the introduction of new parcels.

Applied to subjects from the BLSA data for a range of label cost values, we obtained parcellations of three resolutions - 127, 596, 3351. The parcellation with 596 parcels was chosen for our study, and is shown in Figure 3.

3. SVM Primal Formulation

Consider the standard primal version of the soft-margin SVM formulation, which is as follows:

$$\begin{aligned}
&\underset{\mathbf{w}^k}{\text{minimize}}\ \frac{1}{2}\left|\left|\mathbf{w}^k\right|\right|_1 + C\sum_{i=1}^{N}\xi_i^p \\
&\text{subject to} \\
&\qquad y_i\left(\mathbf{w}^T\mathbf{x}_i + b\right) \geq 1 - \xi_i \\
&\qquad \xi_i \geq 0 \quad i \in \{1, 2, \ldots N\}
\end{aligned} \tag{2}$$

When $p = 1$, the loss is linear ("Hinge" or ℓ_1 loss), and when $p = 2$, loss is quadratic (ℓ_2 loss).

The first inequality can be written as $\xi_i \geq 1 - y_i \left(\mathbf{w}^T\mathbf{x}_i + b\right)$. Given that the $\sum_{i=1}^{N} \xi_i$ term in the objective needs to be minimized, the optimal choice for ξ_i would be $1 - y_i \left(\mathbf{w}^T\mathbf{x}_i + b\right)$. However, the second constraint $\xi_i \geq 0$ cannot be violated. Hence the slack variable ξ_i becomes $\max\left(0, 1 - y_i \left(\mathbf{w}^T\mathbf{x}_i + b\right)\right)$. Plugging this into the objective, we get:

$$\underset{\mathbf{w}^k}{\text{minimize}} \ \frac{1}{2} \left|\left|\mathbf{w}^k\right|\right|_1 + C \sum_{i=1}^{N} \max\left(0, 1 - y_i \left(\mathbf{w}^T\mathbf{x}_i + b\right)\right)^p \tag{3}$$

This is an alternate (but equivalent) version of the SVM primal, which is used in our paper. The ℓ_2-loss primal is obtained by squaring the slack variables. In this case, the cost for mis-classification increases quadratically.

4. Cluster Validity Measures

4.1. Cluster reproducibility for fuzzy cluster membership

We evaluate the reproducibility of the sub-groups across repeated runs of the proposed method. We use the Adjusted-Rand Index (ARI) for fuzzy cluster assigments, as defined in Brouwer (2009). The ARI is a scalar value between $[-1, 1]$ which measures the extent to which two fuzzy cluster assignments are similar, after adjusting for chance.

Let $\mathbf{M} = [m_1, m_2, \ldots m_N], m_n \in \mathbf{R}^K$ define the membership values for N data-points for K clusters. Let $\mathbf{M}_1$ and $\mathbf{M}_2$ be two solutions to a clustering method.

Then define the bonding matrix $\mathbf{B} \in \mathbf{R}^{N \times N}$ as $\mathbf{B}_{i,j} = \langle m_i, m_j \rangle$ where $<>$ denotes normalized inner-product. The bonding matrix measures the similarity between each pair of memberships m_i, m_j.

Define

$$f(\mathbf{B}) = \frac{1}{N} \sum_{i,j} \mathbf{B}_{i,j}$$

$$g(\mathbf{B}) = f(\mathbf{B}) - \frac{N}{2}$$

Define the four measures of overlap a, b, c and d as follows:

$$
\begin{aligned}
a &= g(\mathbf{B}^1, (\mathbf{B}^2)^T) \\
b &= f(1 - \mathbf{B}^1, (\mathbf{B}^2)^T) \\
c &= f(\mathbf{B}^1, 1 - (\mathbf{B}^2)^T) \\
b &= f(1 - \mathbf{B}^1, 1 - (\mathbf{B}^2)^T)
\end{aligned}
$$

Then the Adjusted Rand Index (ARI) is defined as:

$$\text{ARI} = \frac{2(ad - bc)}{c^2 + b^2 + 2ad + (a + d)(b + c)}$$

For more details, see Brouwer (2009).

4.2. Cluster separation for fuzzy cluster membership

We evaluate the extent to which the K clusters are separated across repeated runs of the proposed method. We use the Bezdek Partition Coefficient (BPC) (Bezdek, 1981; Dave, 1996) which provides a scalar value between $[0, 1]$ for each fuzzy clustering assignment.

Let $m_n(k)$ define the membership value for the nth data-point for the kth cluster. Let N be the number of data-points, and K be the number of clusters.

Then BPC is defined as follows:

$$\text{BPC} = 1 - \frac{K}{K - 1}\left(1 - \frac{1}{N}\sum_{k=1}^{K}\sum_{n=1}^{N} m_n^2(k)\right)$$

5. Differences in cognition between groups

We used cognitive scores from the seven domains below to evaluate whether the subgroups obtained using the MOE model (using the SCP data alone) showed differences in cognition as well.

1. California Verbal Learning Task (CVLT) was used to assess verbal learning and memory. Higher values indicate better performance.
2. Benton Visual Retention Test (BVRT) quantifies figural memory and visuo-constructional ability. ability. Lower values indicate better performance.

3. CARD Rotation Test (CRT) measure the ability to mentally manipulate figures. Higher values indicate better performance.
4. Letter Fluency (FLULET) measures phonemic fluency. Higher values indicate better performance.
5. Category Fluency (FLUCAT) measures semantic fluency. Higher values indicate better performance.
6. Trail Making Test Part A (TRATS) was used as an indicator of visual attention and processing speed. Lower values indicate better performance.
7. Trail Making Test Part B (TRBTS) was used to evaluate executive function. Lower values indicate better performance.

The ANOVA and pair-wise results for concurrent cognitive data are shown in Table 1. Results obtained for the intercept and slope estimates are shown in Tables 2 and 3.

Test	Number of subjects	Mean estimate (Std. Error)			ANOVA p-value	Pair-wise p-value		
		Younger subjects	Older 1	Older 2		Younger vs Older 1	Younger vs Older 2	Older 1 vs Older 2
CVLT	69	56.8(1.5)	46.7(2.6)	42.5(3.1)	$< .0001$	0.0014	$< .0001$	0.31
BVRT	82	4.7(0.7)	11.1(0.9)	9.6(1.2)	$< .0001$	$< .0001$	0.0007	0.32
CRDROT	80	102.9(5.4)	68.6(7.8)	71.7(9.3)	0.0005	0.0006	0.0049	0.80
FLULET	61	14.6(0.6)	13.5(1.1)	16.1(1.2)	0.29	-	-	-
FLUCAT	61	18.3(0.5)	14.4(0.9)	14.6(1.0)	0.0001	0.0003	0.0018	0.86
TRATS	60	29.7(2.9)	41.9(5.0)	36.7(6.0)	0.097	-	-	-
TRBTS	60	57.8(4.9)	110.8(8.5)	84.4(10.2)	< 0.0001	< 0.0001	0.022	0.051

Table 1: Cross-sectional analysis: Table listing ANOVA results from cross-sectional analysis of concurrent cognitive scores obtained at time-of scan. The number of subjects for which this data was available, and the mean and standard error estimates for each group are also provided. The last four columns show the ANOVA p-value and the three pair-wise p-values respectively.

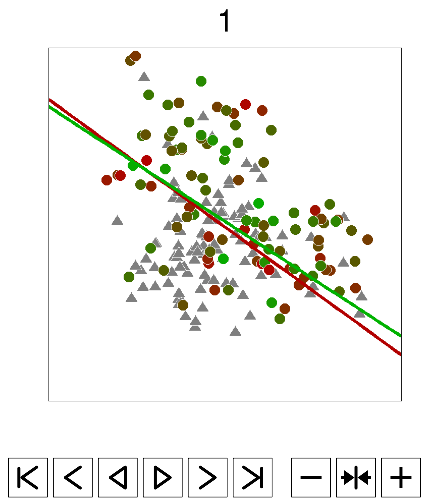
Test	Number of subjects (assessments)	Mean intercept estimate (Std. Error)			p-value	Pair-wise p-value		
		Younger subjects	Older 1	Older 2		Younger vs Older 1	Younger vs Older 2	Older 1 vs Older 2
CVLT	83(321)	57.1(1.7)	48.7(2.3)	49.6(2.8)	0.0070	0.0046	0.027	0.80
BVRT	86(345)	4.2(0.7)	8.7(0.9)	6.1(1.1)	0.0008	0.0002	0.17	0.054
CRDROT	86(341)	101.1(5.4)	68.0(6.9)	77.5(8.4)	0.0009	0.0003	0.020	0.35
FLULET	79(311)	14.4(0.6)	13.7(0.8)	16.2(0.9)	0.10	-	-	-
FLUCAT	79(311)	18.0(0.5)	15.0(0.7)	15.6(0.8)	0.0013	0.0006	0.013	0.55
TRATS	78(309)	28.0(1.9)	41.2(2.4)	30.6(2.9)	0.0002	< 0.0001	0.46	0.0047
TRBTS	79(309)	60.7(5.1)	106.2(6.6)	63.8(7.9)	< 0.0001	< 0.0001	0.90	< 0.0001

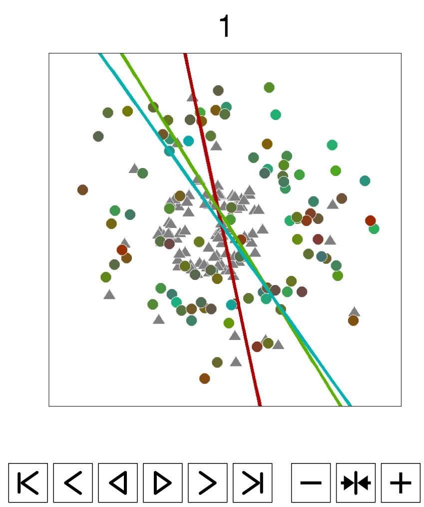
Table 2: Longitudinal Analysis: Table listing ANOVA results from analysis of intercept of longitudinal cognitive trajectories. The number of scans for which this data was available, and the mean and standard error estimates for each group are also provided. The values which showed a significant difference between older sub-group 1 and older sub-group 2 are shown in red.

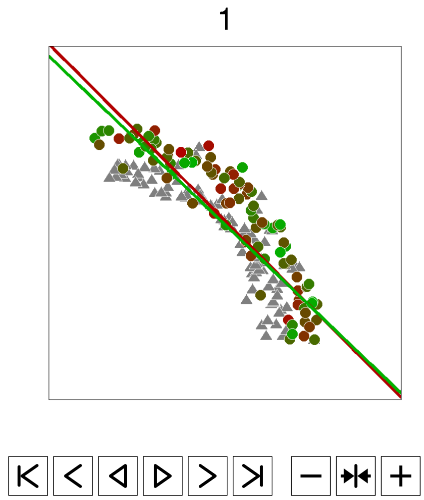
Test	Number of subjects (assessments)	Mean slope estimate (Std. Error)			p-value	Pair-wise p-value		
		Younger subjects	Older 1	Older 2		Younger vs Older 1	Younger vs Older 2	Older 1 vs Older 2
CVLT	83(321)	-0.011(0.17)	-0.16(0.16)	-0.30(0.17)	0.50	-	-	-
BVRT	86(345)	0.058(0.073)	0.099(0.073)	0.32(0.082)	0.049	0.69	0.020	0.0501
CRDROT	86(341)	0.65(0.48)	-0.056(0.44)	-0.080(0.47)	0.32	-	-	-
FLULET	79(311)	0.034(0.12)	-0.035(0.040)	-0.030(0.046)	0.85	-	-	-
FLUCAT	79(311)	0.086(0.098)	-0.038(0.036)	-0.078(0.041)	0.31	-	-	-
TRATS	78(309)	-0.29(0.39)	-0.25(0.13)	0.055(0.16)	0.32	-	-	-
TRBTS	79(309)	-0.88(1.07)	0.079(0.44)	1.24(0.49)	0.12	-	-	-

Table 3: Longitudinal Analysis: Table listing ANOVA results from analysis of slope of longitudinal cognitive trajectories. The number of scans for which this data was available, and the mean and standard error estimates for each group are also provided.

6. Simulation examples - animations







7. Global Signal Regression

SCPs used for the MOE analysis were generated after removing the baseline global signal from each subject's data. This pre-processing step was performed in order to remove the effect of motion and other physiological confounds from the data (Fox et al., 2009). When global signal regression was not performed, we found that many common functional systems were not clearly delineated (Eavani et al., 2015) and were substantively different, as seen in Figure 4. Of the ten SCPs computed, only one SCP had a significant areas of negative correlation - the Dorsal Attention vs. Default mode anti-correlation pattern. The other nine SCPs showed only positive correlations. Due to this lack of interpretability, we did not use these coefficients for the MOE analysis.

References

Bezdek, J. C., 1981. Pattern recognition with fuzzy objective function algorithms. Kluwer Academic Publishers.

Brouwer, R. K., 2009. Extending the rand, adjusted rand and jaccard indices to fuzzy partitions. Journal of Intelligent Information Systems 32 (3), 213–235.

Dave, R. N., 1996. Validating fuzzy partitions obtained through c-shells clustering. Pattern Recognition Letters 17 (6), 613–623.

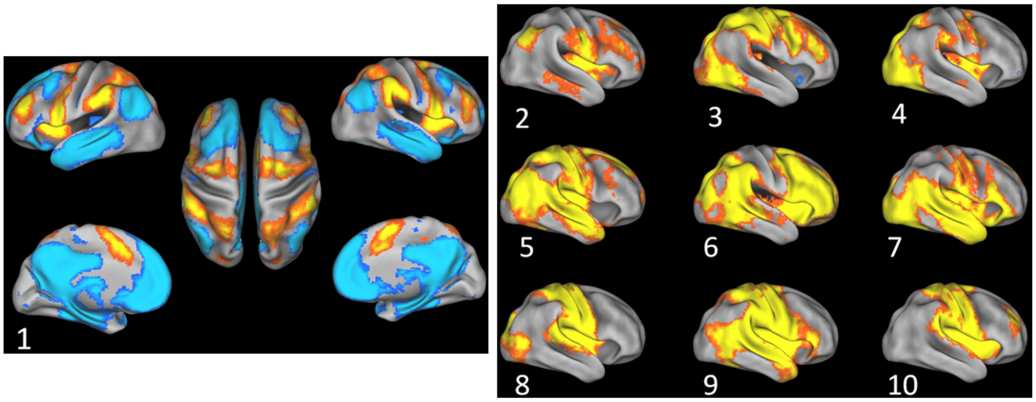


Figure 4: SCPs computed without global signal regression.

Eavani, H., Satterthwaite, T. D., Filipovych, R., Gur, R. E., Gur, R. C., Davatzikos, C., 2015. Identifying sparse connectivity patterns in the brain using resting-state fmri. NeuroImage 105, 286–299.

Fox, M. D., Zhang, D., Snyder, A. Z., Raichle, M. E., 2009. The global signal and observed anticorrelated resting state brain networks. Journal of neurophysiology 101 (6), 3270–3283.

Honnorat, N., Eavani, H., Satterthwaite, T., Gur, R., Gur, R., Davatzikos, C., 2015. Grasp: Geodesic graph-based segmentation with shape priors for the functional parcellation of the cortex. NeuroImage 106, 207–221.

Munkres, J., 1957. Algorithms for the assignment and transportation problems. Journal of the Society for Industrial & Applied Mathematics 5 (1), 32–38.