Identification of Recurrent Patterns in the Activation of Brain Networks
Technical Supplement

Firdaus Janoos[*]   **Weichang Li**   **Niranjan Subrahmanya**
ExxonMobil Corporate Strategic Research
Annandale, NJ 08801

István Á. Mórocz   **William M. Wells (III)**
Harvard Medical School
Boston, MA 02115

Consider the optimization problem of

$$\max_{\mathbf{g}}\langle\mathbf{g},\mathbf{dz}\rangle \qquad \text{such that} \qquad A\mathbf{g} \leq \mathbb{1}_{N_{\mathbb{V}}\times(N_{\mathbb{V}}-1)} \tag{1}$$

as defined in eqn. (2) of the Main text.

A   Similarity Transformation

Let the *distance* between the nodes of the graph be encoded in the matrix $W_{\mathbb{G}}$, and let $\Delta_{\mathbb{G}}$ be the corresponding *adjacency* matrix, such that $\Delta_{\mathbb{G}}[i,j] = W_{\mathbb{G}}[i,j]^{-1}$, $\forall i \neq j$. Therefore, its graph Laplacian is $\mathcal{L}_{\mathbb{G}} = D_{\Delta_{\mathbb{G}}} - \Delta_{\mathbb{G}}$, where $D_{\Delta_{\mathbb{G}}}$ is the diagonal degree matrix with $D_{\Delta_{\mathbb{G}}}[i,i] = \sum_{j\in\mathbb{V}}\Delta_{\mathbb{G}}[i,j]$ and $D_{\Delta_{\mathbb{G}}}[i,j] = 0$, for $i \neq j$.

The similarity transformation $M = A^{-}\text{diag}\{\mathbf{w}_{\mathbb{G}}\}^{-1}A$, where A^{-} is the pseudo-inverse of A. Now, because $A^{\top}A$ is a Toeplitz $N_{\mathbb{V}} \times N_{\mathbb{V}}$ matrix with of rank $N_{\mathbb{V}} - 1$, with all diagonal entries equal to $2(N_{\mathbb{V}} - 1)$ and all non-diagonal elements equal to -2, the SVD of a $A = \sqrt{2N_{\mathbb{V}}}UI_{N_{\mathbb{V}}-1}V^{\top}$ where $I_{N_{\mathbb{V}}-1}$ is the $N_{\mathbb{V}} \times N_{\mathbb{V}}$ identity matrix with 0 on its last entry.

Therefore,

$$A^{-} = \frac{1}{\sqrt{2N_{\mathbb{V}}}}VI_{N_{\mathbb{V}}-1}U^{\top} = \frac{1}{N_{\mathbb{V}}}A^{\top} \qquad \text{giving} \qquad M = \frac{1}{N_{\mathbb{V}}}A^{\top}\text{diag}\{\mathbf{w}_{\mathbb{G}}\}^{-1}A = \frac{1}{N_{\mathbb{V}}}\mathcal{L}_{\mathbb{G}}. \tag{2}$$

B   Error Analysis

Let the term $\mathbf{a}_{i,j}$ denote the row of the constraint matrix A corresponding to the constraint $g[i] - g[j] \leq 1$.

Proposition 1. *If $\mathbf{g}^{*}$ is an optimal solution to eqn. (1), then for every pair of constraints such that*

$$\langle\mathbf{a}_{i,j},\mathbf{g}^{*}\rangle = 1 \qquad and \qquad \langle\mathbf{a}_{l,m},\mathbf{g}^{*}\rangle = 1$$

it is the case that:

$$\langle\mathbf{a}_{i,j},\mathbf{a}_{j,k}\rangle \geq 0$$

---

[*]Corresponding Author. firdaus.janoos@exxonmobil.com

Restated, this theorem implies that an extreme point is formed by the intersection of hyper-planes whose normal vectors (*i.e.* constraints) either make an inner product of 0 or 1 with each other.

Proof. Here, $\mathbf{a}_{i,j}$ and $\mathbf{a}_{l,m}$ represent two distinct constraint vectors of magnitude $\sqrt{2}$ corresponding to the edges (i,j) and (l,m) of $\mathbb{G}$, where each has one $+1$ and one -1 entry. The inner product between them can therefore have 3 values: (a) 0 (*i.e.* orthogonal), (b) $+1$ (*i.e.* $\pi/3$ radians), (c) -1 (*i.e.* $2\pi/3$ radians), or (d) -2 (*i.e.* π radians).

In case (d), the two hyper-planes do not intersect, while the $N_{\mathbb{V}} - 2$ dimensional space arising from case (c), violates the constraints of the dual problem, because $\langle \mathbf{a}_{i,j}, \mathbf{a}_{l,m} \rangle = -1$ is equivalent to an inner product of the form $\langle \mathbf{a}_{i,j}, \mathbf{a}_{l,i} \rangle = -1$. This leads to:

$$+\, g[i] - g[j] = 1$$
$$+\, g[l] - g[i] = 1$$
$$\Rightarrow +\, g[l] - g[j] = 2 \qquad \text{which violates constraint } g[l] - g[j] \leq 1$$

$\square$

Proposition 2. *If $\mathbf{g}^*$ is an optimal solution to eqn. (1), then*

$$\mathbf{g}^* = \frac{1}{N_{\mathbb{V}}} \left(+(d_1^+, d_2^+, \ldots d_{N_{\mathbb{V}}}^+)^\top - (d_1^-, d_2^-, \ldots d_{N_{\mathbb{V}}}^-)^\top \right). \tag{3}$$

Here, $(d_1^+ \ldots d_{N_{\mathbb{V}}}^+)$ are the number of $+1$ entries and $(d_1^- \ldots d_{N_{\mathbb{V}}}^-)$ are the number of -1 entries in the $1 \ldots N_{\mathbb{V}}$ columns of the matrix E of active constraints for $\mathbf{g}^*$ (i.e. $\mathrm{E}\mathbf{g}^* = \mathbb{1}$), such that:

$$\left. \begin{array}{cccc} & d_i^+ = N_{\mathbb{V}}^+ & AND & d_i^- = 0 \\ OR & d_i^+ = 0 & AND & d_i^- = N_{\mathbb{V}}^- \end{array} \right\} \qquad \textit{for all} \qquad i = 1 \ldots N_{\mathbb{V}},$$

with $N_{\mathbb{V}}^+ \geq 1$, $N_{\mathbb{V}}^- \geq 1$ and $N_{\mathbb{V}}^+ + N_{\mathbb{V}}^- = N_{\mathbb{V}}$.

Corollary. *A straightforward result of Proposition 2 is that:*

$$\mathbf{g}^* = \frac{1}{N_{\mathbb{V}}} \sum_{i,j | \mathbf{a}_{i,j} \in \mathcal{E}} \mathbf{a}_{i,j}. \tag{4}$$

Here $\mathcal{E} = \{ \mathbf{a}_{i,j} \mid \langle \mathbf{a}_{i,j}, \mathbf{g}^* \rangle = 1 \}$ is the set of constraints that active at $\mathbf{g}^*$.

Proof. By Proposition 1, an optimal extreme point $\mathbf{g}^*$ is constructed by selecting a subset $\mathbb{V}^+ \subset \mathbb{V}$ of as sources nodes corresponding to $+g[i]$ term for the active constraints at $\mathbf{g}^*$ while the remaining $\mathbb{V}^- \triangleq \mathbb{V} \backslash \mathbb{V}^+$ of nodes as destinations corresponding to $-g[j]$ in the active constraint set[1]. Therefore, constraint $\mathbf{a}_{i,j}$ is active $\iff i \in \mathbb{V}^+$ and $j \in \mathbb{V}^-$.

As a result, in the set $\mathcal{E}$ of active constraints at $\mathbf{g}^*$, every source node i will pair with $N_{\mathbb{V}}^- \triangleq |\mathbb{V}^-|$ destination nodes, and therefore participate in $N_{\mathbb{V}}^-$ constraints. Similarly, every destination node j (corresponding to a -1 entry in the j–th column of the active constraint matrix E) will participate in $N_{\mathbb{V}}^+ \triangleq |\mathbb{V}^+|$ constraints. Therefore, the active constraint set $\mathcal{E}$ for $\mathbf{g}^*$ will have $N_{\mathbb{V}}^+$ vectors $\mathbf{a}_{i,.}$ with $+1$ for the i–th node and $N_{\mathbb{V}}^-$ vectors $\mathbf{a}_{.,j}$ with -1 for the j–th node.

Define E as the matrix of constraints $\langle \mathbf{a}_{i,j}, \mathbf{g} \rangle = 1$ active at $\mathbf{g}^*$. This matrix E will have $N_{\mathbb{V}}^+$ entries of $+1$ in the i–th column of source node i and $N_{\mathbb{V}}^-$ entries of -1 in the j–th column of source node j. That is, it will have either $N_{\mathbb{V}}^+ \times +1$ or $N_{\mathbb{V}}^- \times -1$ entries in any column [2].

Now defining the $N_{\mathbb{V}}$–vector

$$\eta \triangleq \sum_{i,j | \mathbf{a}_{i,j} \in \mathcal{E}} \mathbf{a}_{i,j}$$
$$= +\, (d_1^+, d_2^+, \ldots d_{N_{\mathbb{V}}}^+)^\top - (d_1^-, d_2^-, \ldots d_{N_{\mathbb{V}}}^-)^\top,$$

we see that

$$\mathrm{E}\eta = N_{\mathbb{V}} \mathbb{1}_{N_{\mathbb{V}}}$$

---

[1] Note that there should be at least one sink and one source node involved in the construction of any $\mathbf{g}^*$.

[2] The equality constraint $\langle \mathbb{1}_{N_{\mathbb{V}}}, \mathbf{g} \rangle = 0$ can be added as the last row of E to give it a rank of $N_{\mathbb{V}}$, without changing the argument.

Therefore, this yields the desired result that the extreme-point

$$\mathbf{g}^* = \frac{\eta}{N_\mathbb{V}}$$

is the unique solution to $\mathbf{E}\mathbf{g}^* = \mathbb{1}_{N_\mathbb{V}}$. $\qquad\square$

Remark: We observe that this construction which requires selecting a node as source or destination, along with the fact that there must be at least one source and one destination node, implies that there are $2^{N_\mathbb{V}} - 2$ extreme points for the polytope $\mathbf{A}\mathbf{g} \leq 1$.

Proposition 3. *Let $\mathbf{g}^*$ be any extreme-point of the polytope $\mathbf{A}\mathbf{g} \leq 1$. Then the optimal solution $\mathbf{g}^*$ to eqn.* (1) *satisfies the property:*

$$\left\langle \mathbf{g}^*, \frac{\mathbf{dz}}{||\mathbf{dz}||} \right\rangle \geq \frac{1}{\sqrt{2}}$$

Proof. At the optimum point, by the strict duality of LP, there exists a primal solution $\mathbf{f}^* : \mathbb{E} \to \mathbb{R}^+$, such that $\left\langle \mathbf{f}^*, \mathbb{1}_{N_\mathbb{V} \times (N_\mathbb{V} - 1)} \right\rangle = \langle \mathbf{g}^*, \mathbf{dz} \rangle$ and $\mathbf{f}^{*\top} \mathbf{A} = \mathbf{dz}^\top$, and

$$\left. \begin{array}{ll} f^*[i,j] > 0 & \iff \quad g^*[i] - g^*[j] = 1 \\ f^*[i,j] = 0 & \iff \quad g^*[i] - g^*[j] < 1 \end{array} \right\} \qquad \text{by complementary slackness.}$$

Therefore, $\mathbf{dz}$ is in the positive cone of the active constraint set $\mathcal{E} = \{\mathbf{a}_{i,j} \mid \langle \mathbf{a}_{i,j}, \mathbf{g}^* \rangle = 1\}$ as per:

$$\mathbf{dz} = \sum_{i,j|\mathbf{a}_{i,j}\in\mathcal{E}} f^*[i,j]\mathbf{a}_{i,j}, \qquad \text{with} \qquad f^*[i,j] \geq 0. \tag{5}$$

Furthermore, by eqn. (4), $\mathbf{g}^*$ also belongs to this cone:

$$\mathbf{g}^* = \frac{1}{N_\mathbb{V}} \sum_{i,j|\mathbf{a}_{i,j}\in\mathcal{E}} \mathbf{a}_{i,j}. \tag{6}$$

Therefore, by the strict duality property:

$$\begin{aligned}
\frac{1}{||\mathbf{dz}||}\langle \mathbf{g}^*, \mathbf{dz} \rangle &= \frac{\sum_{i,j|\mathbf{a}_{i,j}\in\mathcal{E}} f^*[i,j]}{\left[\sum_{i,j|\mathbf{a}_{i,j}\in\mathcal{E}} \sum_{l,m|\mathbf{a}_{l,m}\in\mathcal{E}} f^*[i,j]f^*[l,m]\langle\mathbf{a}_{i,j},\mathbf{a}_{l,m}\rangle\right]^{\frac{1}{2}}} \\
&\geq \frac{\sum_{i,j|\mathbf{a}_{i,j}\in\mathcal{E}} f^*[i,j]}{\sqrt{2}\left[\sum_{i,j|\mathbf{a}_{i,j}\in\mathcal{E}} f^*[i,j]\right]} \\
&\geq \frac{1}{\sqrt{2}} \tag{7}
\end{aligned}$$

The second last inequality follows from $0 \leq \langle \mathbf{a}_{i,j}, \mathbf{a}_{l,m} \rangle \leq 2$ by Proposition 1, giving:

$$\begin{aligned}
\sum_{i,j|\mathbf{a}_{i,j}\in\mathcal{E}} \sum_{l,m|\mathbf{a}_{l,m}\in\mathcal{E}} f^*[i,j]f^*[l,m]\langle\mathbf{a}_{i,j},\mathbf{a}_{l,m}\rangle &\leq 2 \sum_{i,j|\mathbf{a}_{i,j}\in\mathcal{E}} \sum_{l,m|\mathbf{a}_{l,m}\in\mathcal{E}} f^*[i,j]f^*[l,m] \\
&\leq 2\left[\sum_{i,j|\mathbf{a}_{i,j}\in\mathcal{E}} f^*[i,j]\right]^2.
\end{aligned}$$

$\qquad\square$

Theorem 1. *Defining $\mathbf{dz} \triangleq \mathbf{z}_{t_1} - \mathbf{z}_{t_2}$, the worst-case error ϵ for the spherical relaxation*

$$\widehat{\mathrm{TD}}(\mathbf{z}_{t_1}, \mathbf{z}_{t_2}) = \frac{1}{\sqrt{2}}||\mathbf{dz}|| = \frac{1}{\sqrt{2}}||\mathbf{z}_{t_1} - \mathbf{z}_{t_2}|| \qquad \text{with} \qquad \widehat{\mathbf{g}}^* = \frac{1}{\sqrt{2}}\frac{\mathbf{dz}}{||\mathbf{dz}||} \tag{8}$$

of the transportation distance $\mathrm{TD}(\mathbf{z}_{t_1}, \mathbf{z}_{t_2})$ is

$$\epsilon \leq \frac{1}{\sqrt{2}}||\mathbf{dz}||. \tag{9}$$

3

The worst-case error $|\text{TD}(\mathbf{z}_{t_1}, \mathbf{z}_{t_2}) - \widehat{\text{TD}}(\mathbf{z}_{t_1}, \mathbf{z}_{t_2})|$ for an arbitrary graph is:

$$\epsilon \leq \frac{\lambda_{\min}^{-1}}{\sqrt{2}} ||\mathbf{dz}||, \tag{10}$$

where $\lambda_{\min}$ is smallest non-zero eigenvalue of the unnormalized Laplacian $\mathcal{L}_{\mathbb{G}}$ of the graph $\mathbb{G}$.

Proof. Assuming constant $||\mathbf{dz}|| = \alpha$, the worst-case error is:

$$\begin{aligned}
\epsilon &= \max_{\mathbf{dz}, ||\mathbf{dz}||=\alpha} \left| \langle \mathbf{g}^*, \mathbf{dz} \rangle - \frac{||\mathbf{dz}||}{\sqrt{2}} \right|, \\
&= \max_{\mathbf{dz}} \alpha \left| \left\langle \mathbf{g}^*, \frac{\mathbf{dz}}{\alpha} \right\rangle - \frac{1}{\sqrt{2}} \right|,
\end{aligned} \tag{11}$$

where $\mathbf{g}^*$ is the optimal solution to eqn. (1).

Now, Proposition 3 implies that for any cost vector $\mathbf{dz}$, there exists an extreme point of the polytope $A\mathbf{g}^* \leq 1$, such that

$$\frac{1}{||\mathbf{dz}||} \langle \mathbf{g}^*, \mathbf{dz} \rangle \geq \frac{1}{\sqrt{2}}.$$

Therefore, eqn. (11) becomes

$$\epsilon = \alpha \left[\max_{\mathbf{dz}} \left\langle \mathbf{g}^*, \frac{\mathbf{dz}}{\alpha} \right\rangle - \frac{1}{\sqrt{2}} \right],$$

where now $\mathbf{g}^*$ is now *any extreme point* of the polytope $A\mathbf{g}^* \leq 1$.

This achieves a maxima for $\mathbf{dz}/\alpha = \mathbf{g}^*/||\mathbf{g}^*||$, and therefore, we get:

$$\begin{aligned}
\epsilon &= \alpha \left[\left\langle \mathbf{g}^*, \frac{\mathbf{g}^*}{||\mathbf{g}^*||} \right\rangle - \frac{1}{\sqrt{2}} \right], \\
&= ||\mathbf{dz}|| \left[||\mathbf{g}^*|| - \frac{1}{\sqrt{2}} \right] \qquad \text{(because } \alpha = ||\mathbf{dz}||) \\
&\leq ||\mathbf{dz}|| \left[\sqrt{\frac{N_{\mathbb{V}} - 1}{N_{\mathbb{V}}}} - \frac{1}{\sqrt{2}} \right], \\
&\lessapprox \frac{1}{\sqrt{2}} ||\mathbf{dz}||
\end{aligned} \tag{12}$$

The second-last inequality is obtained from eqn. (3), which gives that:

$$\begin{aligned}
||\mathbf{g}^*|| &\leq \left[\frac{(N_{\mathbb{V}} - 1)^2 + \sum_{k=1}^{N_{\mathbb{V}}-1} 1^2}{N_{\mathbb{V}}^2} \right]^{\frac{1}{2}} \\
&\leq \left[\frac{N_{\mathbb{V}}^2 - N_{\mathbb{V}}}{N_{\mathbb{V}}^2} \right]^{\frac{1}{2}} \\
&\leq \left[\frac{N_{\mathbb{V}} - 1}{N_{\mathbb{V}}} \right]^{\frac{1}{2}}.
\end{aligned}$$

Now, for an arbitrary graph, substitute $\widehat{\mathbf{dz}} = \Lambda V^{\top} \mathbf{dz}$ in eqn. (12), where $V \Lambda V^{\top}$ is the eigen-system of the graph Laplacian $\mathcal{L}_{\mathbb{G}}$, to give:

$$\epsilon \lessapprox \frac{1}{\sqrt{2}} ||\widehat{\mathbf{dz}}|| = \frac{1}{\sqrt{2}} ||\Lambda^{-} V^{\top} \mathbf{dz}|| \leq \frac{\lambda_{\min}^{-1}}{\sqrt{2}} ||\mathbf{dz}||. \tag{13}$$

$\square$

C   Functional Networks

Functional networks are routinely defined by the "temporal correlations between spatially remote neurophysiological events" [3]. The following section discusses an algorithm of computing the functional connectivity (i.e. correlations) ρ between voxels that is consistent, sparse and computationally efficient. Because $N \gg T$, the standard covariance estimator is badly conditioned, and its eigen-system is inconsistent [4]. Therefore, regularization is required to impose sensible structure on the estimated covariance matrix while being computationally efficient.

First, the images are smoothed with a Gaussian kernel (FWHM=8mm) to increase spatial coherence of the time-series data. Next, spatially proximal voxels are grouped into a set of $\tilde{N} < N$ spatially contiguous clusters using hierarchical agglomerative clustering (HAC) [1] as described in C.1. Then, cluster-wise covariances are computed and regularized using adaptive soft shrinkage [4] using the procedure detailed in C.2. Finally, estimates of voxel-wise correlations are then recomputed from the regularized cluster-wise correlations as per C.3.

If $i, j = 1 \dots N$ index two cortical voxels, then the functional connectivity map $\rho[i,j] \in [-1,1]$ for all $1 \leq i, j \leq N$ is consistent and extremely sparse. It is also easy to verify that ρ is positive definite. The results of this procedure on the distribution of the functional connectivity estimates on the data-set of Section. 4.2 are shown in Fig. 1
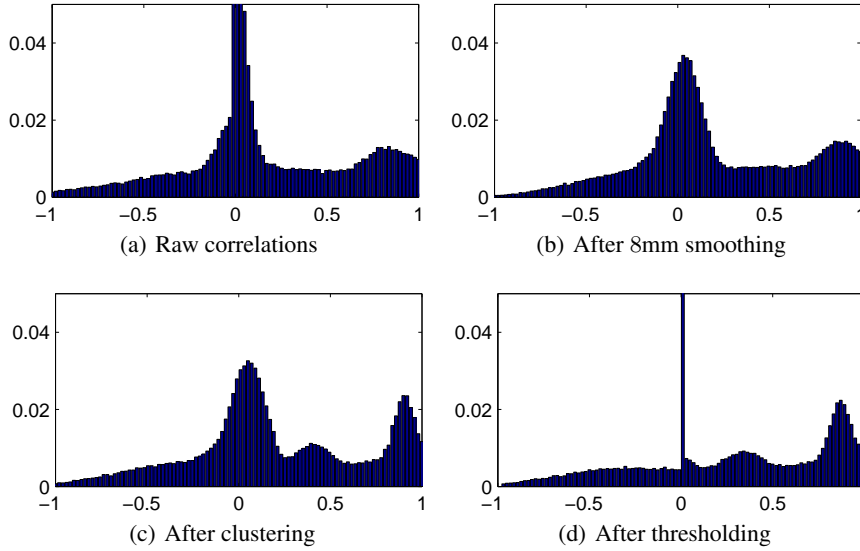

(a) Raw correlations

(b) After 8mm smoothing

(c) After clustering

(d) After thresholding

Figure 1: (a) Without any regularization, most of the mass of the distribution is in concentrated in small non-zero correlations, while the strong correlations are only a fraction of the total. (b) The smoothing procedure shifts the whole distribution towards the right, by strengthening all correlations. (c) The hierarchical clustering procedure boosts strong correlations without affecting weak correlations. (d) Finally, the shrinkage step sparsifies the correlation matrix, with most correlations set to zero.

In the discussion that follows define $\mathbf{y} \triangleq \{\mathbf{y}_1 \dots \mathbf{y}_T\}$ as the fMRI time–series data with N voxels and T scans, where $\mathbf{y}[i]$ is the time–series data of voxel $i = 1 \dots N$.

C.1   Hierarchical Agglomerative Clustering

Algorithm 1 describes the HAC procedure used to group together spatial proximal and functionally correlated voxels.

The time–series for the new cluster c_k is defined as $\mathbf{y}[k] = 1/n_k \sum_{c_i \in c_k} \mathbf{y}[i]$, and for a new cluster $c_k = (c_i, c_j)$ can be efficiently updated according to $\mathbf{y}[k] = (n_i\mathbf{y}[i] + n_j\mathbf{y}[j])/(n_i + n_j)$.

begin // `Initialization`
   For each voxel i, create one cluster c_i of size $n_i = 1$
   Each c_i is associated with a time–course $\mathbf{y}[i]$
end
while *Number of clusters greater than specified value* **do**
   Find two clusters c_i and c_j that are spatially adjacent to each other and merge them into a new cluster
   $c_k = (c_i, c_j)$, if and only if $\mathbb{V}\mathrm{ar}\{c_k\}$ is minimum over all i, j
   Remove clusters c_i and c_j from the set of clusters, and add c_k
end

Algorithm 1: Hierarchical Agglomerative Clustering

The variance of a cluster c_k is $\mathbb{V}\mathrm{ar}\{c_k\} = (1/n_k T)\sum_{c_i \in c_k}\sum_{t=1}^{T}(\mathbf{y}[i] - \mathbf{y}[k])^2$, and is efficiently updated through the variance separation theorem:

$$\mathbb{V}\mathrm{ar}\{c_k\} = \frac{n_i \mathbb{V}\mathrm{ar}\{c_i\} + n_j \mathbb{V}\mathrm{ar}\{c_j\}}{n_i + n_j} - \frac{\sum_{t=1}^{T}(\mathbf{y}[i] - \mathbf{y}[k])^2}{T(n_i + n_j)}.$$

After hierarchical clustering, the covariance $\sigma[k_1, k_2] \triangleq \mathbb{V}\mathrm{ar}\{c_{k_1}, c_{k_2}\}$ between two clusters c_{k_1} and c_{k_2} is estimated as:

$$\sigma[k_1, k_2] = \frac{1}{T}\sum_{t=1}^{T}\mathbf{y}_t[k_1]\mathbf{y}_t[k_2] - \left(\frac{1}{T}\sum_t \mathbf{y}_t[k_1]\right)\frac{1}{T}\left(\sum_t \mathbf{y}_t[k_2]\right).$$

HAC is repeated until the number of clusters $\widetilde{N} \approx 0.25 \times N$. This procedure has a two-fold benefit of reducing the dimensionality of the estimation problem while simultaneously increasing the SNR of the data through averaging. Table 1 shows that the clusters, after Gaussian smoothing, are larger and their sizes are more uniform for the same number of HAC–steps as compared to those without smoothing.

HAC–steps	$0.5 \times N$			$0.75 \times N$			$0.875 \times N$		
FWHM	$\widetilde{N}$	Avg mm^3	Std.dev	$\widetilde{N}$	Avg mm^3	Std.dev	$\widetilde{N}$	Avg mm^3	Std.dev
0mm	0.69	11.59	6.32	0.51	15.68	14.62	0.36	22.22	30.33
4mm	0.63	12.69	4.85	0.42	19.04	9.16	0.29	27.58	17.41
8mm	0.58	13.79	3.98	0.34	23.52	7.70	0.22	36.36	12.56

Table 1: **Effect of FWHM of the Gaussian kernel on the number of clusters $\widetilde{N}$ (as a fraction of N).** The mean and standard deviation of cluster sizes after a certain number of HAC–steps are shown. Values are for the data-set described in Section. 4.2.

C.2 Shrinkage

Next, cluster-wise covariances are computed and regularized using adaptive soft shrinkage [4].

$$s_\lambda(\sigma[k_1, k_2]) = \mathrm{sgn}(\sigma[k_1, k_2])\left(|\sigma[k_1, k_2]| - \lambda|\sigma[k_1, k_2]|^{-1}\right)_+. \tag{14}$$

This estimator has the property that the shrinkage is continuous with respect to $\sigma[k_1, k_2]$, but the amount of shrinkage decreases as $\sigma[k_1, k_2]$ increases resulting in less bias than the standard soft shrinkage estimator. The threshold parameter λ is selected by minimizing the risk function $R(\lambda) = \mathbb{E}||s_\lambda(\sigma) - \sigma||_2$. Under certain regularity assumptions about the data, a closed–form estimate of the optimal threshold is obtained as [2]:

$$\lambda \approx \frac{\sum_{k_1 \neq k_2}\mathbb{V}\mathrm{ar}\{\sigma_{k_1, k_2}\}}{\sum_{k_1 \neq k_2}\sigma_{k_1, k_2}^2}, \tag{15}$$

where $\mathbb{V}\mathrm{ar}\{\}\sigma_{k_1, k_2}]$ is estimated as:

$$\frac{T}{(T-1)^3}\sum_{t=1}^{T}\left(\mathbf{y}_t[i]\mathbf{y}_t[j] - \sum_{t'=1}^{T}\mathbf{y}_{t'}[i]\mathbf{y}_{t'}[j]\right)^2.$$

This estimator is "sparsistent" [4], that is, in addition to being consistent, it estimates true zeros as zeros and non-zero elements as non-zero with the correct sign, with probability tending to 1.

C.3 Voxel-wise Correlations

The covariance between the time-series $\mathbf{y}[i]$ of a voxel i belonging to cluster c_k and the cluster average time-series $\mathbf{y}[k]$ is:

$$\sigma[i,k] = \frac{1}{T}\sum_t \mathbf{y}_t[i]\mathbf{y}_t[k] - \left(\frac{1}{T}\sum_t \mathbf{y}_t[i]\right)\left(\frac{1}{T}\sum_t \mathbf{y}_t[k]\right),$$

and the correlation coefficient is:

$$\rho[i,k] = \frac{\sigma[i,k]}{\sqrt{\sigma[i,i]\sigma[k,k]}}. \tag{16}$$

Also, the smoothed (i.e. conditionally expected) time-series $\mathbf{y}[i|k] \triangleq \mathbb{E}\{[\}\mathbf{y}[i] \mid \mathbf{y}[k]]$ is:

$$\mathbf{y}[i|k] = \frac{1}{T}\sum_t \mathbf{y}_t[i] + \sigma[i,k]\sigma[k,k]^{-1}\left(\mathbf{y}[k] - \frac{1}{T}\sum_t \mathbf{y}_t[k]\right), \quad \text{and} \quad \sigma[i|k] = \sigma[i,i] - \frac{\sigma[i,k]^2}{\sigma[k,k]}. \tag{17}$$

is its conditional variance $\sigma[i|k] \triangleq \mathbb{V}\mathrm{ar}\{[\}\mathbf{y}[i] \mid \mathbf{y}[k]]$.

Therefore, the expected (smoothed) correlation between two voxels i and j belonging to clusters c_{k_i} and c_{k_j} respectively are obtained by substituting eqns. 17 and 16 to get:

$$\rho[i,j] = \frac{\mathbb{C}\mathrm{ov}\{\mathbf{y}[i|k_i], \mathbf{y}[j|k_j]\}}{\sqrt{\sigma[i|k_i]\sigma[j|k_j]}} = s_\lambda(\sigma[k_i,k_j]) \cdot \frac{\rho[i,k_i]\rho[j,k_j]}{\sqrt{(1-\sigma[k_i,k_i]\rho[i,k_i]^2)(1-\sigma[k_j,k_j]\rho[j,k_j]^2)}}. \tag{18}$$

References

[1] Jain, A.K.: Data clustering: 50 years beyond k-means. Pat Recog Letters 31(8), 651 – 666 (2010) 5

[2] Ledoit, O., Wolf, M.: Honey, I shrunk the sample covariance matrix. J Portfolio Man 30(4), 110–119 (2004) 6

[3] Li, K., Guo, L., Nie, J., Li, G., Liu, T.: Review of methods for functional brain connectivity detection using fMRI. J Comp Med Imag Graphics. 33(2), 131 – 139 (2009) 5

[4] Rothman, A., Levina, E., Zhu, J.: Generalized thresholding of large covariance matrices. J Am Stat Assoc (Theory Methods) 104, 177–186 (2009) 5, 6