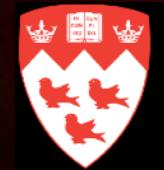




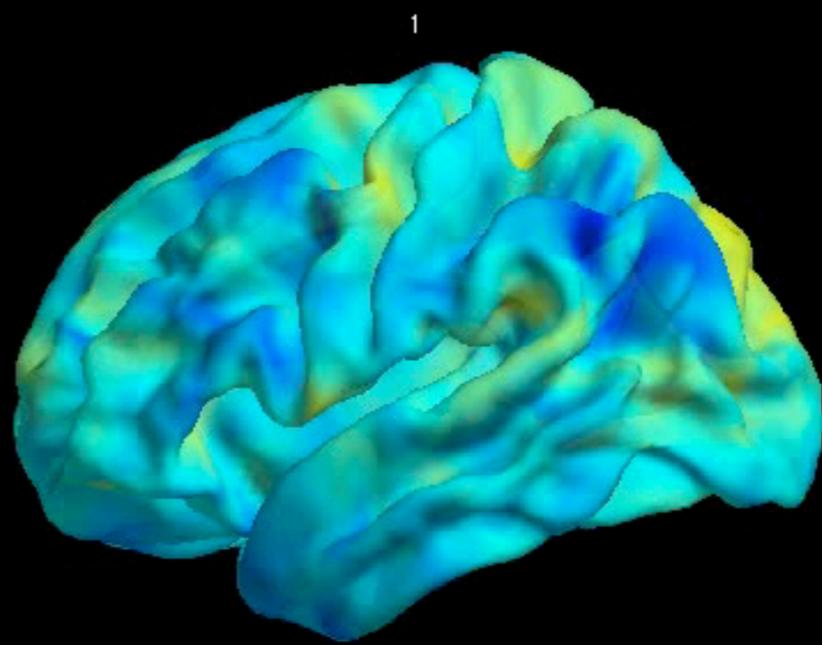
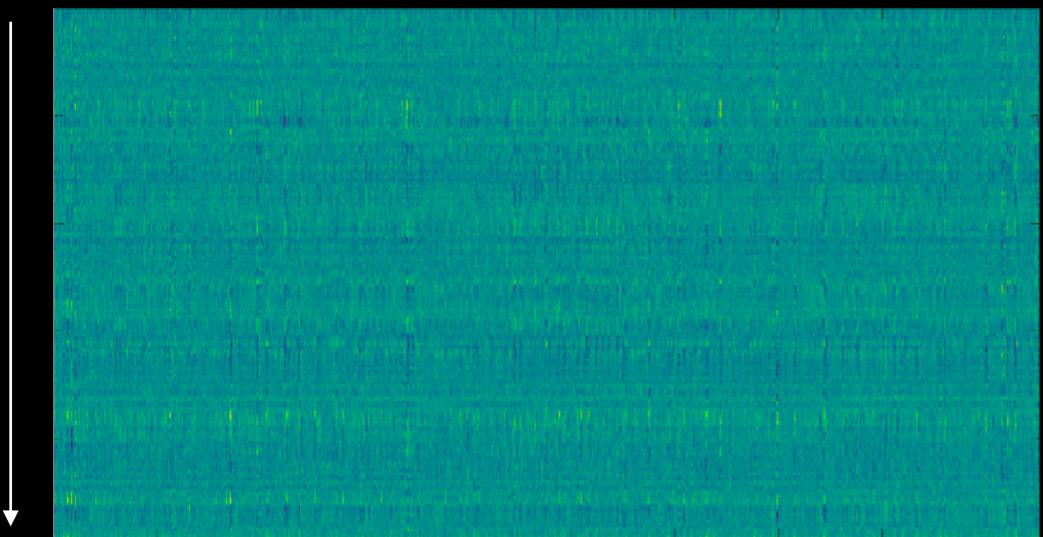
CLUSTERING TECHNIQUES

NEUR-608
BORIS BERNHARDT

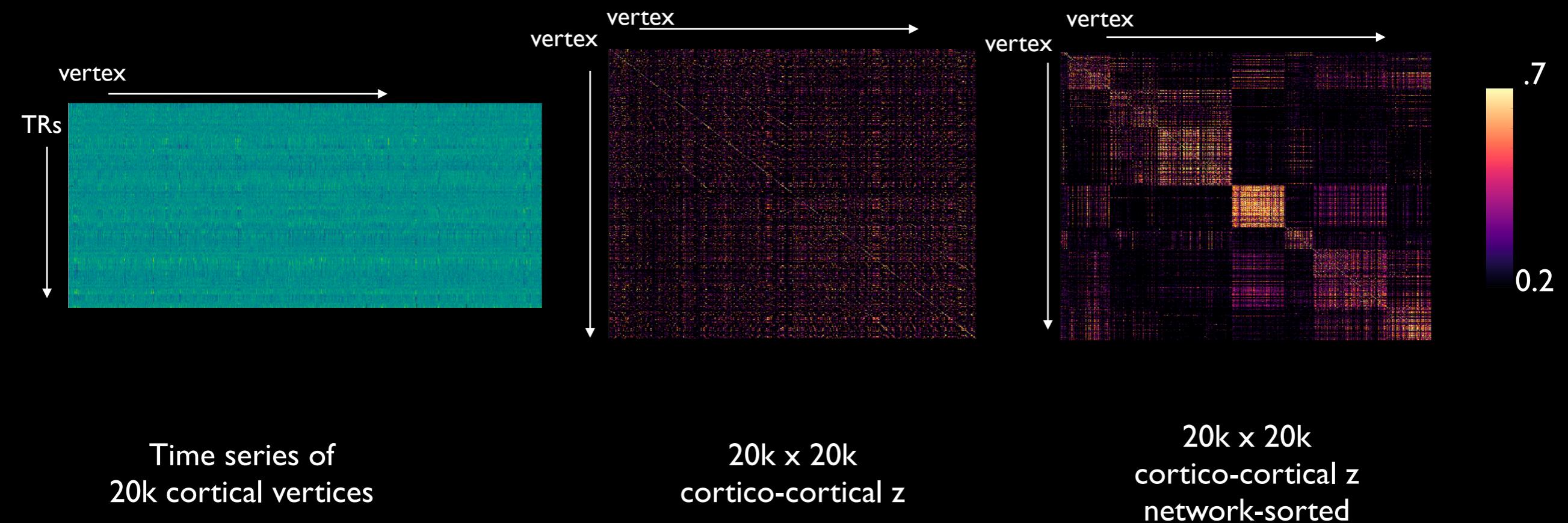


BRAIN AT REST

Space/Vertices
Time/TRs



IMAGING DATA: HIGH DIMENSIONAL AND MULTIVARIATE



HOW TO IDENTIFY NETWORKS HOW TO REDUCE DATA DIMENSIONALITY

COMPRESSION

LINEAR (PCA, ICA, FA, MDS,...)

NON-LINEAR (LE, DME,.....)

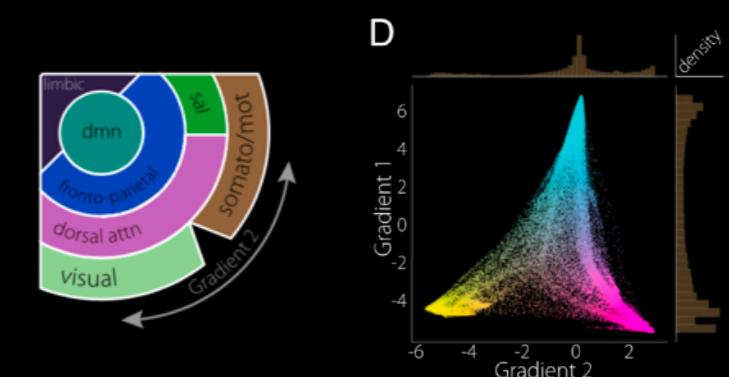
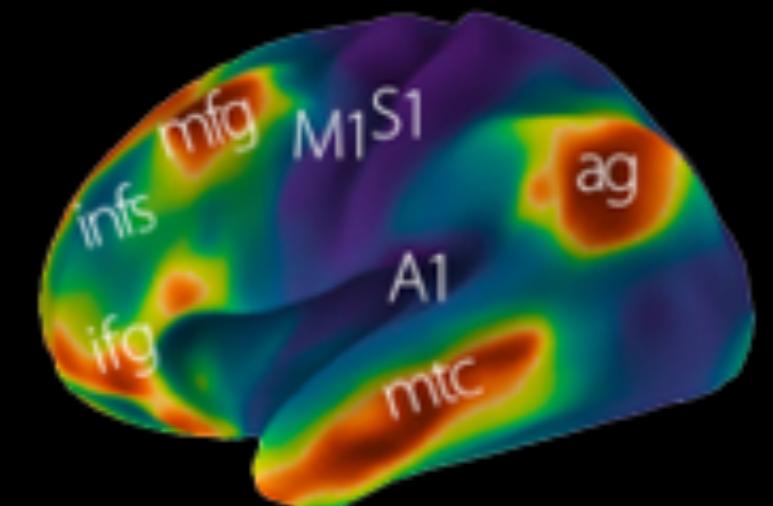
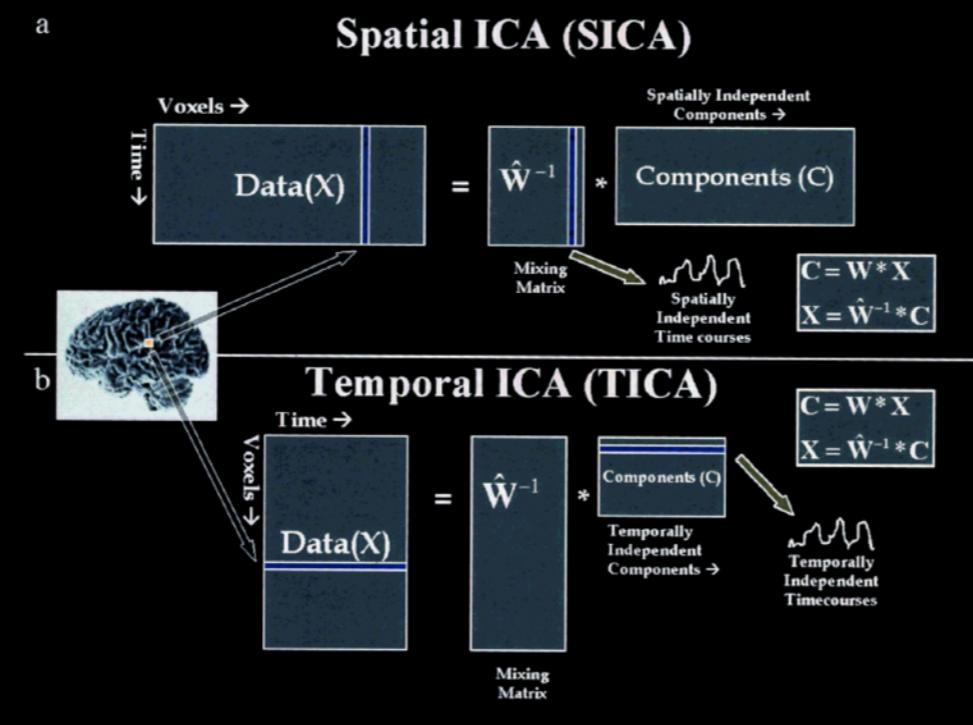
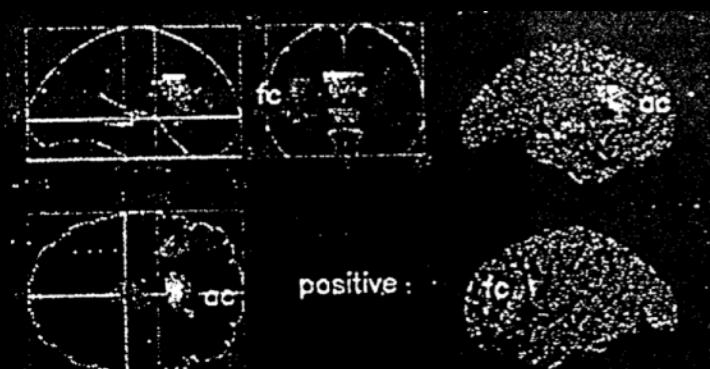
CLUSTERING

K-MEANS

HIERARCHICAL

SPECTRAL

COMPRESSION

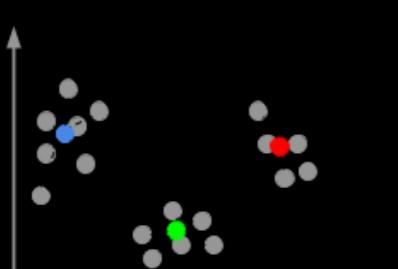
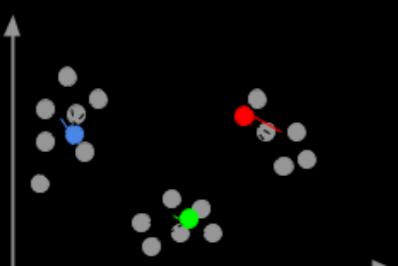
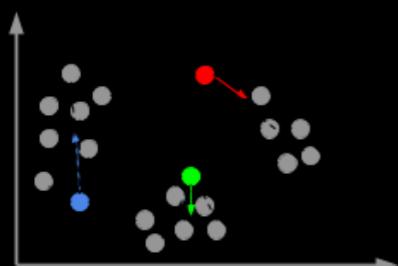
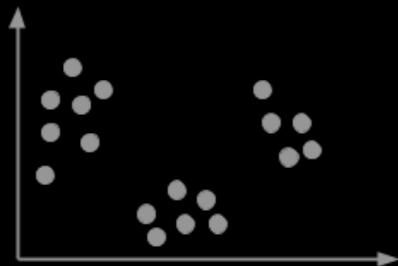


Friston 1993

McKeown et al 1998 HBM
Calhoun 2001
Beckmann 2012 NIMG

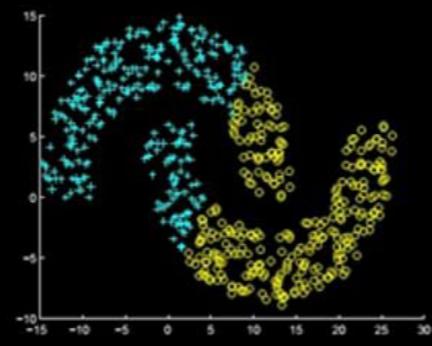
Margulies et al. 2016 PNAS

CLUSTERING ALGORITHMS

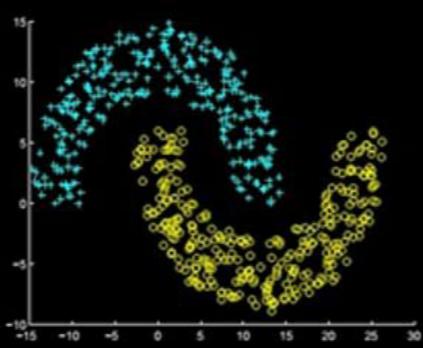


K-means clustering

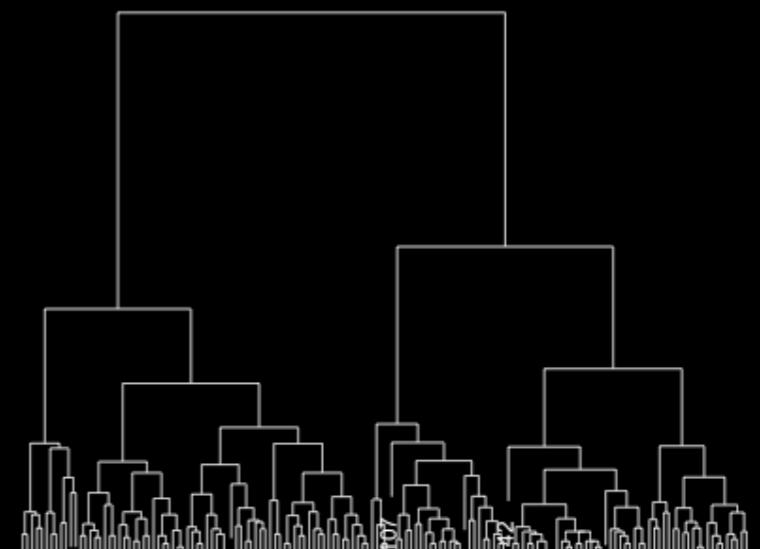
Spectral clustering
Shi Malik 2000
Von Luxburg 2007



(a) K-means

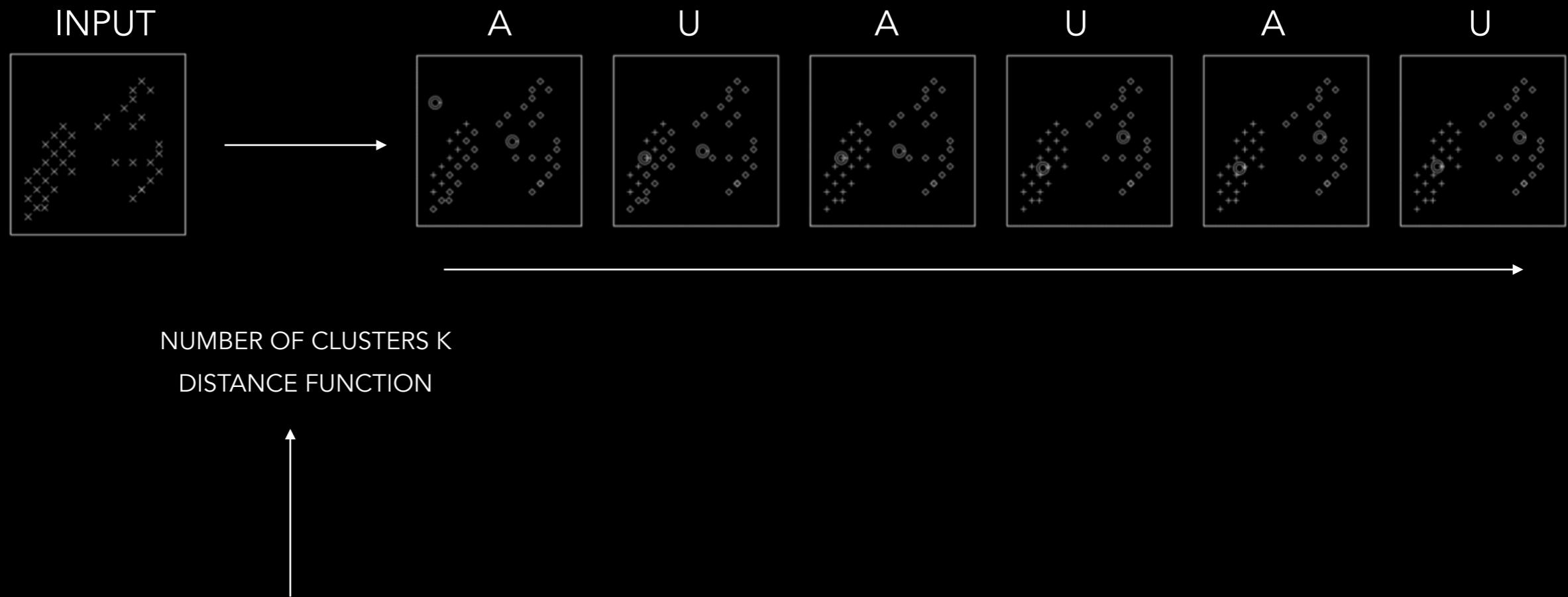


(b) Spectral Clustering



Hierarchical clustering

K-MEANS

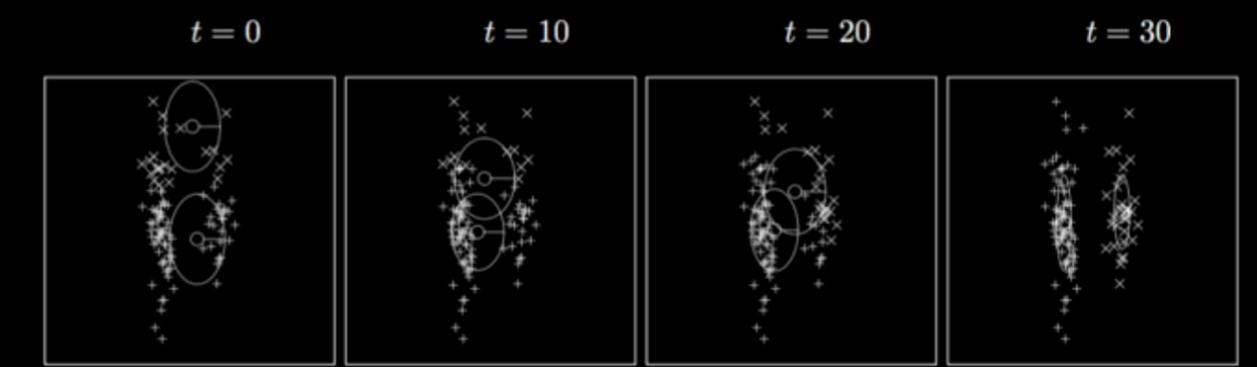
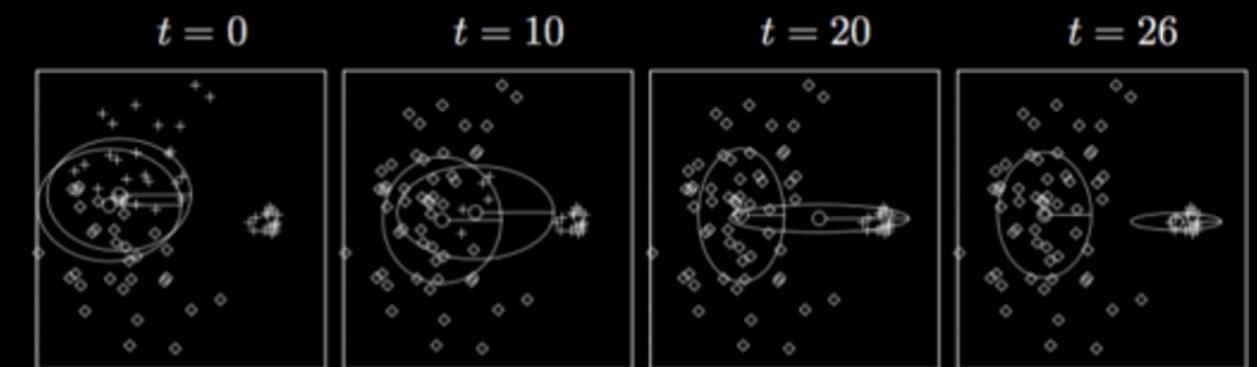
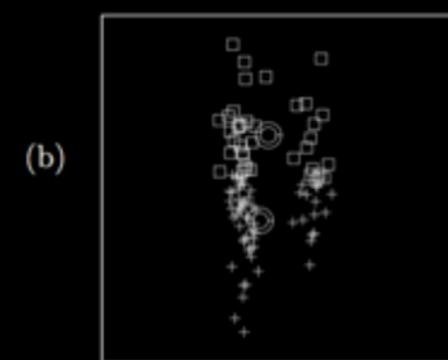
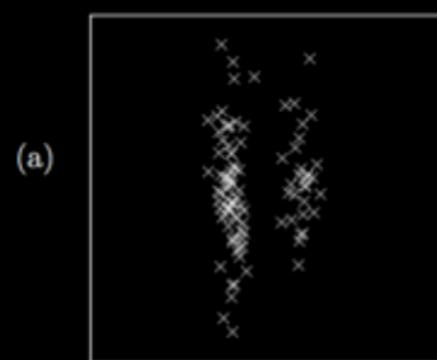
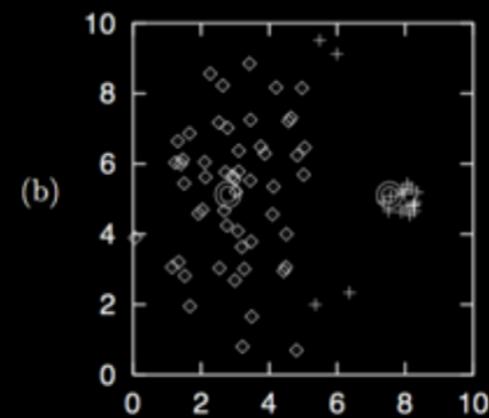
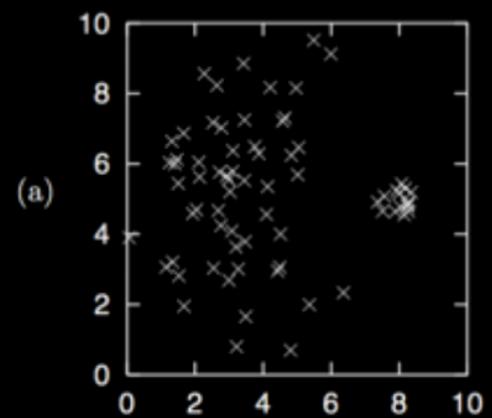


CHALLENGE I: INITIALIZATION



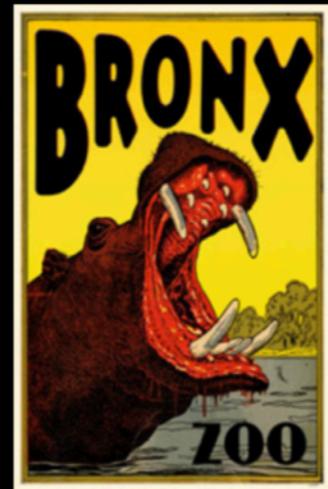
SOME SOLUTIONS: RUN MULTIPLE TIME AND IDENTIFY MOST STABLE SOLUTION; BOOTSTRAP

CHALLENGE 2: VERY DISSIMILAR CLUSTERS



SOME SOLUTIONS: SOFTEN THE ASSIGNMENT RULES

ON DISTANCE MEASURES



A ZOO OF DISTANCE MEASURES

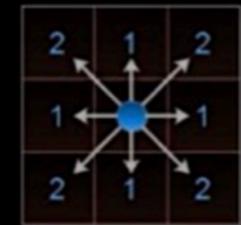
EUCLIDEAN, MANHATTAN, CHEBYCHEW
MINKOWSKI, CANBERRA, MAXIMUM, COSINE

Euclidean Distance



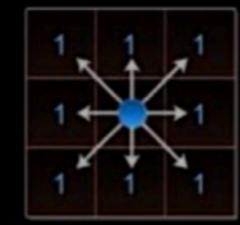
$$\sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$$

Manhattan Distance



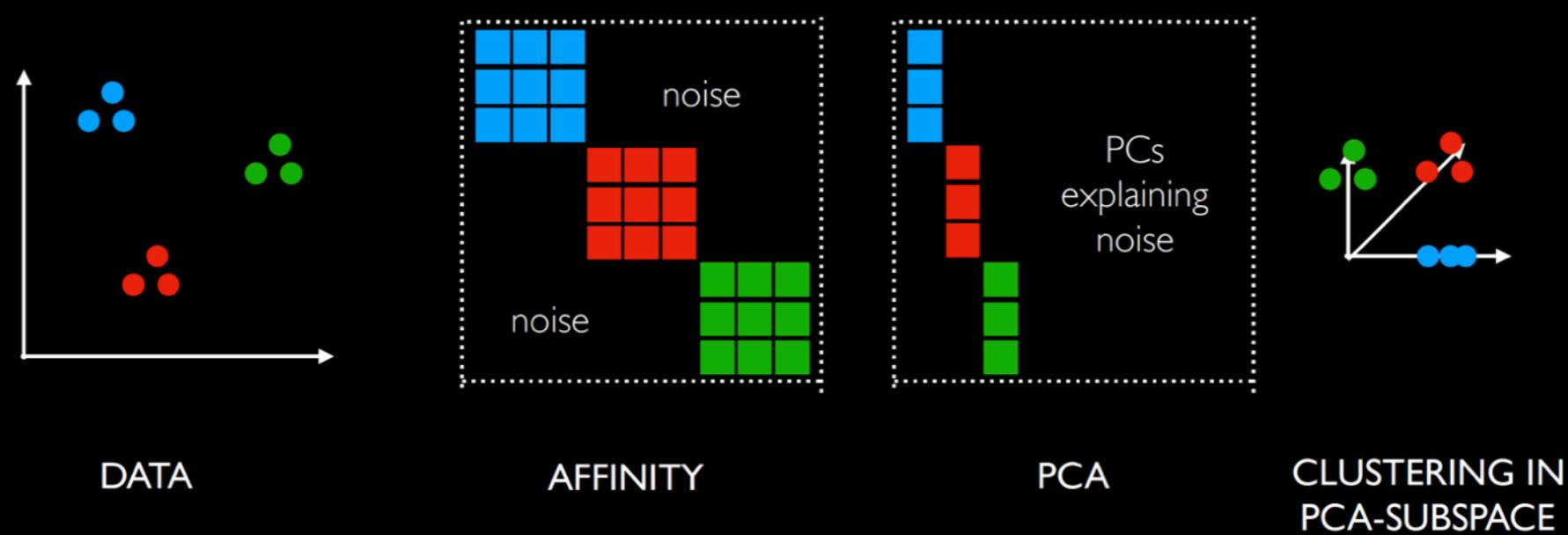
$$\|x_1 - x_2\| + |y_1 - y_2|$$

Chebyshev Distance

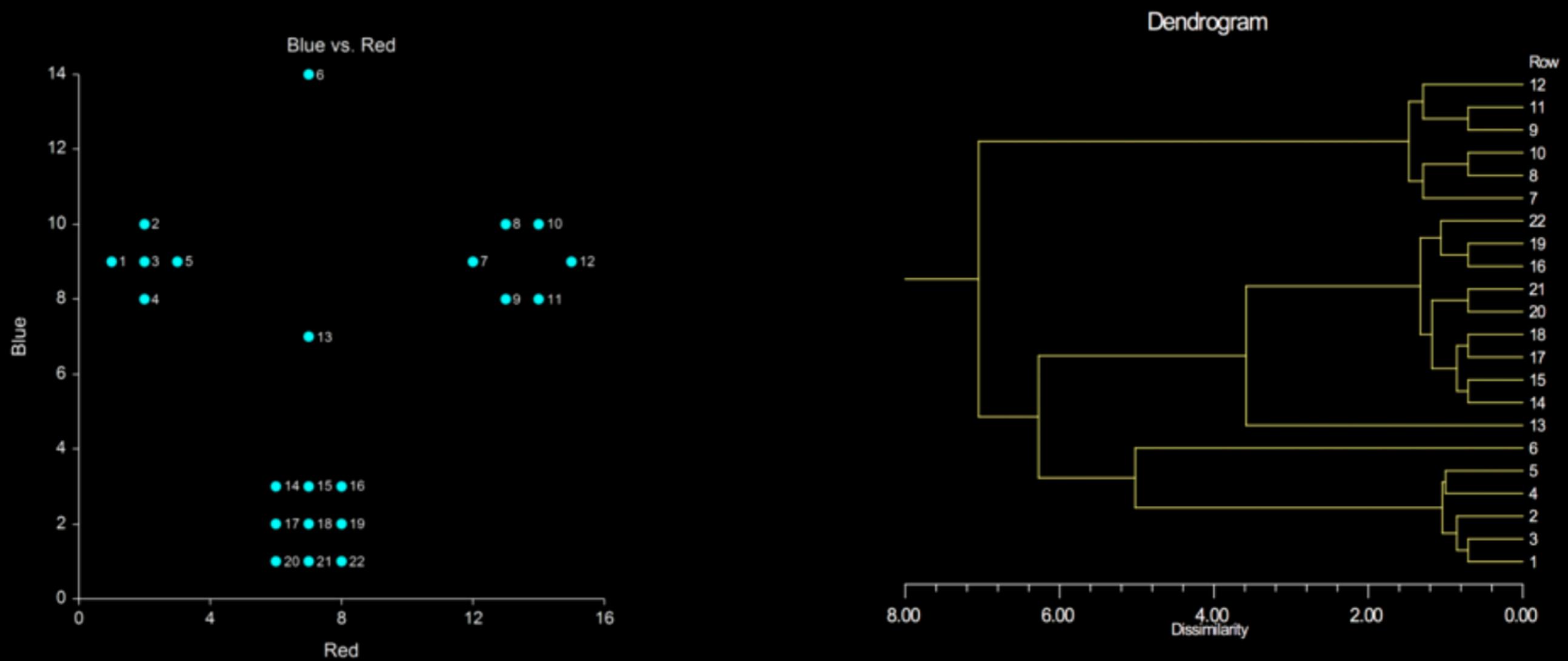


$$\max(|x_1 - x_2|, |y_1 - y_2|)$$

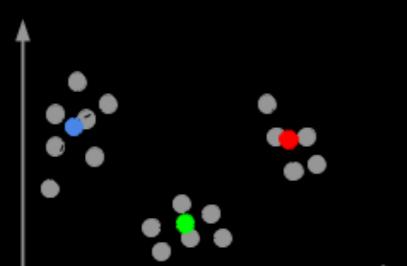
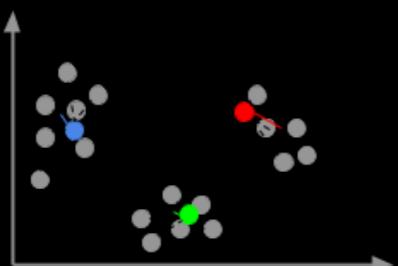
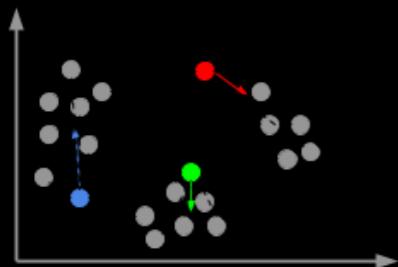
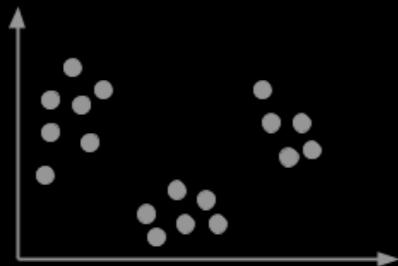
SPECTRAL CLUSTERING



HIERARCHICAL CLUSTERING

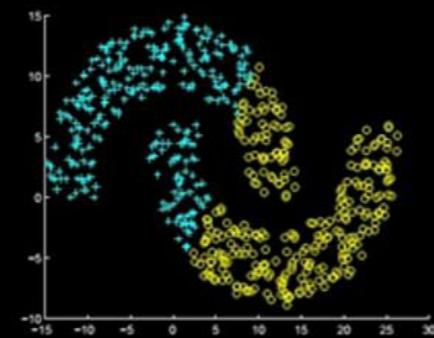


CLUSTERING ALGORITHMS

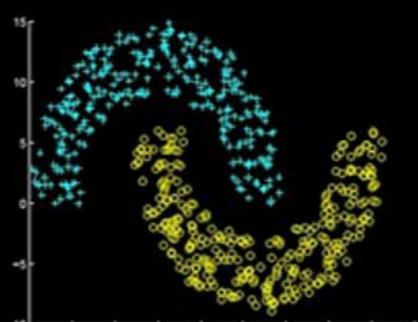


K-means clustering

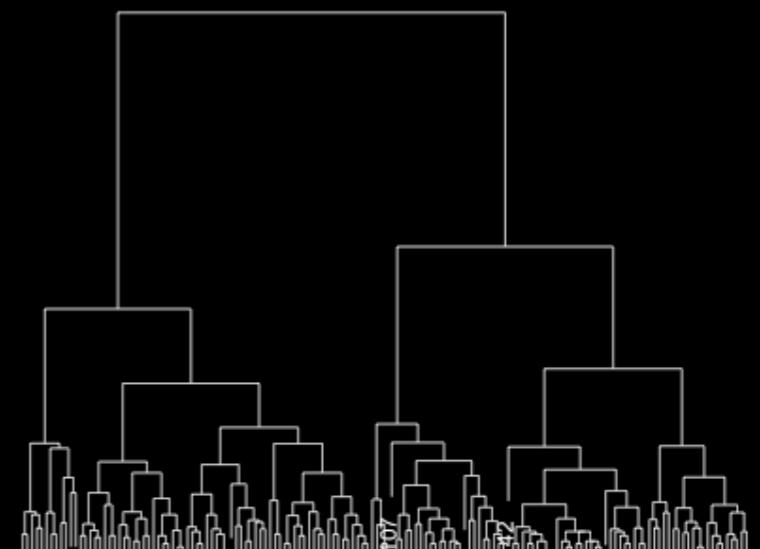
Spectral clustering
Shi Malik 2000
Von Luxburg 2007



(a) K-means

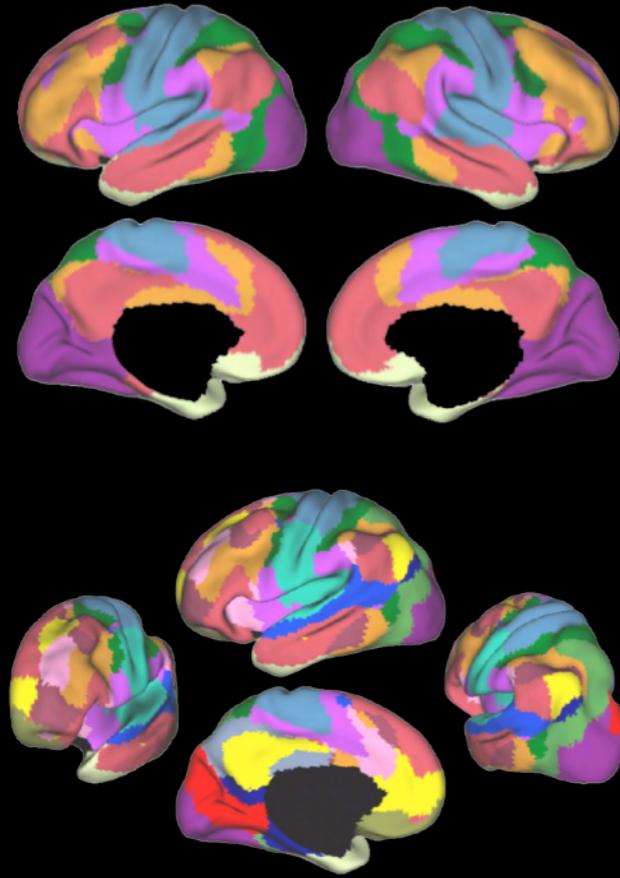
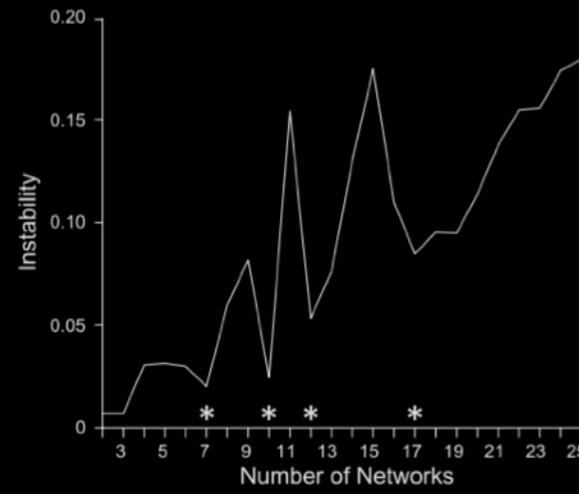


(b) Spectral Clustering

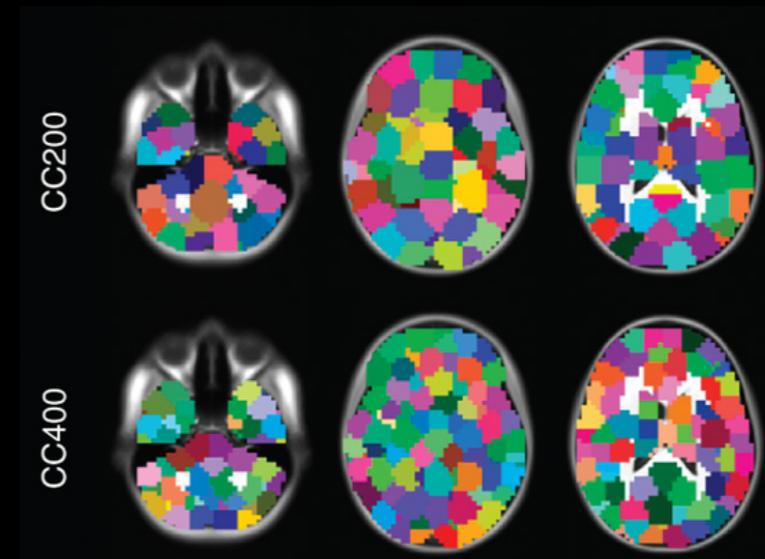


Hierarchical clustering

NETWORK CLUSTERING APPLICATIONS



$$w_{ij} = \begin{cases} s(v_i, v_j) & d_{ij} \leq \varepsilon \\ 0 & d_{ij} > \varepsilon \end{cases}.$$



MODULARITY DETECTION

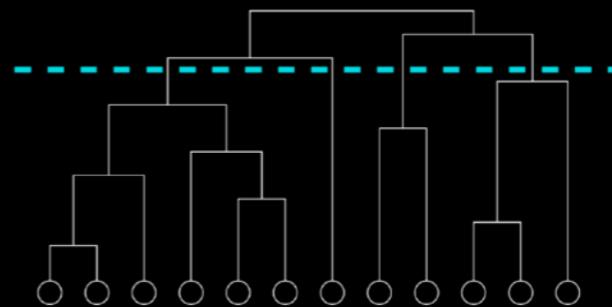
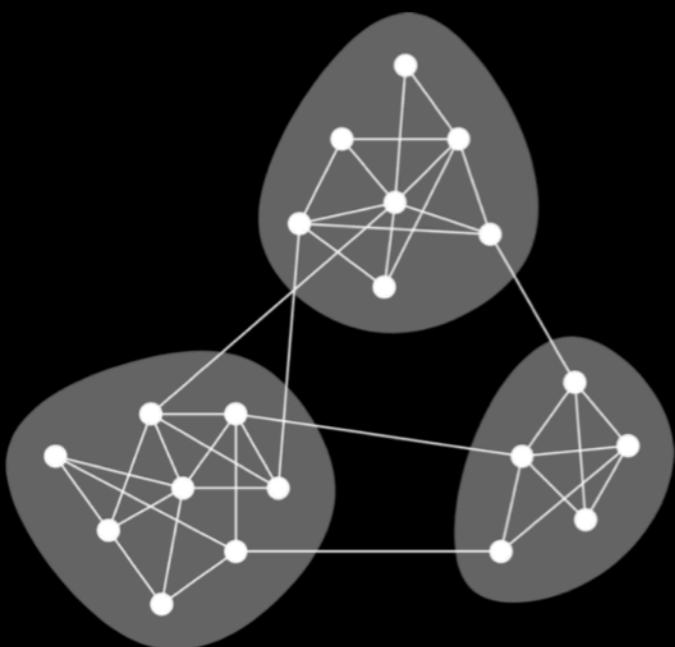
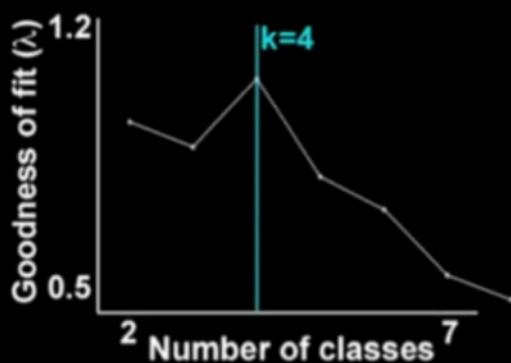


Fig. 1. The vertices in many networks fall naturally into groups or communities, sets of vertices (shaded) within which there are many edges, with only a smaller number of edges between vertices of different groups.

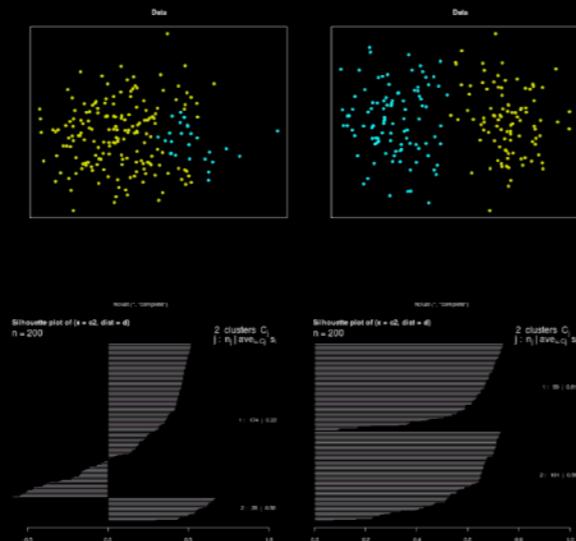
Girvan and Newman 2002, Newman 2006 PNAS, Blondel et al. 2008, Fortunato 2009

HOW MANY KS?

VARIANCE
WITHIN/BETWEEN CLUSTERS



SILHOUETTE



MULTI-CRITERIA

NbClust {NbClust}

Description
NbClust package provides 30 indices for determining the number of clusters and proposes to user the scheme from the different results obtained by varying all combinations of number of clusters, distance clustering methods.

Usage
`NbClust(data, diss = NULL, distance = "euclidean", min.nc = 2, max.nc = 15, method = "ward.D2", index = "all", alphaDeale = 0.1)`

Arguments

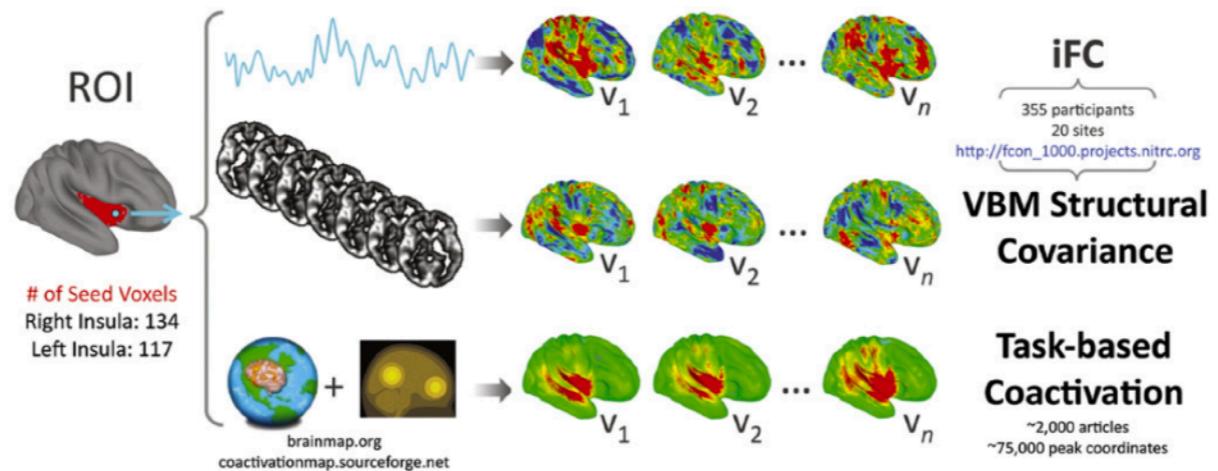
- data** matrix or dataset (the only mandatory argument)
- diss** dissimilarity matrix to be used. By default, `diss=NULL`, but if it is replaced by a dissimil distance should be "NULL".
- distance** the distance measure to be used to compute the dissimilarity matrix. This must be one o "maximum", "manhattan", "canberra", "binary", "minkowski" or "NULL". By default, d If the distance is "NULL", the dissimilarity matrix (diss) should be given by the user. If "NULL", the dissimilarity matrix should be "NULL".
- min.nc** minimal number of clusters, between 1 and (number of objects - 1)

STABILITY
REPLICABILITY

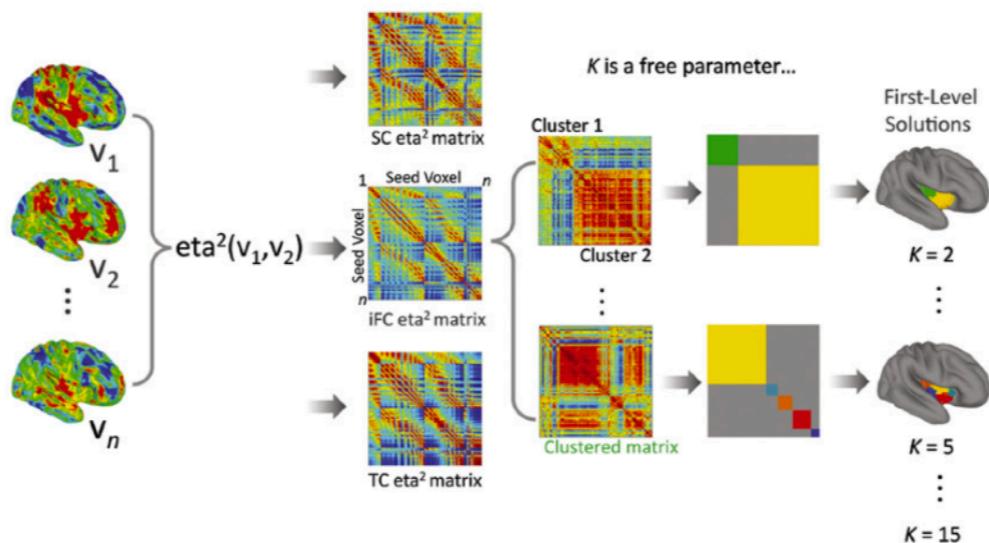


KELLY ET AL

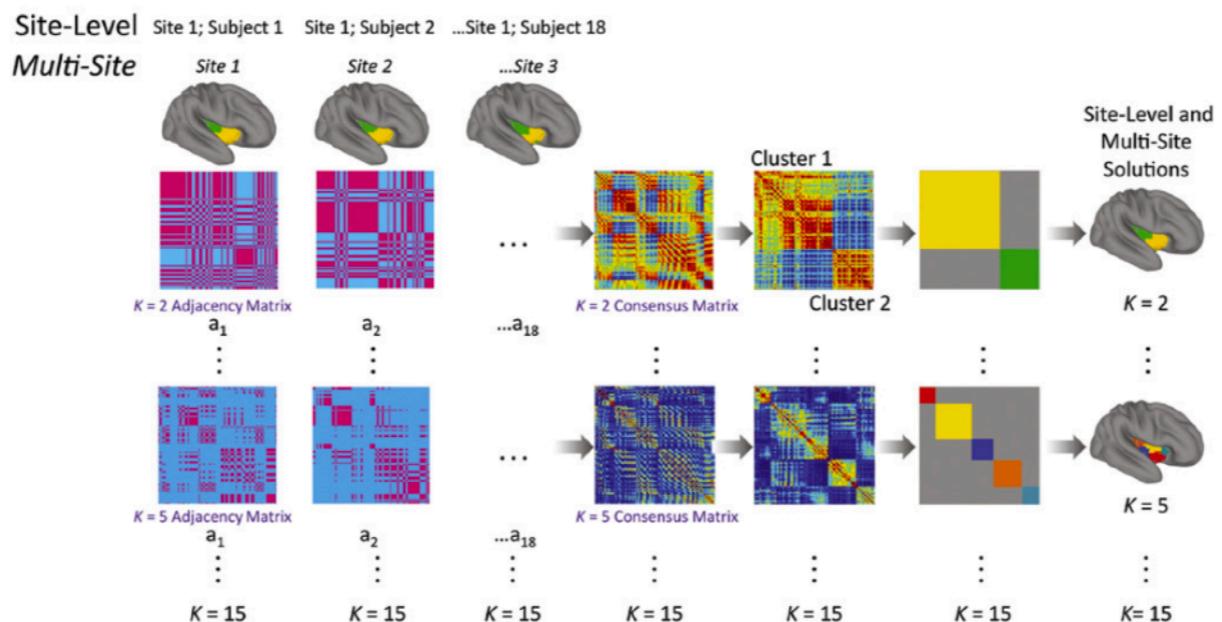
Step 1: Covariance-Based Measures



Step 2: η^2 and First-Level Clustering



Step 3: Consensus (Site-Level and Multi-Site) Clustering



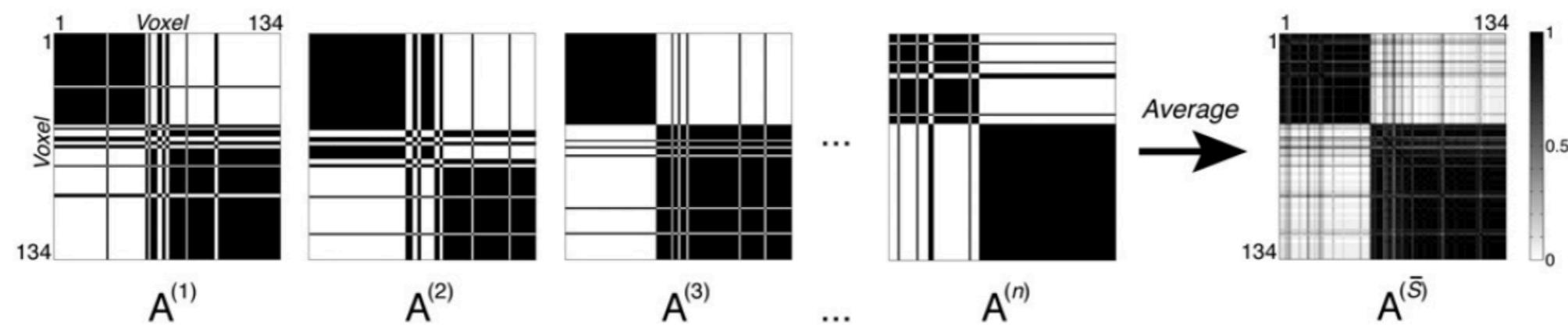
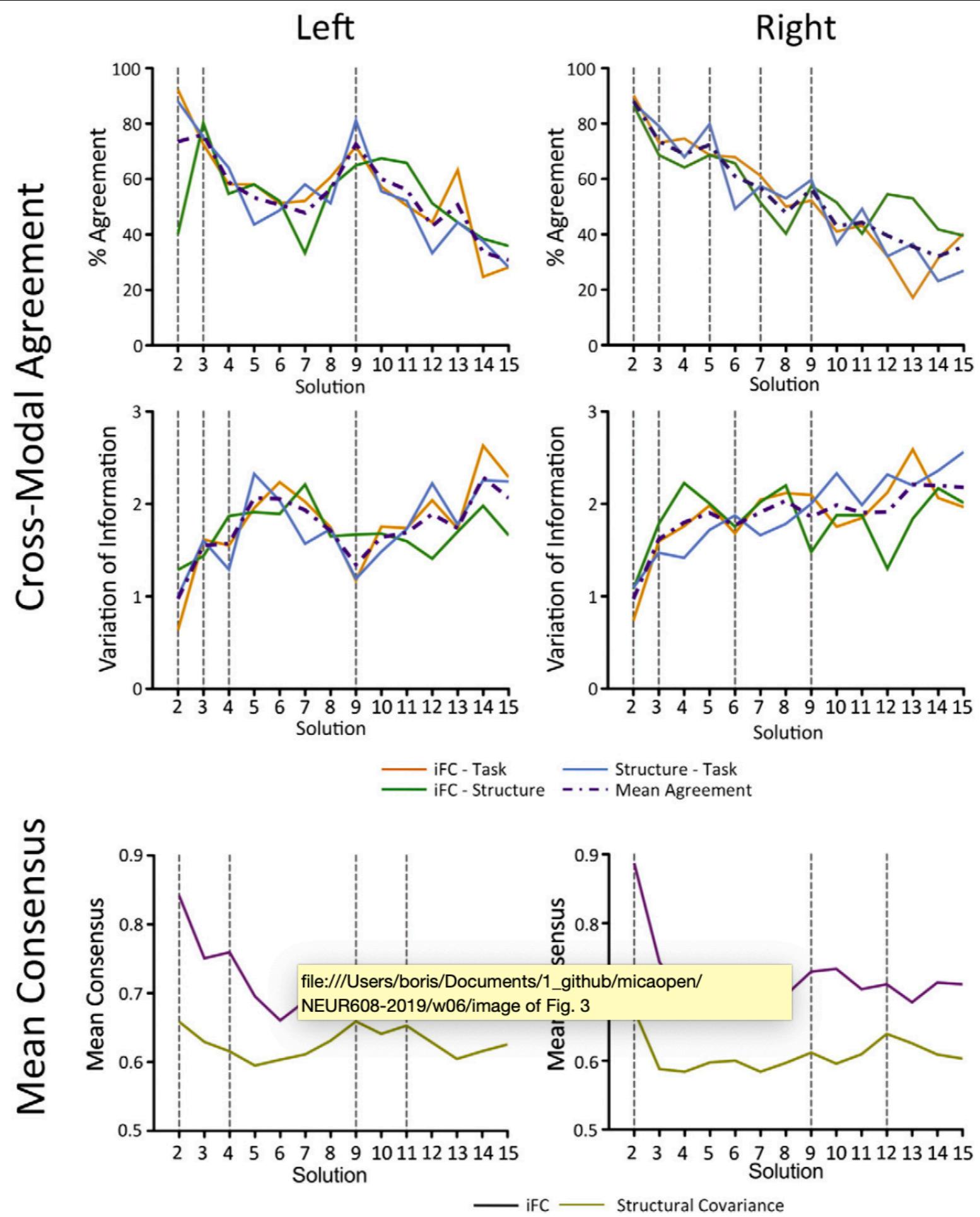
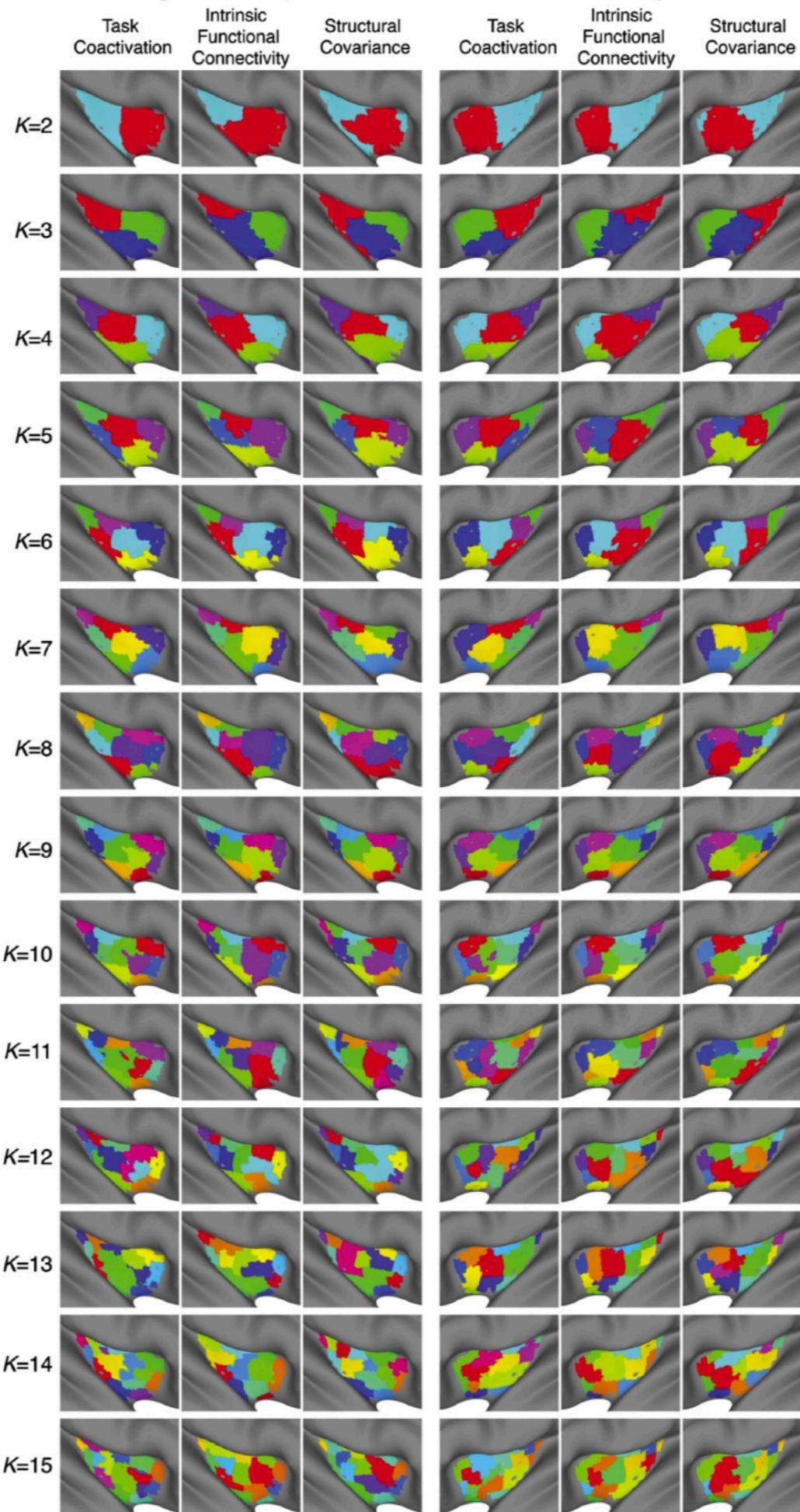


Fig. 2. Consensus clustering schematic for $K=2$. The schematic illustrates the consensus clustering process. For each scale K , each clustering instance contributes an adjacency matrix $A^{(s)}$, each element $a_{ij}^{(s)}$ of which contains a value of 1 if voxels i and j are assigned to the same cluster k , and 0 otherwise. In this example, let each instance be a data collection site, so $A^{(1)}$ is contributed by Bangor; $A^{(2)}$ is contributed by Berlin, etc. A consensus matrix $A^{(\bar{S})}$ is derived by averaging across adjacency matrices. Each element of the consensus matrix thus contains a number between 0 and 1, corresponding to the proportion of times a given pair of voxels appeared in the same cluster, across instances (here, data collection sites). The spectral clustering algorithm can then be applied to identify the most stable pattern of cluster assignments across instances, using the same scale K that was used to generate the consensus matrix (here, $K=2$).



Right Hemisphere





YEO, KRIENEN ET AL

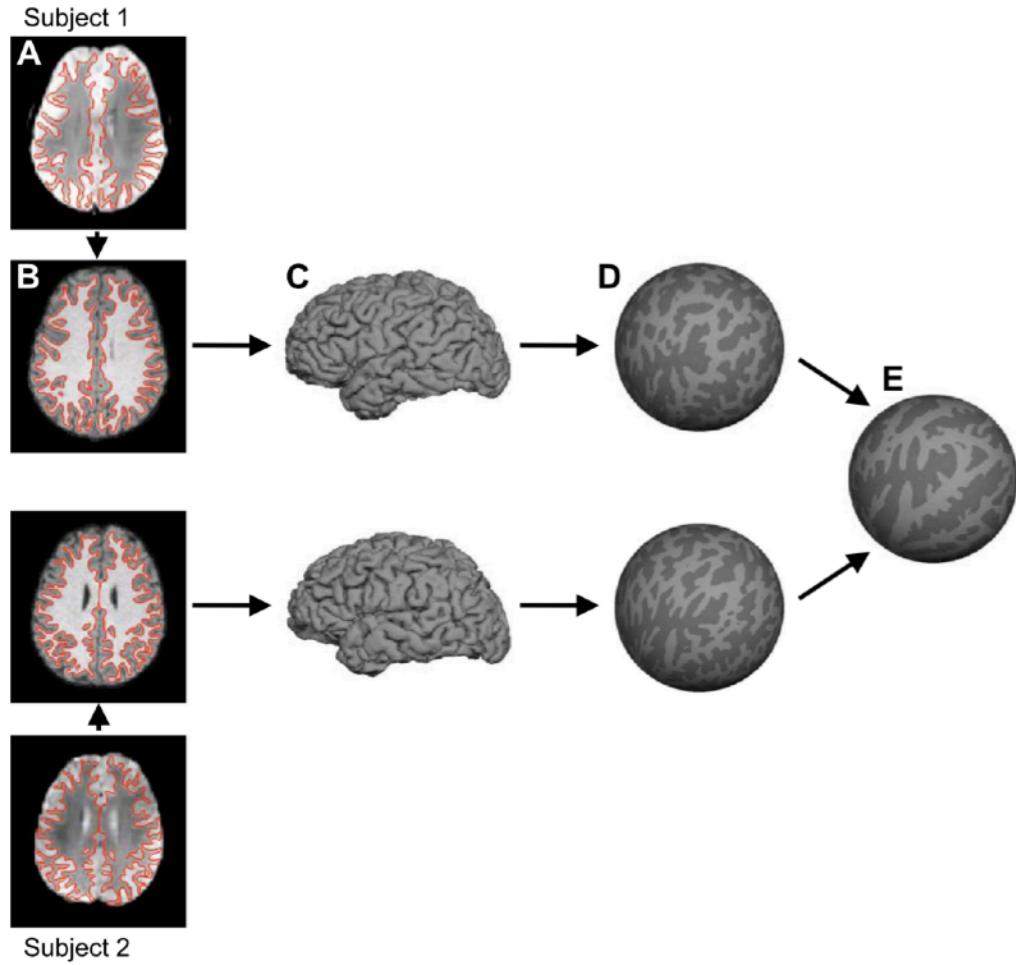


Fig. 1. Surface coordinate system for functional magnetic resonance imaging (fMRI) analysis. For each subject, the $T2^*$ images yielding blood oxygenation level-dependent (BOLD) contrast fMRI data (*A*) were registered to the $T1$ -weighted structural data (*B*). The cortical gray-white and pial surfaces were estimated from the structural data. The red lines show the estimated gray-white surface (*A* and *B*). Pial surface is shown in *C*. The gray-white surface was inflated into a sphere (*D*). The inflated spheres were then aligned across subjects using surface-based registration of the cortical folding pattern, resulting in a common spherical coordinate system (*E*). BOLD data of individual subjects (*A*) can then be projected onto the spherical coordinate system (*E*) in a single transformation step to reduce artifacts due to multiple interpolations.

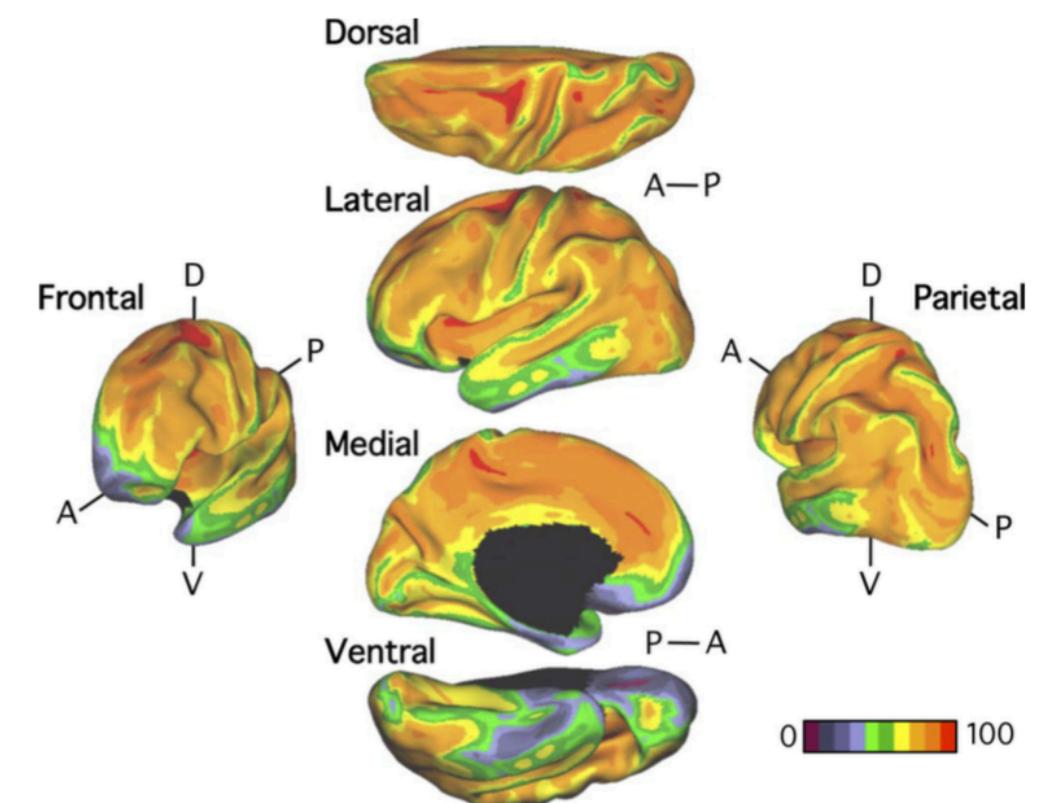
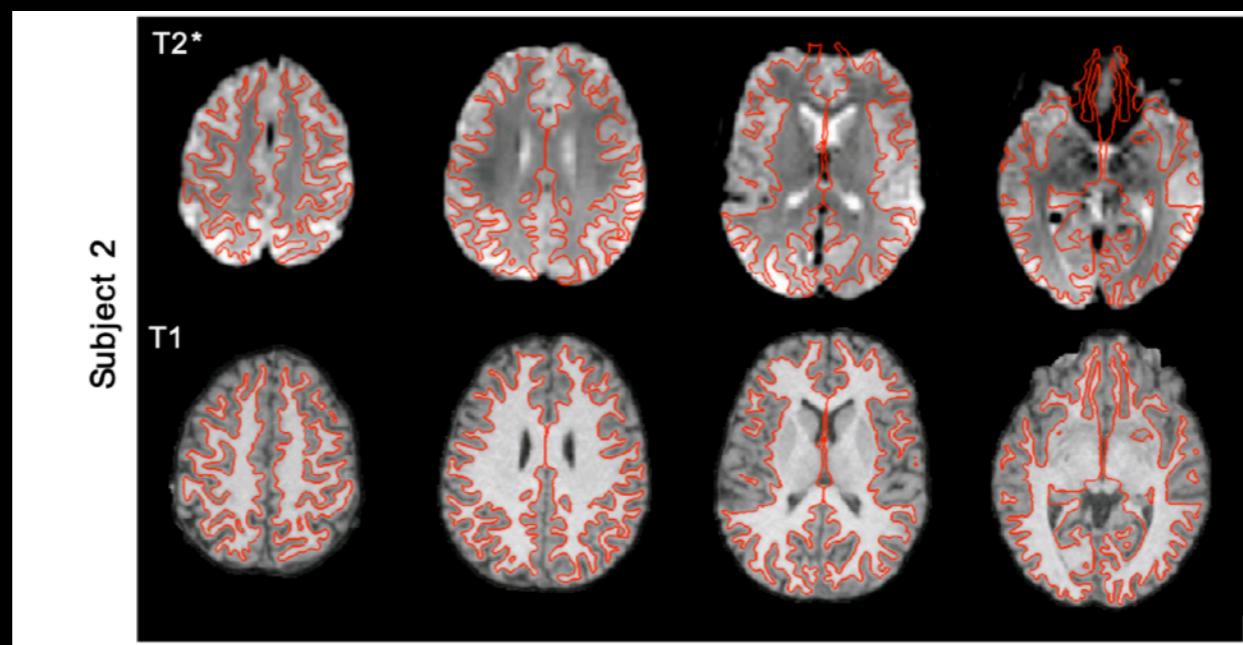
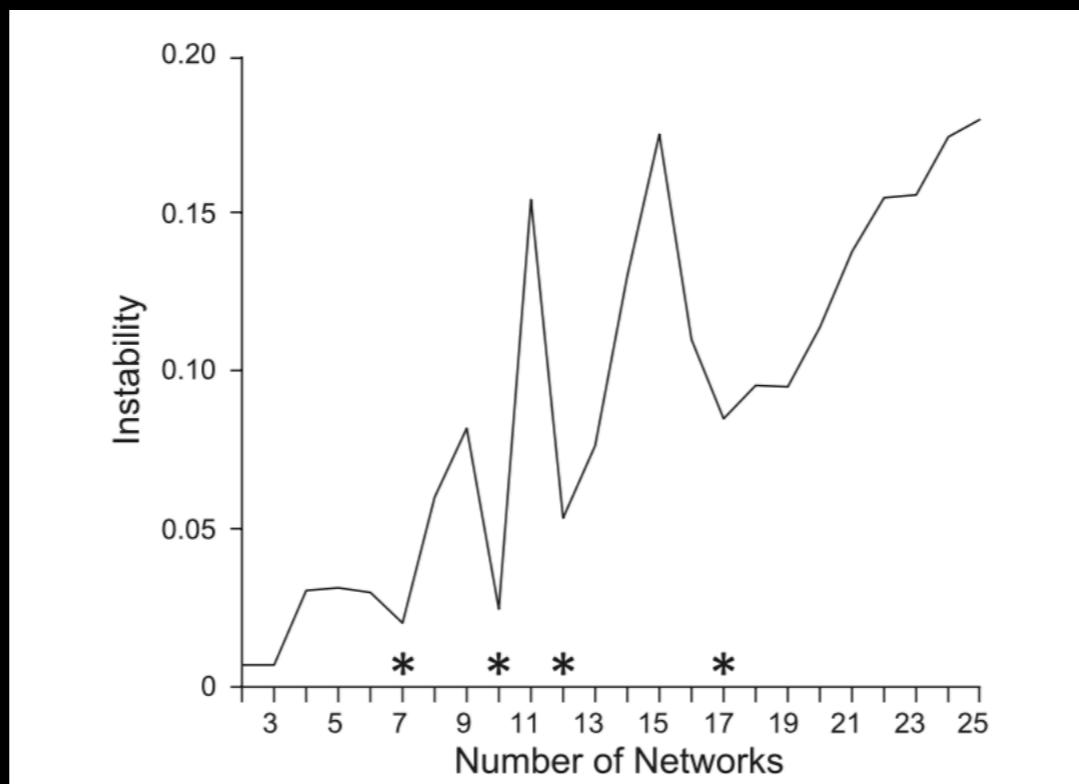
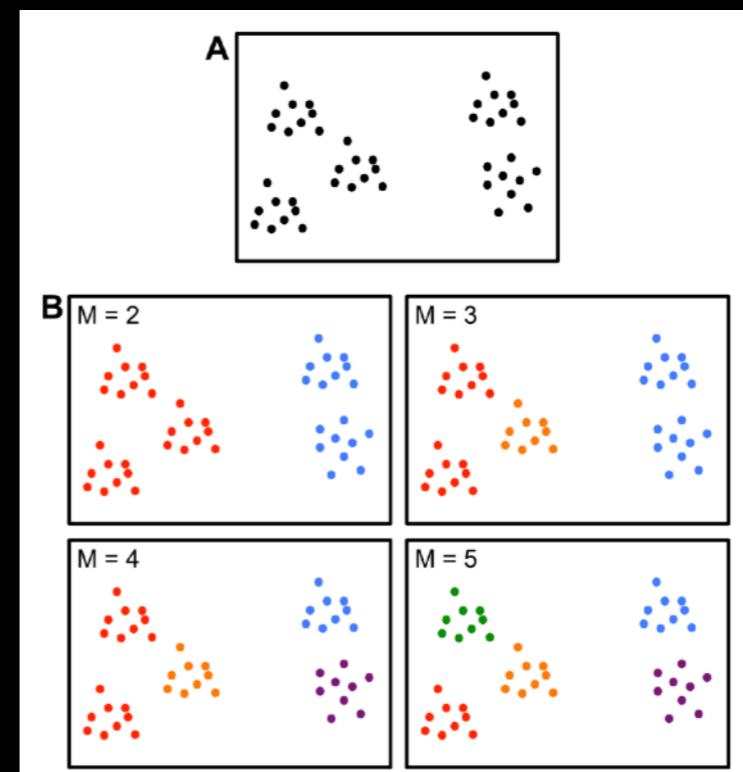
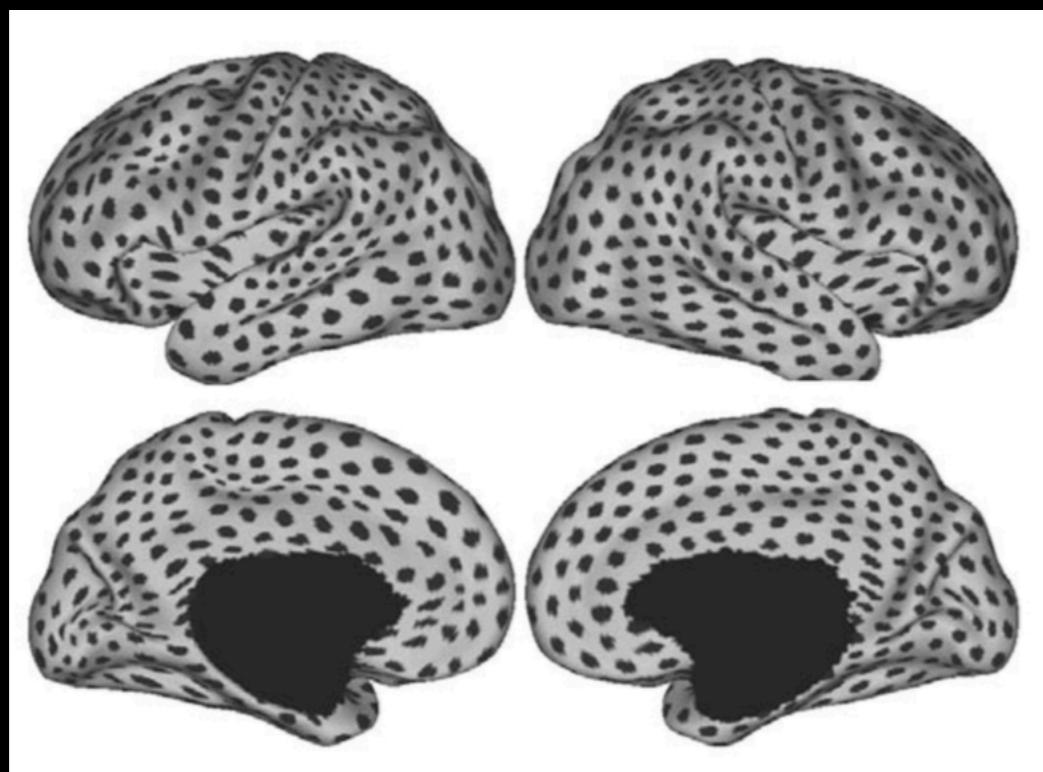


Fig. 3. Signal-to-noise ratio (SNR) maps of the functional data from the full sample ($N = 1,000$). The mean estimate of the BOLD fMRI data SNR is illustrated for multiple views of the left hemisphere in Caret PALS space. A, anterior; P, posterior; D, dorsal; V, ventral.





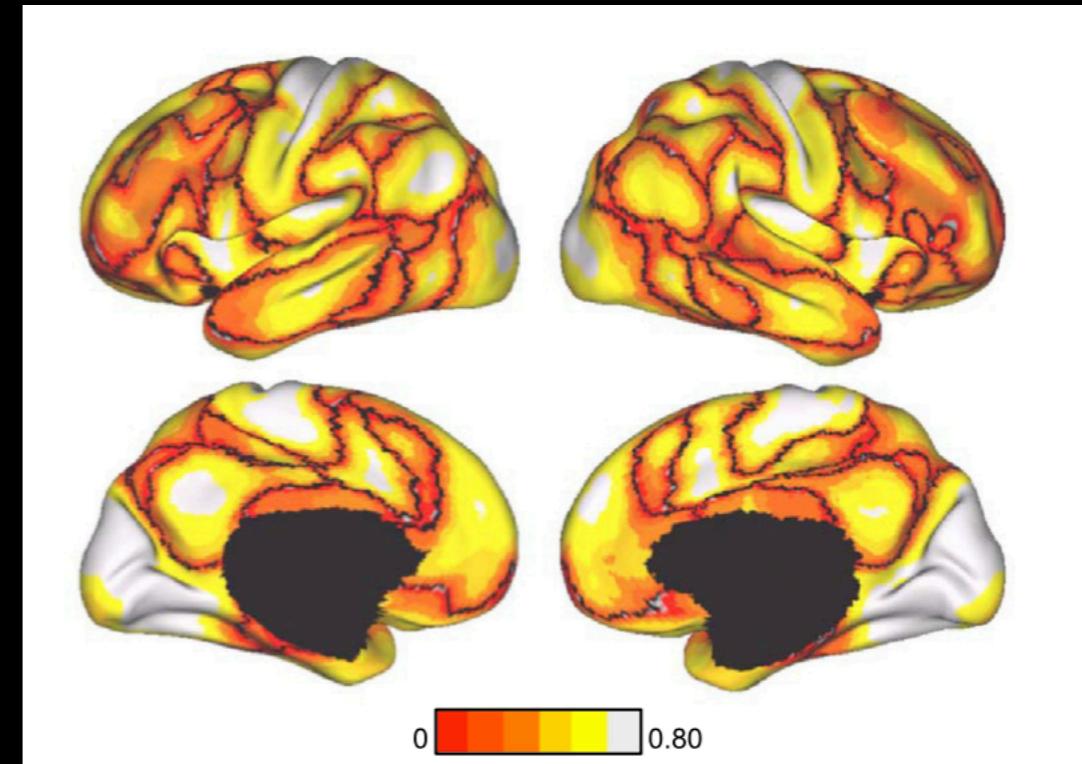
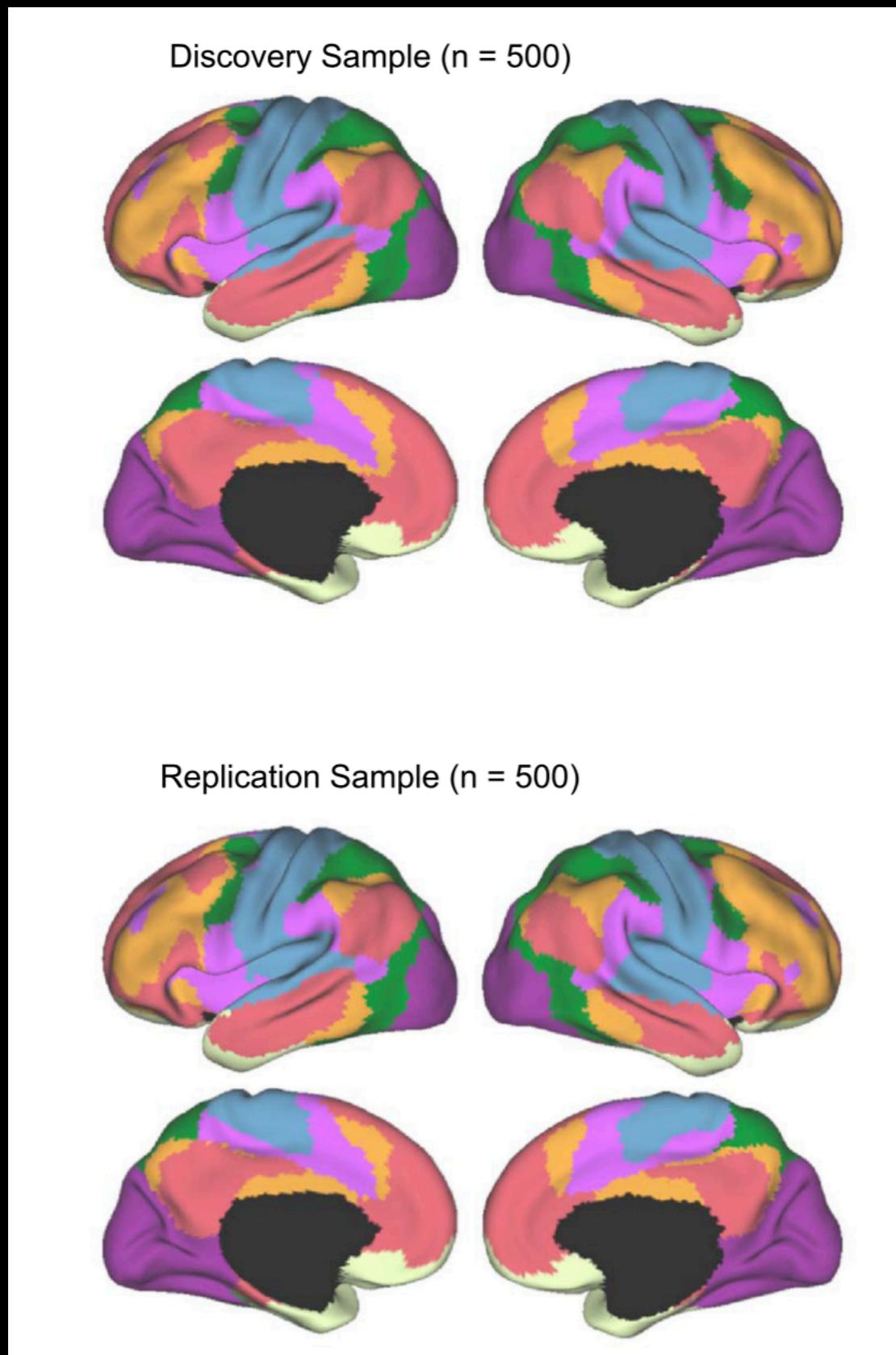
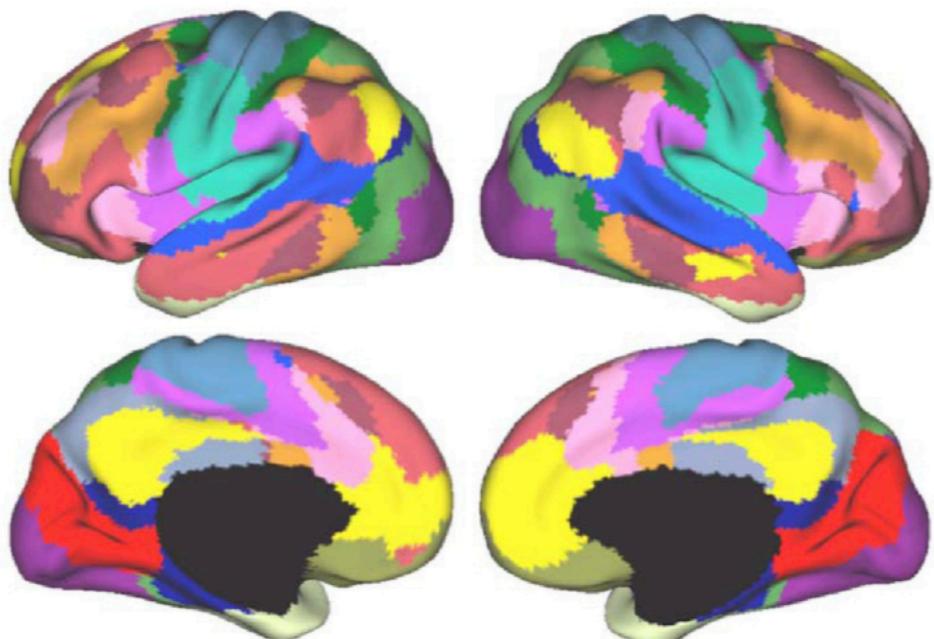


Fig. 8. Confidence of the 7-network estimate in the discovery data set. Confidence (silhouette) value for each vertex with respect to its assigned network is shown for the discovery data set. Regions close to the boundaries between networks were less confident of their assignment, although we also observed structured spatial variation within individual components of the estimated networks, such as lateral prefrontal cortex, which foreshadows its division in the 17-network estimate (see Fig. 9).

Discovery Sample (n = 500)



Replication Sample (n = 500)

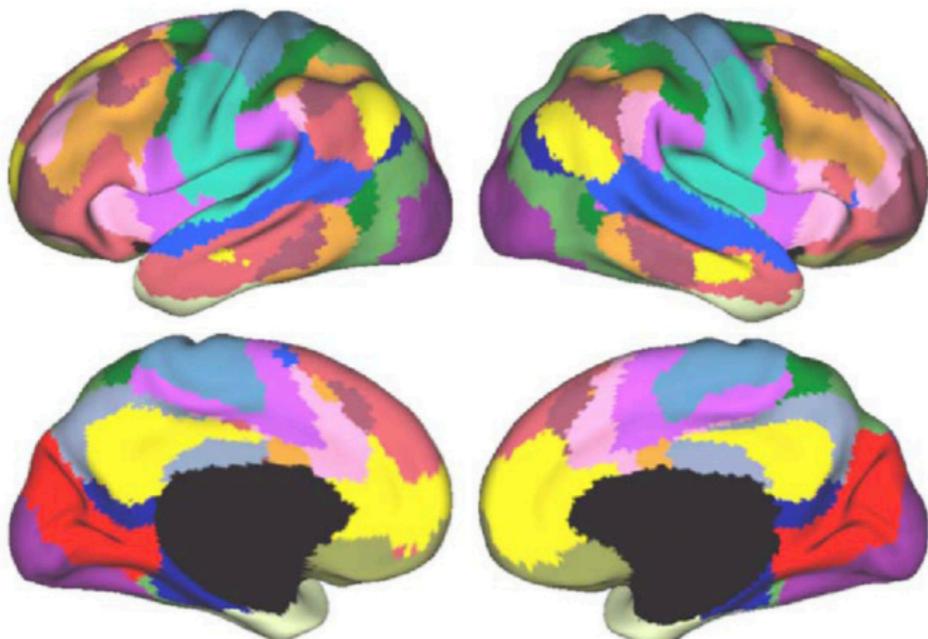
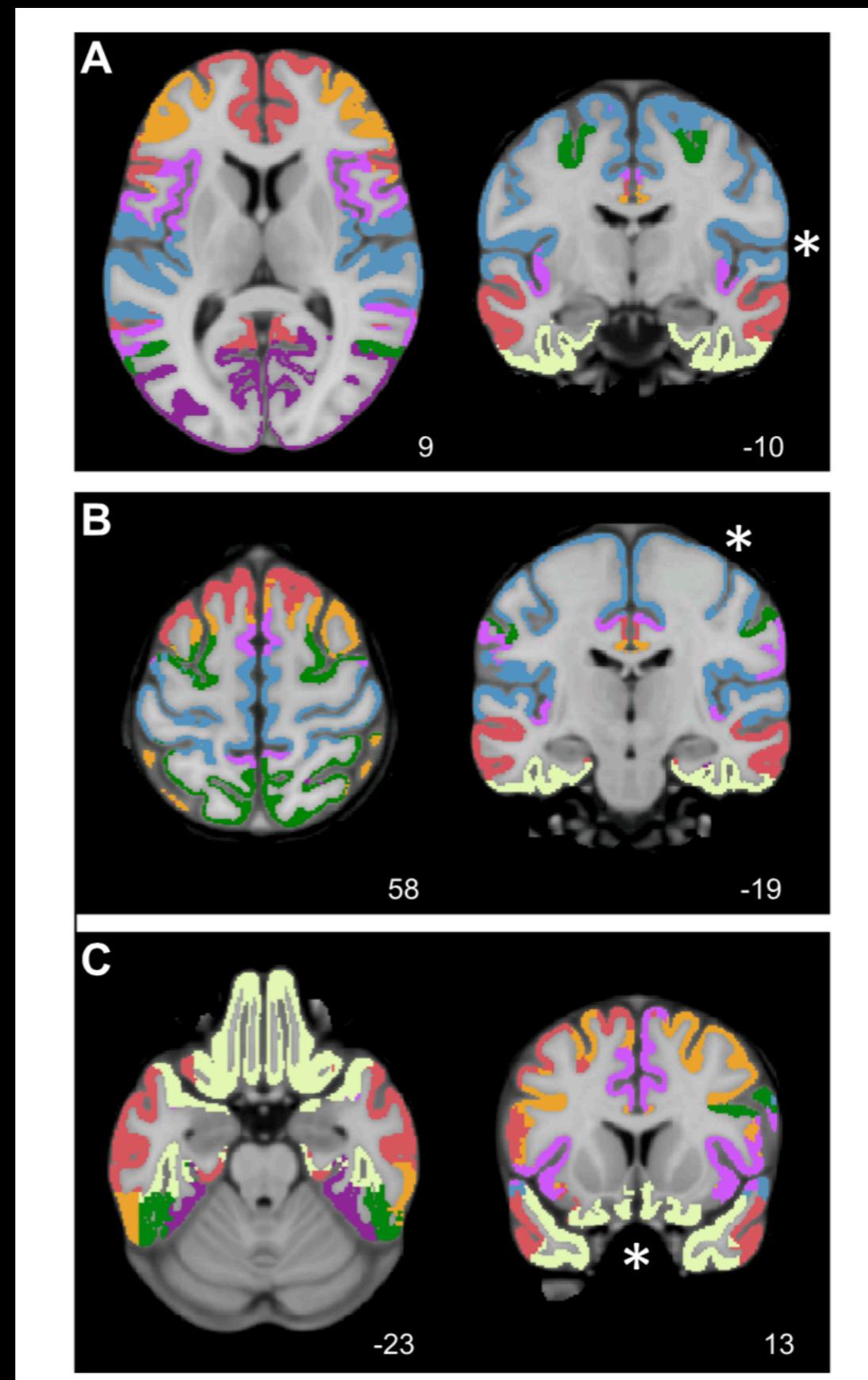


Fig. 10. Confidence of 17-network estimate in the discovery data set. Confidence (silhouette) value for each vertex with respect to its assigned network is shown for the discovery data set. Again, regions close to the boundaries between networks were less confident of their assignment.



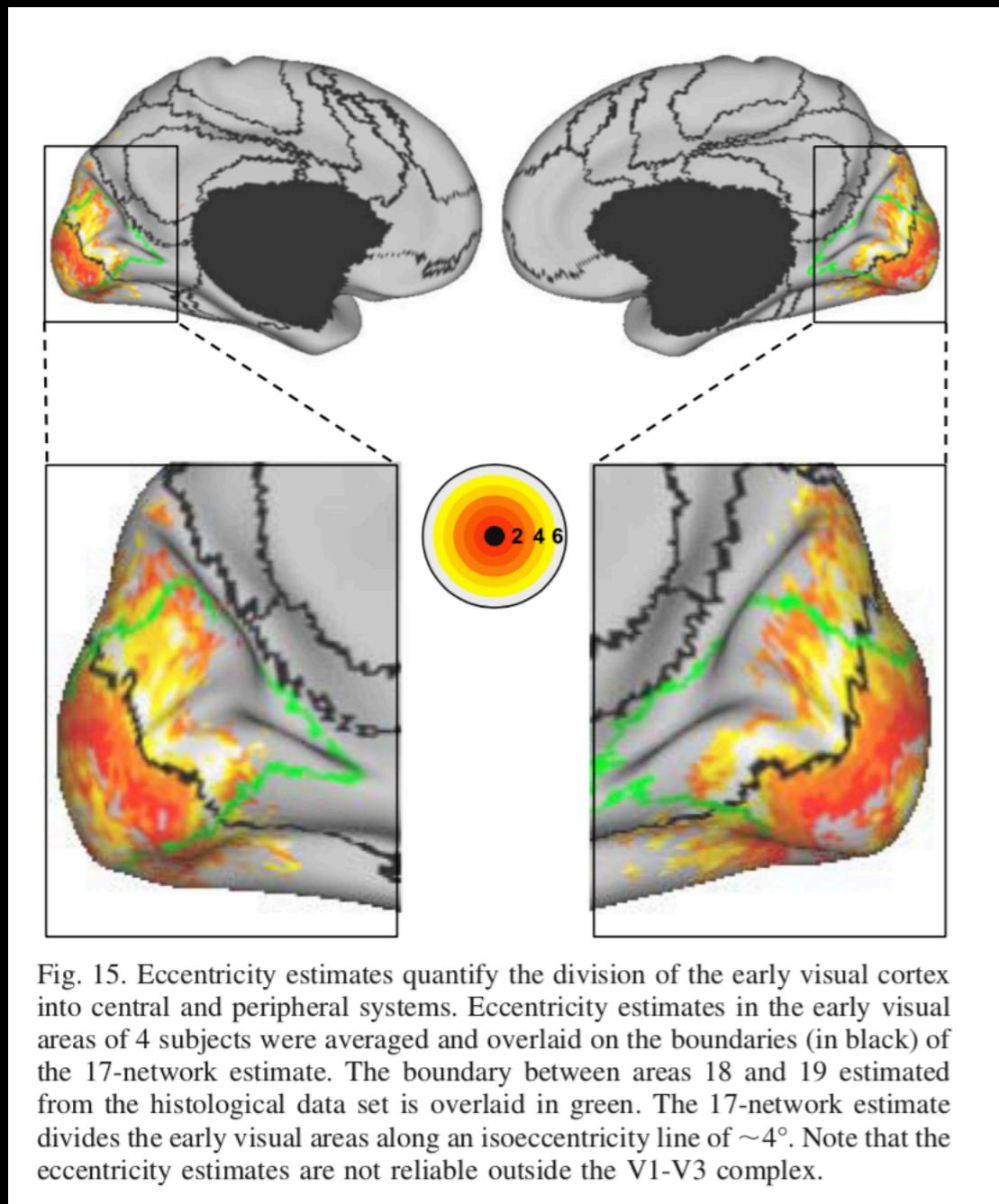


Fig. 15. Eccentricity estimates quantify the division of the early visual cortex into central and peripheral systems. Eccentricity estimates in the early visual areas of 4 subjects were averaged and overlaid on the boundaries (in black) of the 17-network estimate. The boundary between areas 18 and 19 estimated from the histological data set is overlaid in green. The 17-network estimate divides the early visual areas along an isoeccentricity line of $\sim 4^\circ$. Note that the eccentricity estimates are not reliable outside the V1-V3 complex.

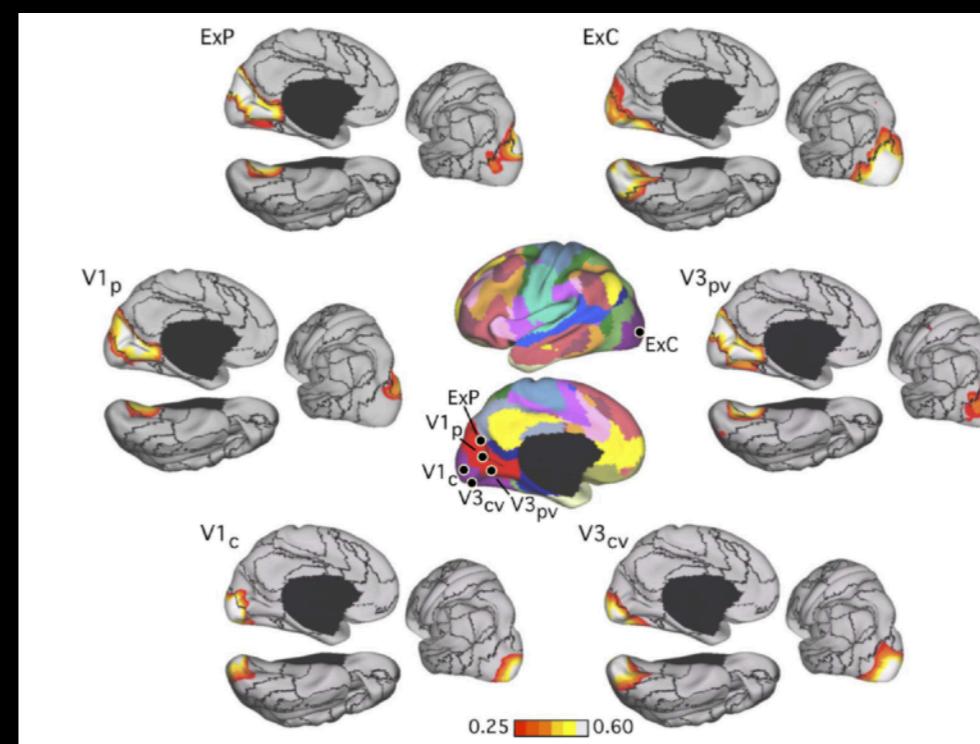


Fig. 16. Evidence that the fractionation of the visual system reflects functional connectivity MRI (fcMRI) topography within the visual cortex. Six left hemisphere seed regions were picked from the discovery dataset: V1_c and V1_p correspond to central and peripheral visual field representation within V1, respectively; V3_{cv} and V3_{pv} correspond to central and peripheral V3v, respectively; ExC and ExP correspond to 2 seed regions within the extrastriate visual cortex in the estimated locations of the central and peripheral visual fields (purple and bright red at center). The 6 seed regions are illustrated at center, and their coordinate locations are reported in Table 1. Their left hemisphere fcMRI maps were computed using the replication data set and arranged around the center images. Note that the central visual seed regions are selectively correlated with the central visual representation, whereas the peripheral visual seed regions are selectively correlated with the peripheral visual representation.

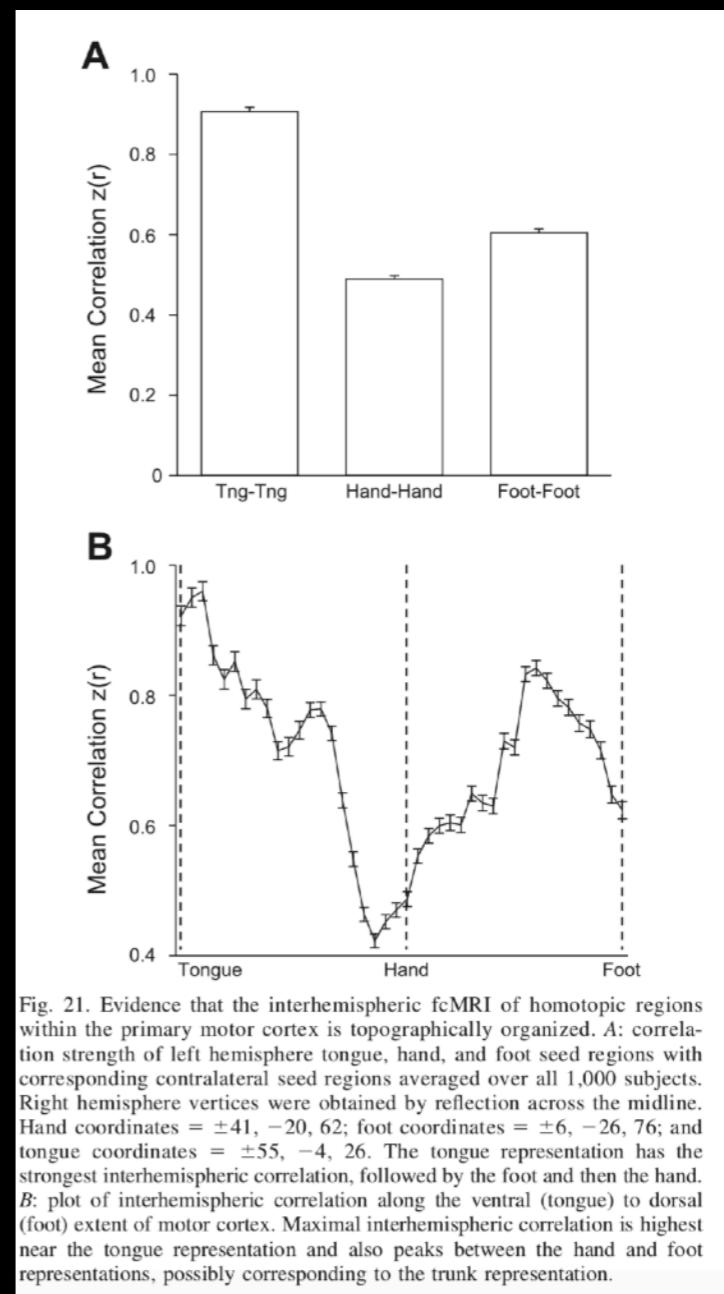
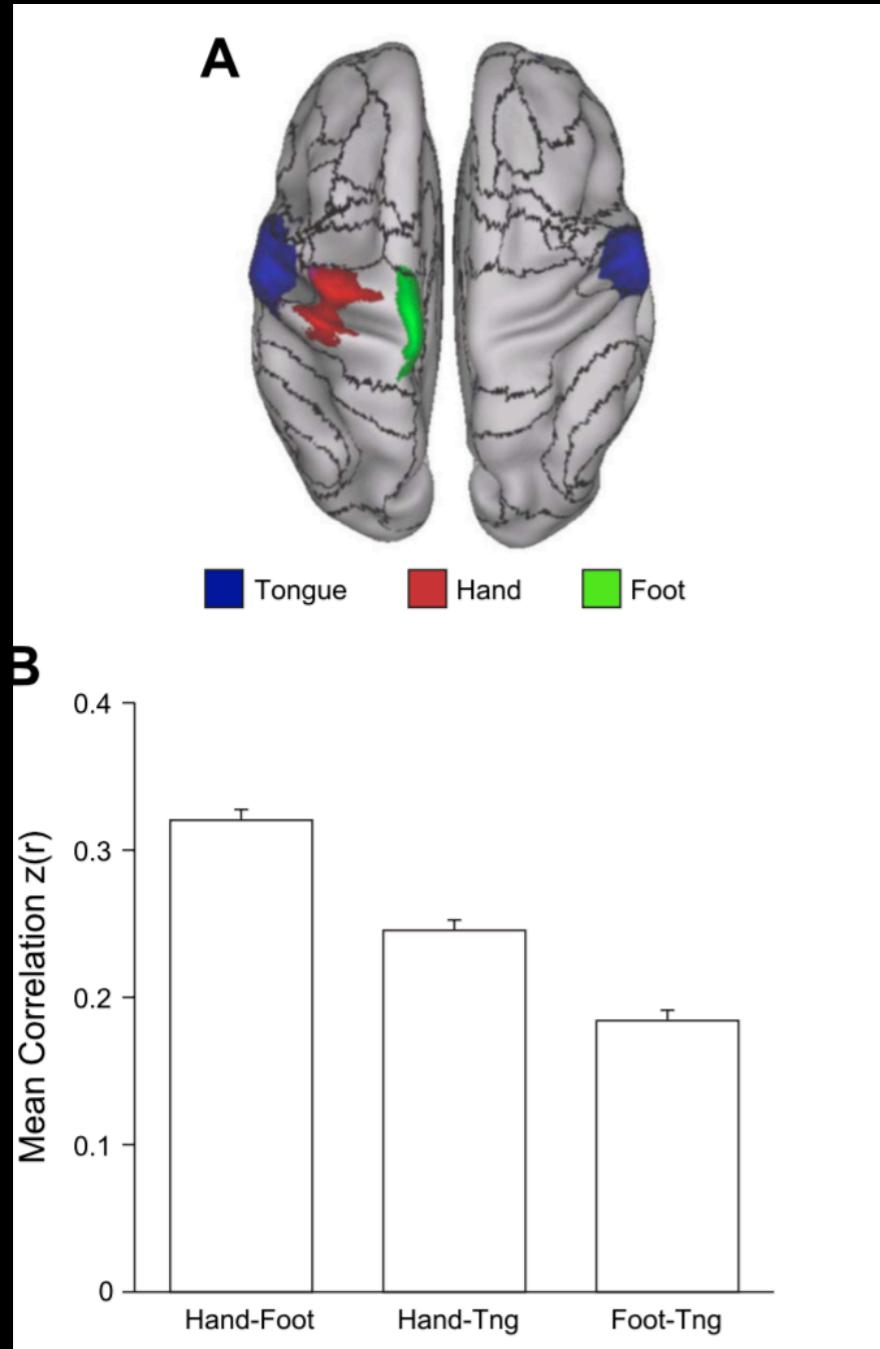
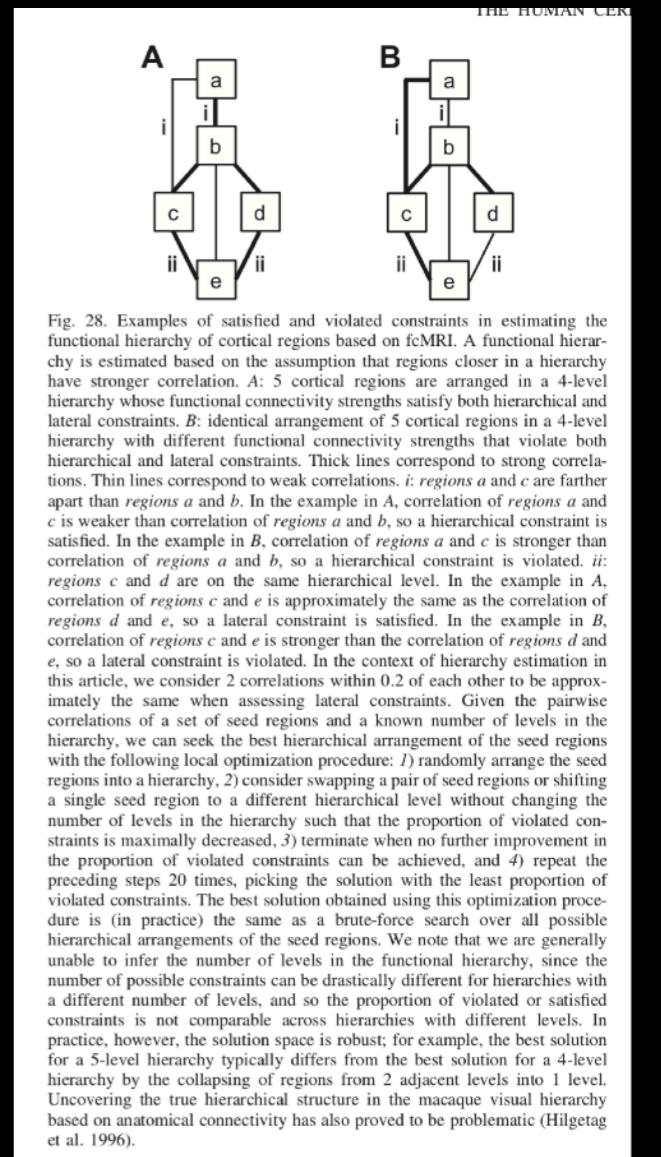
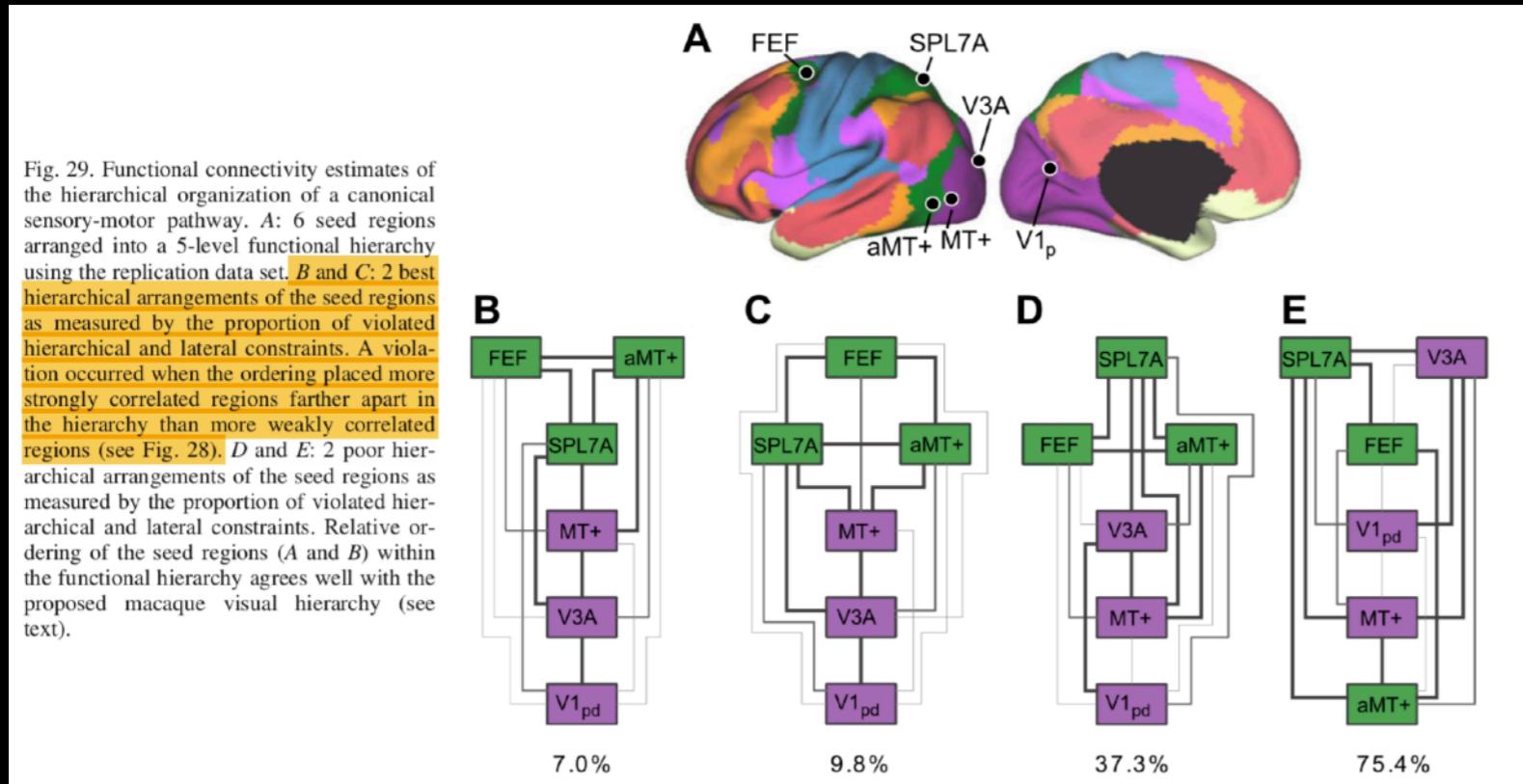
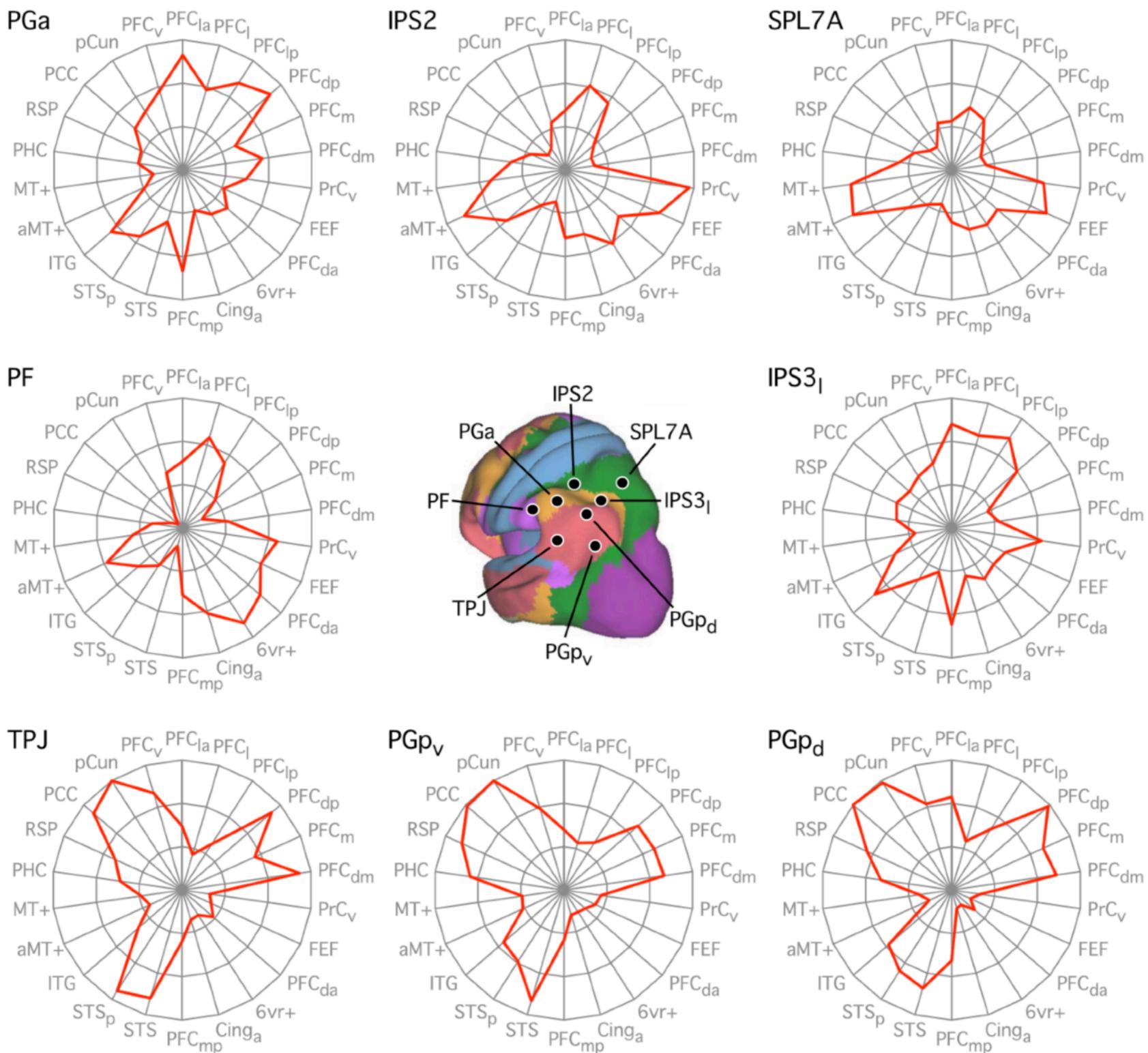
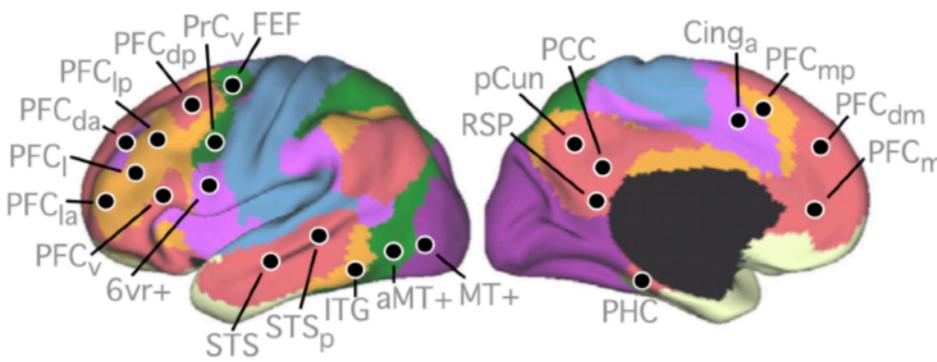


Fig. 21. Evidence that the interhemispheric fcMRI of homotopic regions within the primary motor cortex is topographically organized. **A:** correlation strength of left hemisphere tongue, hand, and foot seed regions with corresponding contralateral seed regions averaged over all 1,000 subjects. Right hemisphere vertices were obtained by reflection across the midline. Hand coordinates = $\pm 41, -20, 62$; foot coordinates = $\pm 6, -26, 76$; and tongue coordinates = $\pm 55, -4, 26$. The tongue representation has the strongest interhemispheric correlation, followed by the foot and then the hand. **B:** plot of interhemispheric correlation along the ventral (tongue) to dorsal (foot) extent of motor cortex. Maximal interhemispheric correlation is highest near the tongue representation and also peaks between the hand and foot representations, possibly corresponding to the trunk representation.





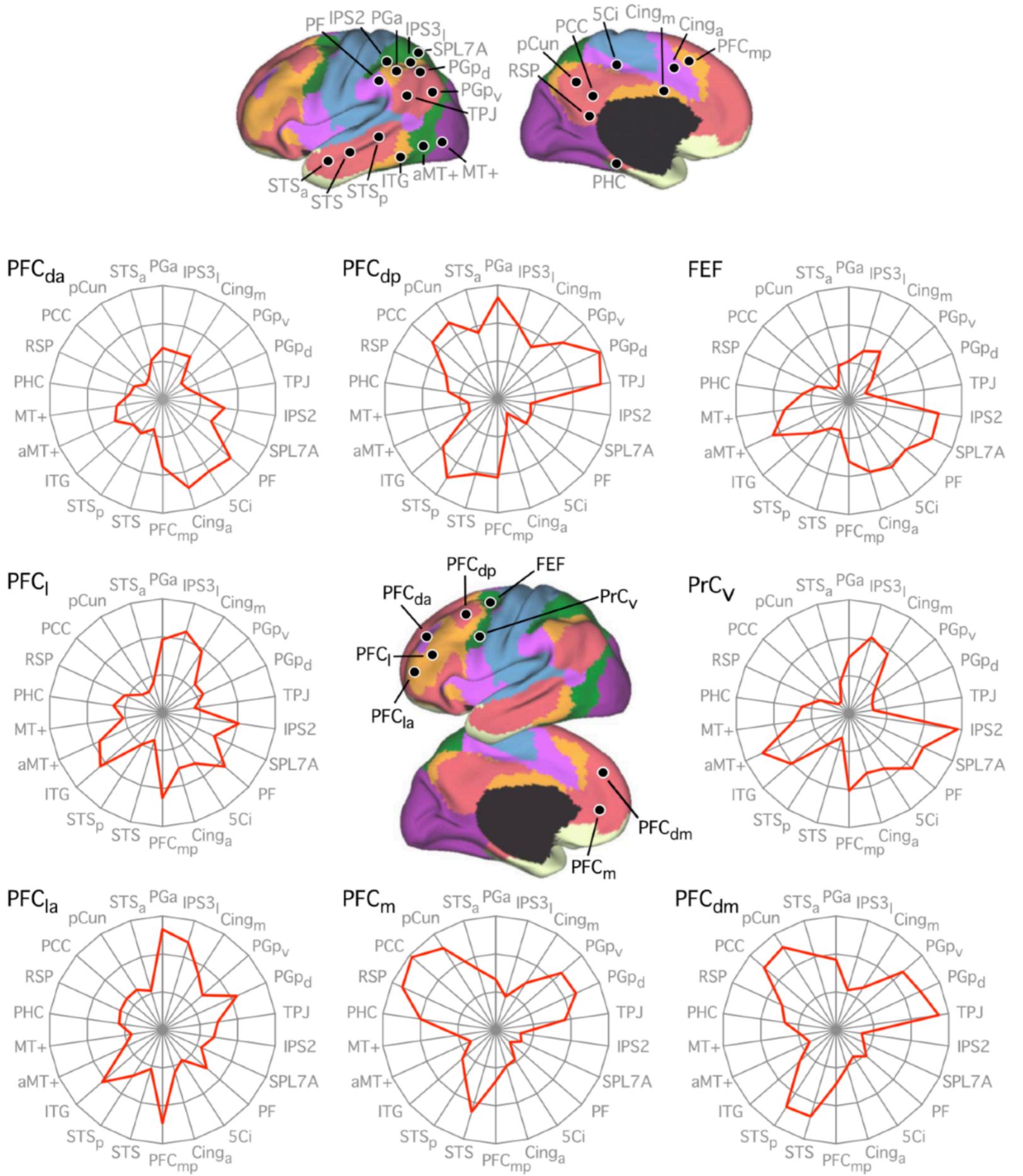


Fig. 31. Adjacent frontal regions exhibit distinct functional connectivity fingerprints. The format and plotting are the same as for Fig. 30 with regions tailored for exploration of frontal cortex. The coordinate locations are reported in Table 4. The polar scales range from $r = -0.4$ (center) to $r = 0.5$ (outer boundary) in 0.3-step increments.