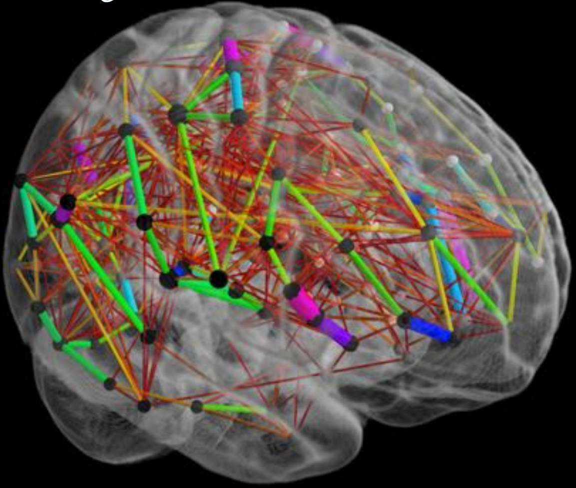
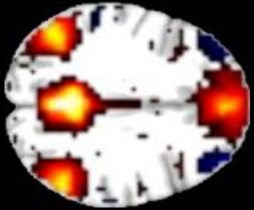
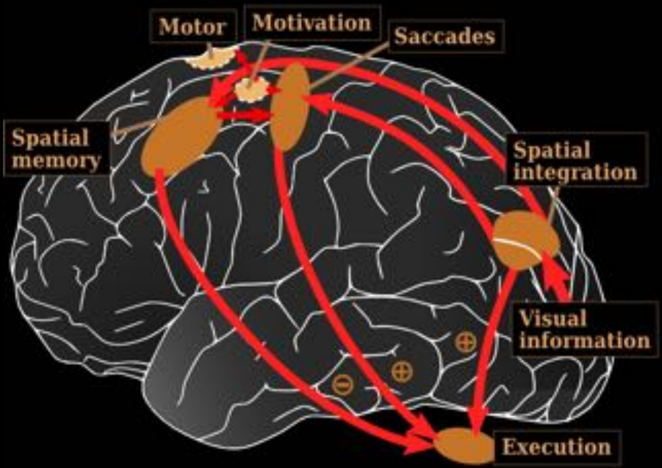
Functional connectomics:
Extracting and quantifying connectivity in fMRI

Gaël Varoquaux

PARIETAL Inria NeuroSpin

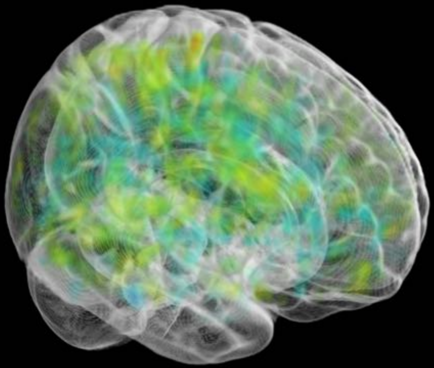






Captures functional interactions

Probing rest



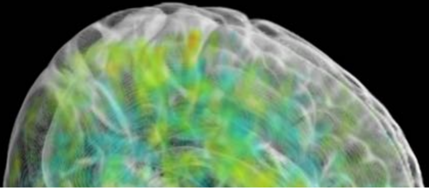
- Activation mapping is paradigm-driven: not ecologic

  Resting-state probes *intrinsic* structure

- Activation mapping requires demanding tasks

  $\Rightarrow$ inapplicable to diminished subjects

  Resting-state is easily applicable everywhere



Population imaging

- Scanning many subjects to study variability
- Links with neuropsychological profiling, genomics...
- A window to imaging epidemiology

Rest fMRI on **dozens of thousands** of subjects

- Activation mapping requires demanding tasks

$\Rightarrow$ inapplicable to diminished subjects

Resting-state is easily applicable everywhere

The brain at rest?

Metabolism (measured via PET)

- The brain represents 2% of body weight, but 20% of energy consumed
- Difficult cognitive tasks modulate consumption by less than 10%

[Raichle and Mintun 2006]

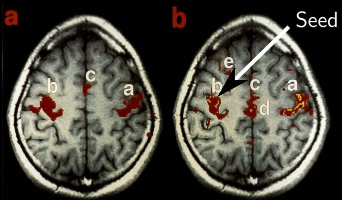
- Neural firing never stop (EEG/MEG evidence)

Study of brain activity in the absence of task

"Resting state"

Shared structured between on-going and evoked

[Biswal... 1995]



Task activation   Rest correlation

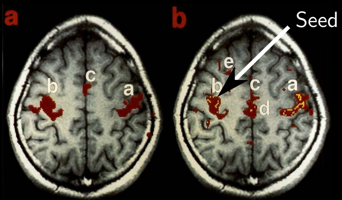
[Biswal... 1995]: **fMRI**

- Finger-tapping task to map the motor finger cortex
- During rest: which voxels correlate to the activity of this region?

Resting-state activity to study cognition?

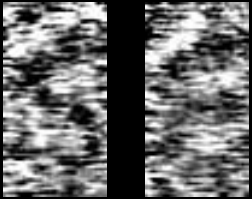
Shared structured between on-going and evoked

[Biswal... 1995]



Task activation   Rest correlation

[Kenet... 2003]



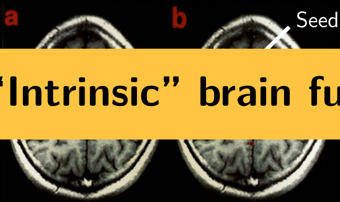
Spont.   Evoked

[Kenet... 2003]: **Voltage sensitive dye imaging**

- Visual cortex: cortical columns related to stimuli orientation
- Without stimuli, similar activity maps sometimes appear

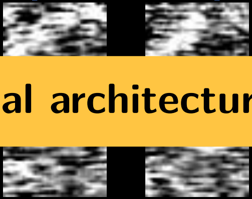
Shared structured between on-going and evoked

[Biswal... 1995]



Task activation   Rest correlation

[Kenet... 2003]

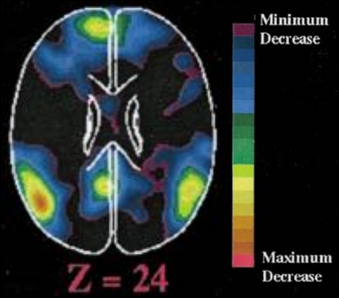


Spont.   Evoked

"Intrinsic" brain functional architecture

- The physical brain architecture (connections, cortical columns) is present in the absence of stimuli

The “default mode network”



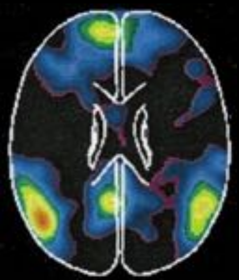
- Brain regions that deactivate during task

  [Raichle... 2001]

- Appear as an integrated network during rest

  [Greicius... 2003]

Notion of resting-state network



- Resting-state activity can be decomposed into networks
- How to do it systematically is a difficult question...

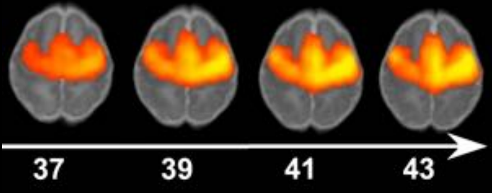
Capturing behavior or phenotype

[Lewis... 2009] *Learning sculpts the spontaneous activity of the resting human brain*

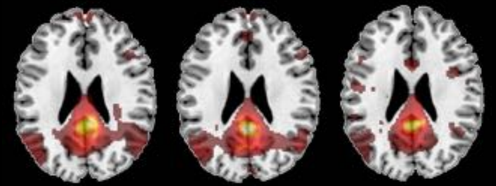
Strong perceptual training changed resting-state correlations

---

Cognition-less intrinsic activity



[Doria... 2010] *Emergence of resting state networks in the preterm human brain*



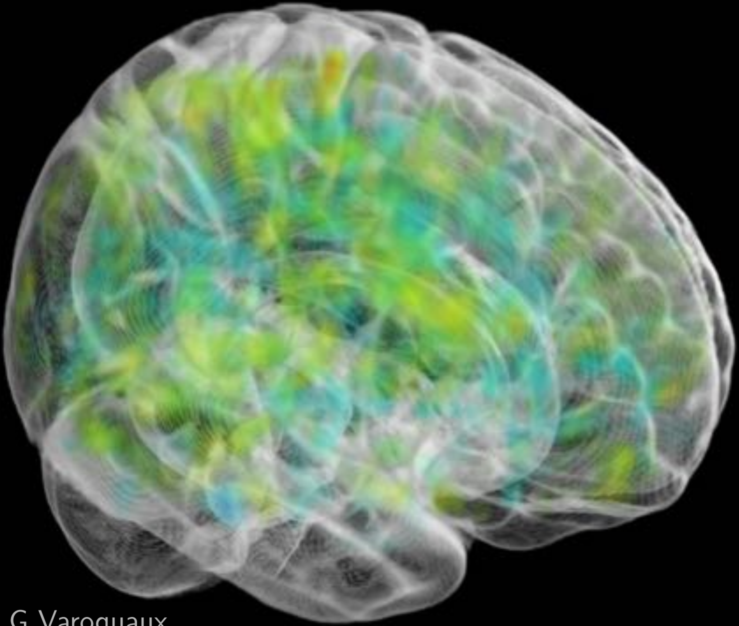
Awake Low Moderate

[Stamatakis... 2010] *Changes in resting neural connectivity during propofol sedation*

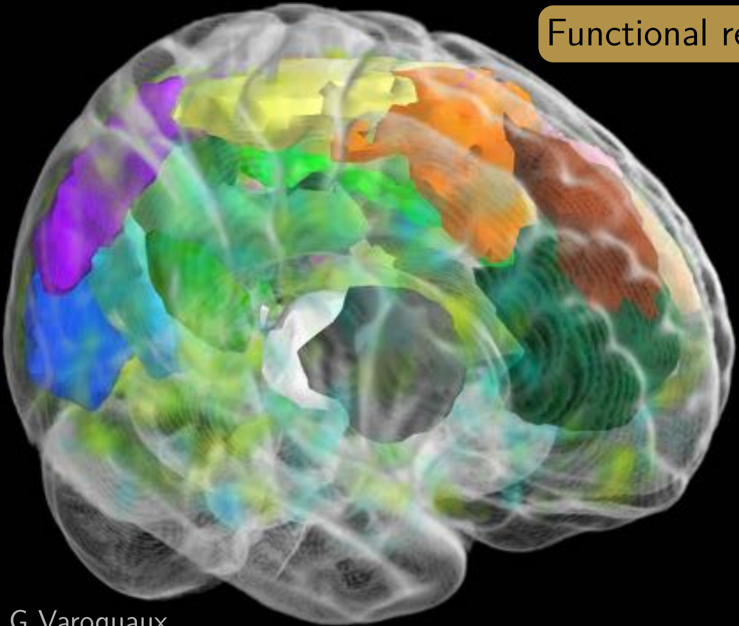
Functional connectivity and resting-state

- Notion of distributed functional networks
- “Functional connectivity” links and reveals them
- They correspond to an “intrinsic” brain architecture
- They can capture phenotype with simple experiments applicable disabled patient

Data analysis: conceptual models

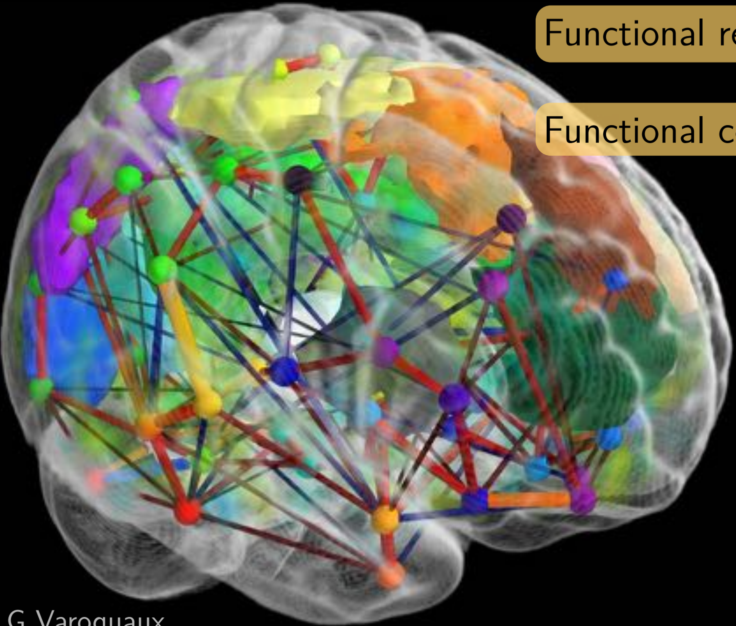


Data analysis: conceptual models



Functional regions

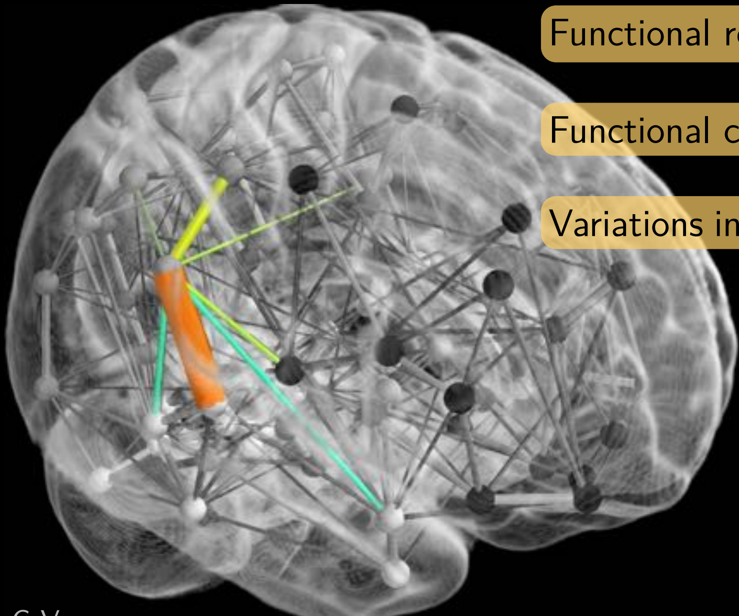
Data analysis: conceptual models



Functional regions

Functional connections

Data analysis: conceptual models



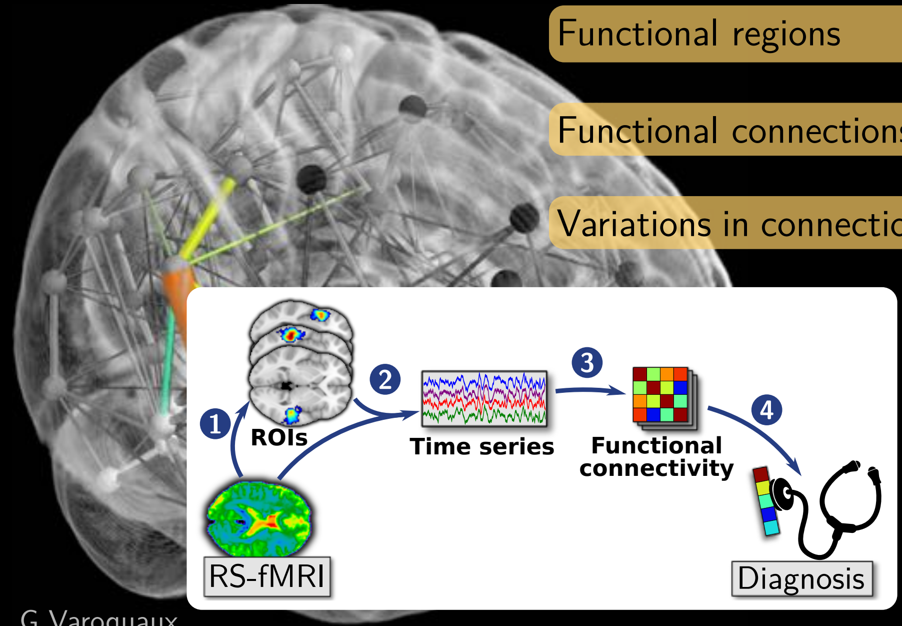
Functional regions

Functional connections

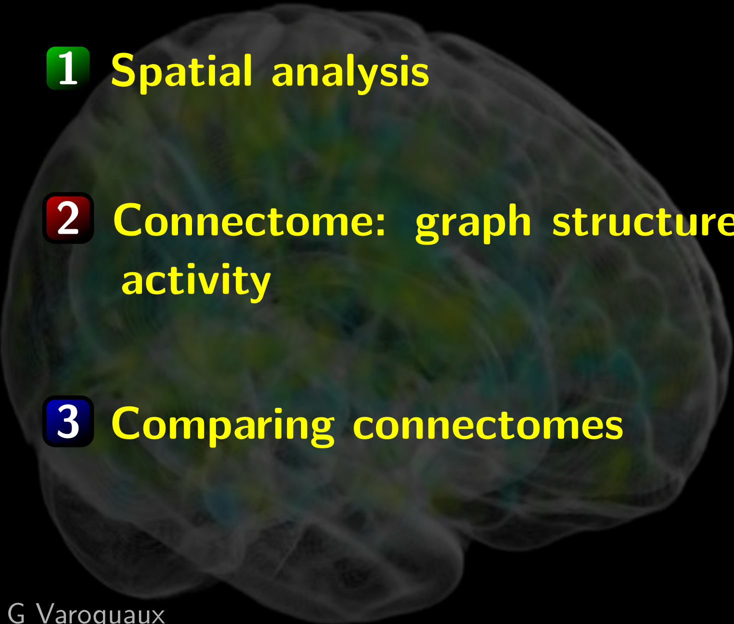
Variations in connections

Data analysis: conceptual models

- Functional regions
- Functional connections
- Variations in connections



Outline

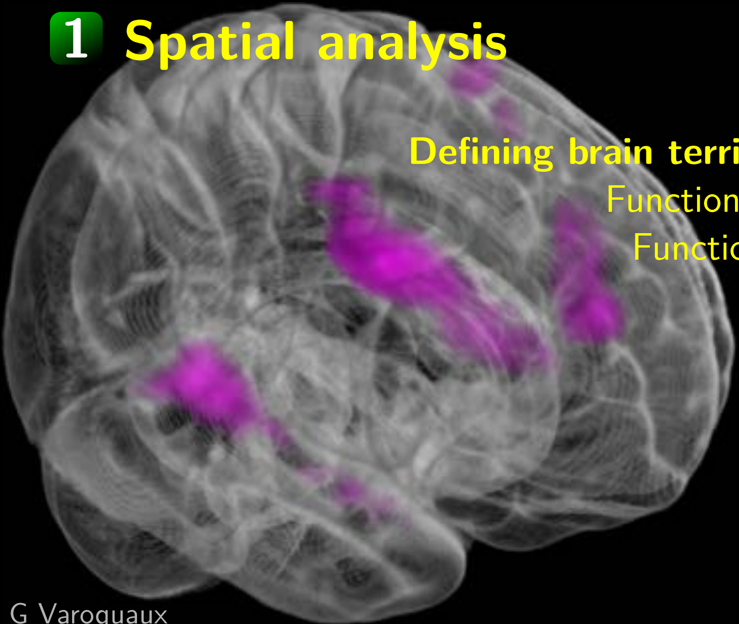


1. **Spatial analysis**

2. **Connectome: graph structure of brain activity**

3. **Comparing connectomes**

1 Spatial analysis



Defining brain territories:

Functional networks

Functional regions

Defining functional regions

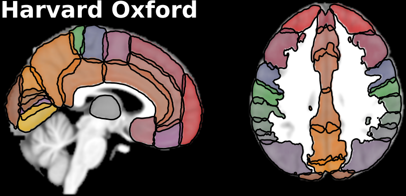
Dividing the brain in regions

anatomical atlases, functional atlases, region extraction methods

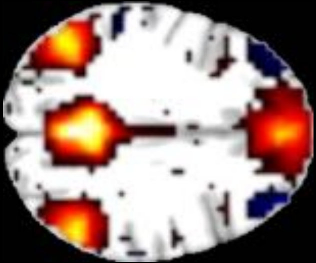
Some examples

- Anatomical atlases do not resolve functional structures

Harvard Oxford

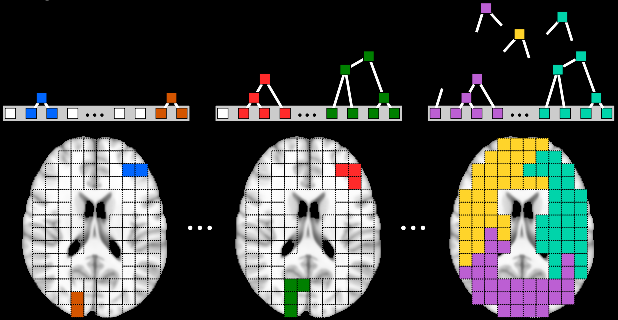


AAL



1 Clustering approaches

- Group together voxels with similar time courses



Give a *parcellation*:

each and every voxel is affected to one cluster

[Thirion... 2014]

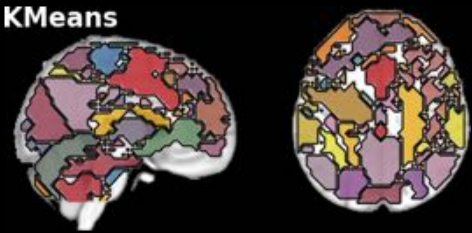
Finds cluster centers (prototype time-series) and assignements to minimize squared residuals

Pros

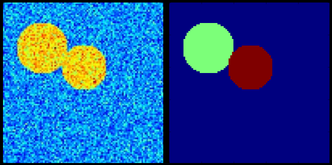
- There exists fast variants
- Good for few clusters

Cons

- No spatial constraint
  ⇒ (smooth the data)



[Thirion... 2014]



A variant of *spectral clustering*

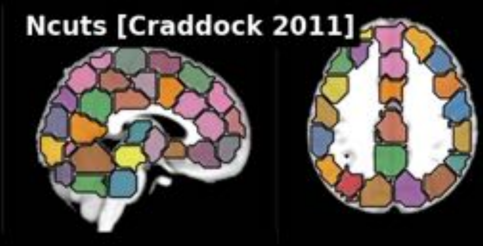
Adds a “surface energy” term: cost of cutting the graph of neighbors

Pros

- Spatial constraints
- Good for few clusters

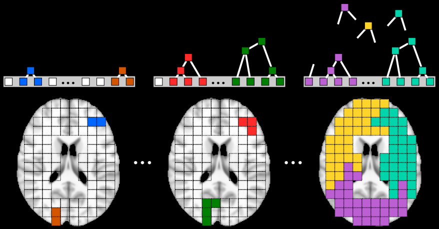
Cons

- Slow
- Very geometrical



[Craddock... 2012, Thirion... 2014]

1 Clustering approaches: Ward clustering



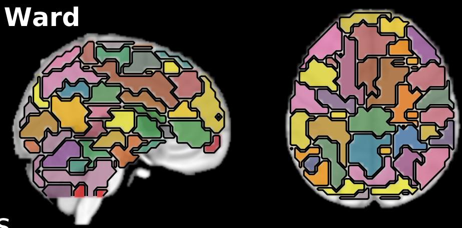
An algomerative clustering approach that minimizes variance

Pros

- Fast spatial constraints
  (even with many clusters)
- Good for many clusters

Cons

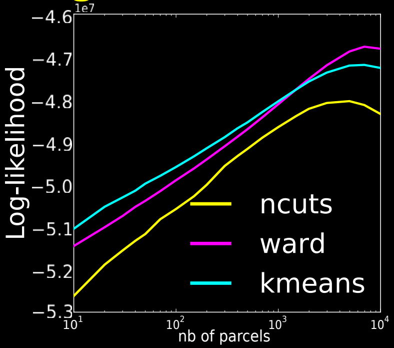
- Capture noise in big clusters



[Thirion... 2014]

Validation is hard

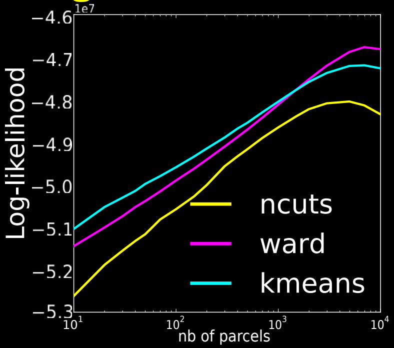
Signal fit



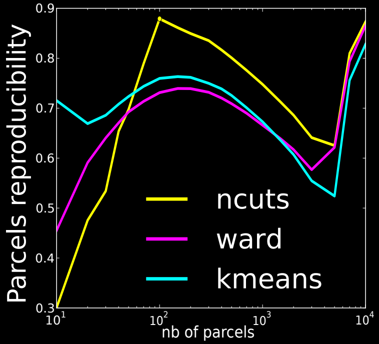
[Thirion... 2014]

Validation is hard

Signal fit



Parcel reproducibility



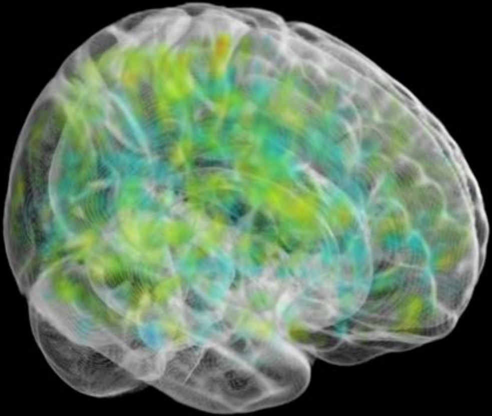
- K-means for small # of clusters
- Ward for large # of clusters

[Thirion... 2014]

Working hypothesis:

Observing linear mixtures of networks at rest

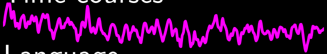
Time courses



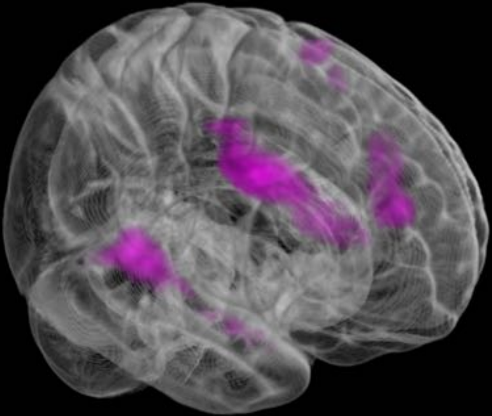
Working hypothesis:

Observing linear mixtures of networks at rest

Time courses

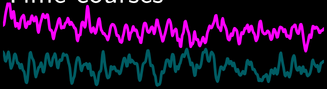


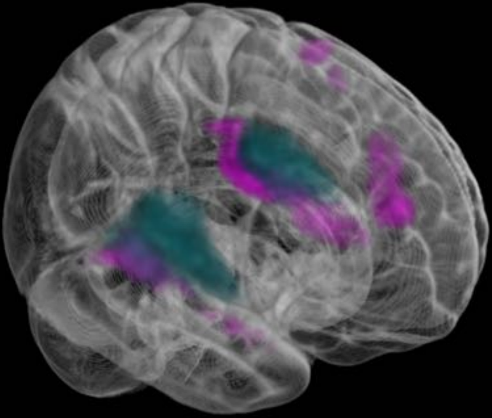
Language



Working hypothesis:

Observing linear mixtures of networks at rest

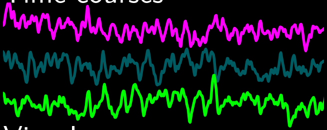




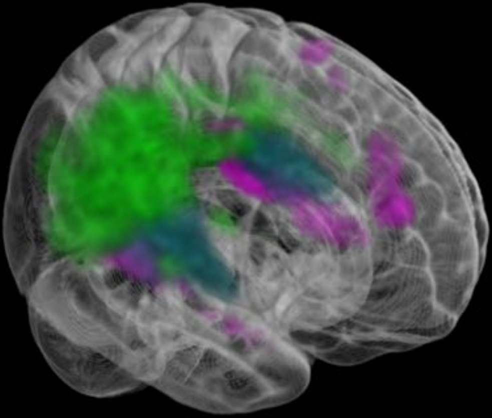
Working hypothesis:

Observing linear mixtures of networks at rest

Time courses



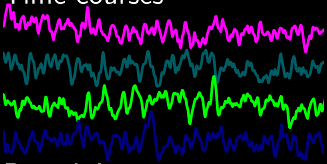
Visual



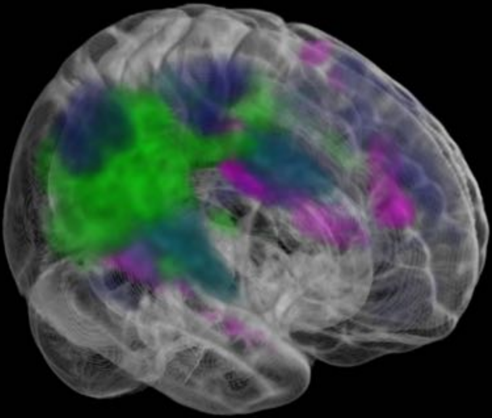
Working hypothesis:

Observing linear mixtures of networks at rest

Time courses

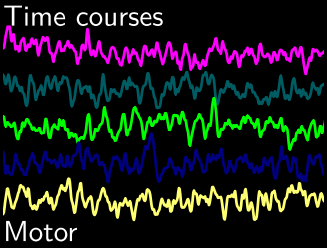


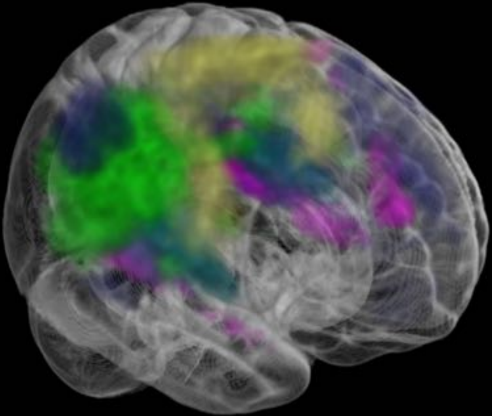
Dorsal Att.



Working hypothesis:

Observing linear mixtures of networks at rest

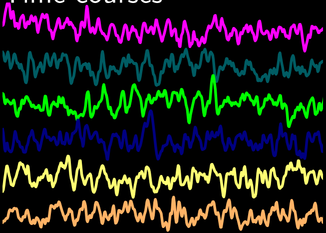




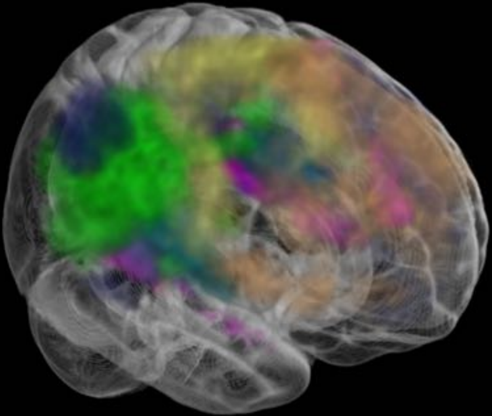
Working hypothesis:

Observing linear mixtures of networks at rest

Time courses

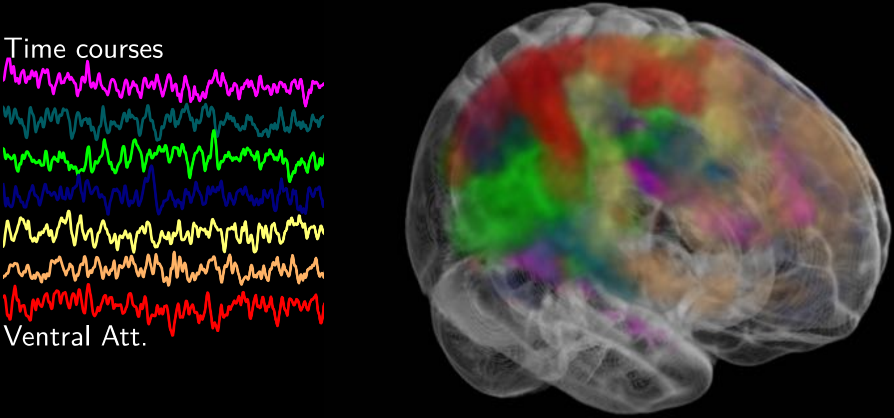


Salience



Working hypothesis:

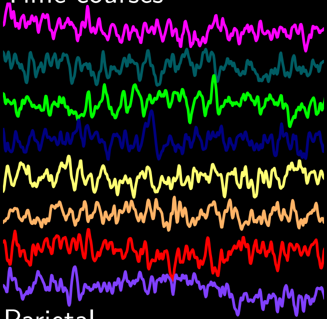
Observing linear mixtures of networks at rest



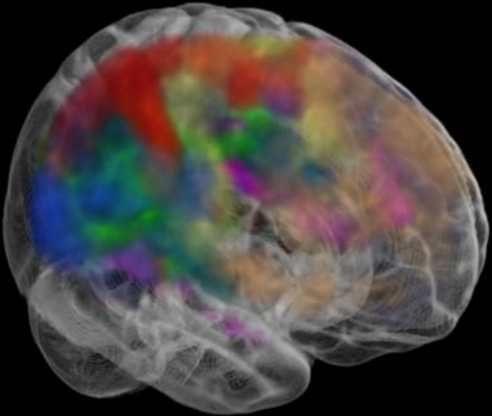
Working hypothesis:

Observing linear mixtures of networks at rest

Time courses

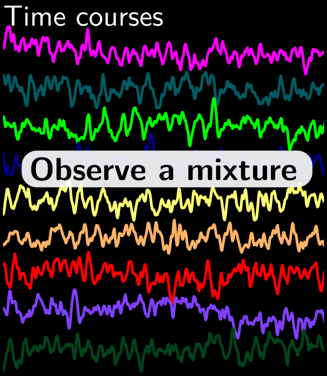


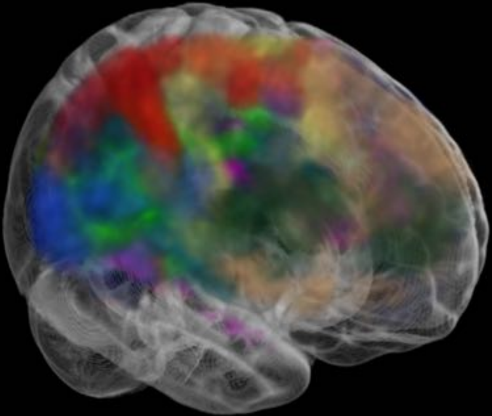
Parietal



Working hypothesis:

Observing linear mixtures of networks at rest

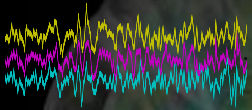


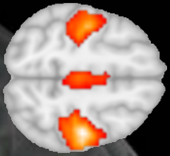


How to unmix networks?

1 Spatial modes: ICA decomposition

$$\text{time}\overset{\text{voxels}}{\left[\quad Y \quad\right]} = \text{time}\left[E \right] \cdot \overset{\text{voxels}}{\left[\quad S \quad\right]} + \text{time}\overset{\text{voxels}}{\left[\quad N \quad\right]}$$





Decomposing time series into:

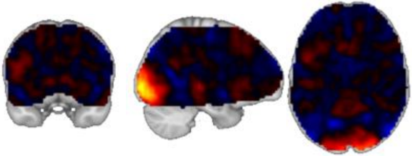
- covarying spatial maps, **S**
- uncorrelated residuals, **N**

ICA: minimize mutual information across S

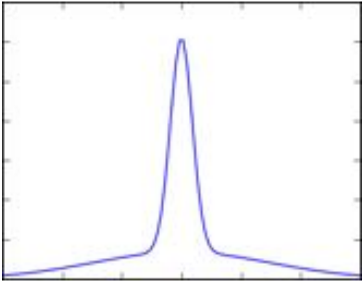
[Kiviniemi... 2003, Beckmann and Smith 2004, Varoquaux... 2010c]

1 Spatial modes: ICA decomposition

voxels voxels



Histogram of interesting maps are non-Gaussian

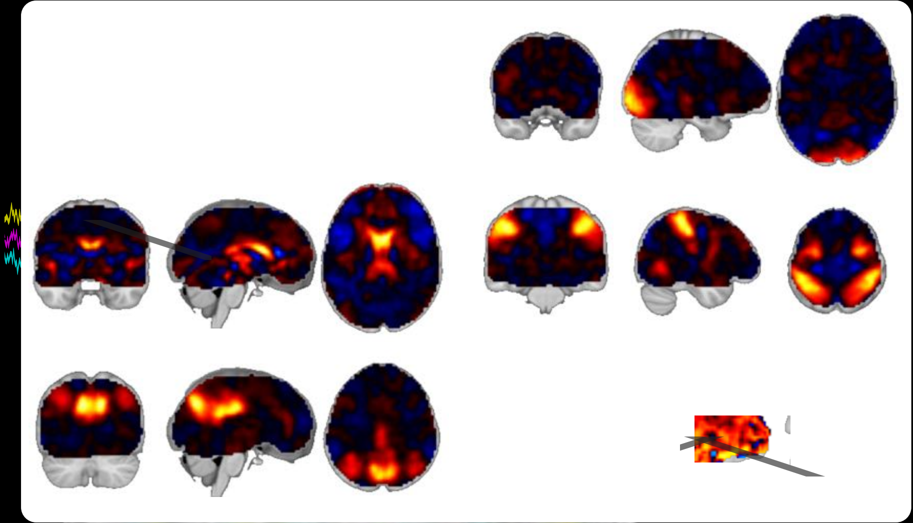


ICA: minimize mutual information across S

1 Spatial modes: ICA decomposition

voxels

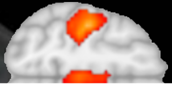
voxels



ICA: minimize mutual information across S

1 Spatial modes: Sparse decomposition

$$\text{time}\overset{\text{voxels}}{\left[\quad Y \quad\right]} = \text{time}\left[\, E \,\right] \cdot \overset{\text{voxels}}{\left[\quad S \quad\right]} + \text{time}\overset{\text{voxels}}{\left[\quad N \quad\right]}$$



Estimation via minimization:

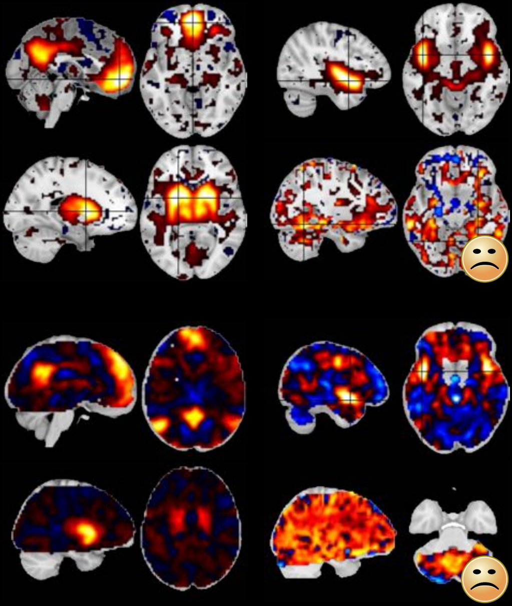
loss (error term) + penalty

$$\mathbf{E}, \mathbf{S} = \underset{\mathbf{E},\mathbf{S}}{\operatorname{argmin}} \|\mathbf{Y} - \mathbf{E}\,\mathbf{S}^T\|^2 + \lambda\|\mathbf{S}\|_1$$

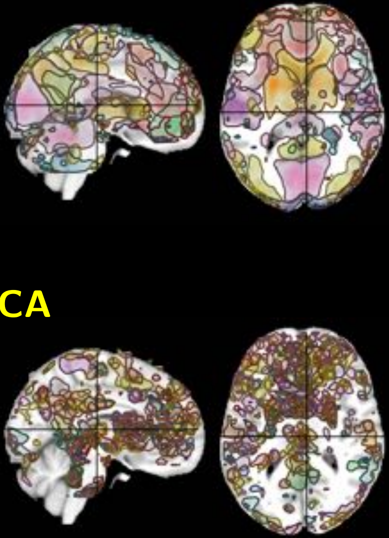
ℓ_1 **norm on S creates sparsity**

Sparse decompositions: sparse penalty on maps

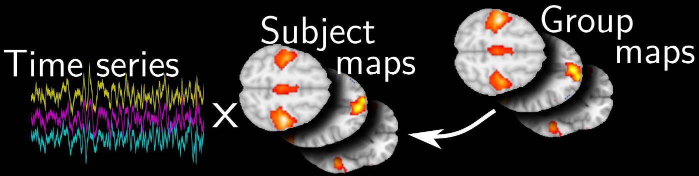
Spatial modes: Sparse decomposition



ICA



[Varoquaux in prep]



- Subject level spatial patterns:

$$\mathbf{Y}^s = \mathbf{U}^s {\mathbf{V}^s}^T + \mathbf{E}^s, \qquad \mathbf{E}^s \sim \mathcal{N}(0, \sigma \mathbf{I})$$

- Group level spatial patterns:

$$\mathbf{V}^s = \mathbf{V} + \mathbf{F}^s, \qquad \mathbf{F}^s \sim \mathcal{N}(0, \zeta \mathbf{I})$$

- Sparsity and spatial-smoothness prior:

$$\mathbf{V} \sim \exp\left(-\xi\, \Omega(\mathbf{V})\right), \qquad \Omega(\mathbf{v}) = \|\mathbf{v}\|_1 + \frac{1}{2} \mathbf{v}^T \mathbf{L} \mathbf{v}$$

[Varoquaux... 2011]

Estimation: maximum a posteriori

$$\underset{\mathbf{U}^s,\mathbf{V}^s,\mathbf{V}}{\operatorname{argmin}} \sum_{sujets} \left(\|\mathbf{Y}^s - \underbrace{\mathbf{U}^s{\mathbf{V}^s}^T}_{\text{Data fit}}\|^2_{\text{Fro}} + \mu \underbrace{\|\mathbf{V}^s - \mathbf{V}\|^2_{\text{Fro}}}_{\text{Subject variability}} \right) + \lambda\, \underbrace{\Omega(\mathbf{V})}_{\text{Penalization: sparse and smooth maps}}$$

[Varoquaux... 2011]

Estimation: maximum a posteriori

$$\underset{\mathbf{U}^s,\mathbf{V}^s,\mathbf{V}}{\operatorname{argmin}} \sum_{sujets} \Big(\underbrace{\|\mathbf{Y}^s - \mathbf{U}^s {\mathbf{V}^s}^T\|^2_{\text{Fro}}}_{\text{Data fit}} + \mu \underbrace{\|\mathbf{V}^s - \mathbf{V}\|^2_{\text{Fro}}}_{\text{Subject variability}} \Big) + \underbrace{\lambda\, \Omega(\mathbf{V})}_{\text{Penalization: sparse and smooth maps}}$$

Alternate optimization on $\mathbf{U}^s$, $\mathbf{V}^s$, $\mathbf{V}$:

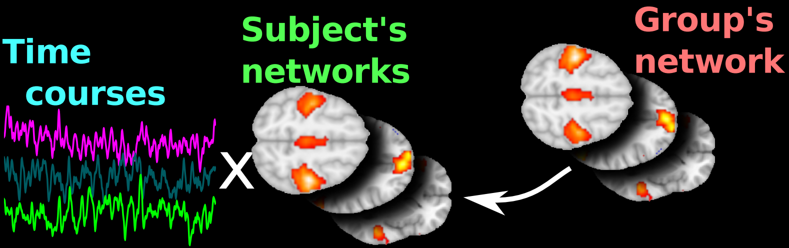
Update $\mathbf{U}^s$: standard dictionary learning procedure [Mairal... 2010]

Update $\mathbf{V}^s$: ridge regression on $(\mathbf{V}^s - \mathbf{V})^T$

Update $\mathbf{V}$: proximal operator for $\lambda\,\Omega$:

$$\underset{\mathbf{v}}{\operatorname{argmin}} \sum_{s=1}^{S} \frac{1}{2}\|\mathbf{v}^s - \mathbf{v}\|^2_2 + \gamma\,\Omega(\mathbf{v}) = \operatorname{prox}_{\gamma/S\,\Omega} \bar{\mathbf{v}}, \qquad \bar{\mathbf{V}} = \operatorname{mean}_s \mathbf{V}^s$$

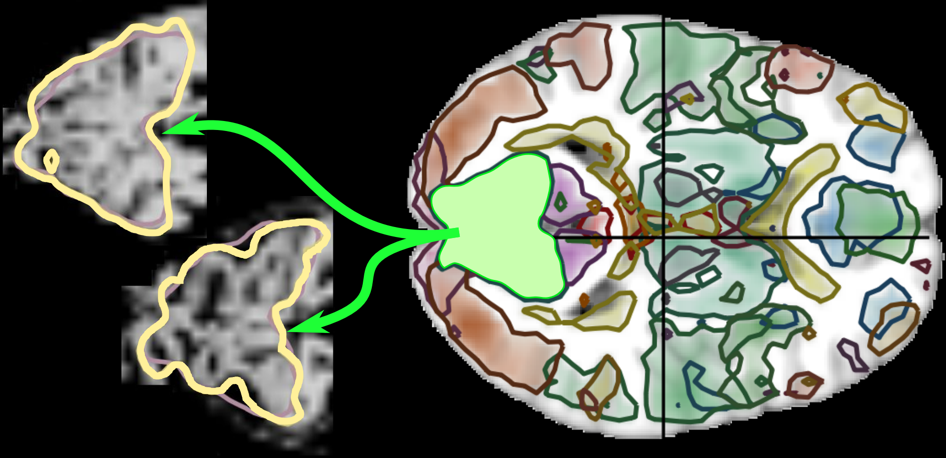
[Varoquaux... 2011]



Multi-Subject Dictionary Learning

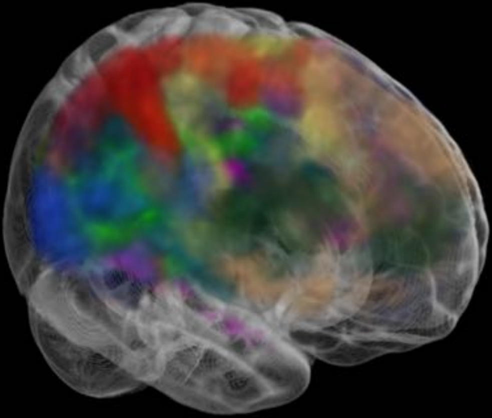
[Varoquaux... 2011]

Individual maps + Population-level atlas



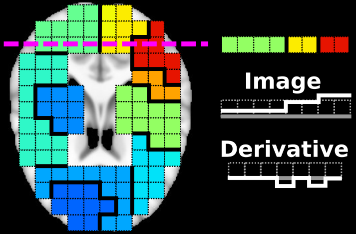
[Varoquaux... 2011]

[Kiviniemi... 2009] Extracting many networks with ICA almost forms a brain parcellation



1 In dictionary learning: Total-variation MSDL

Create a region-forming penalty:



Total-variation penalization

Impose sparsity on the gradient of the image:

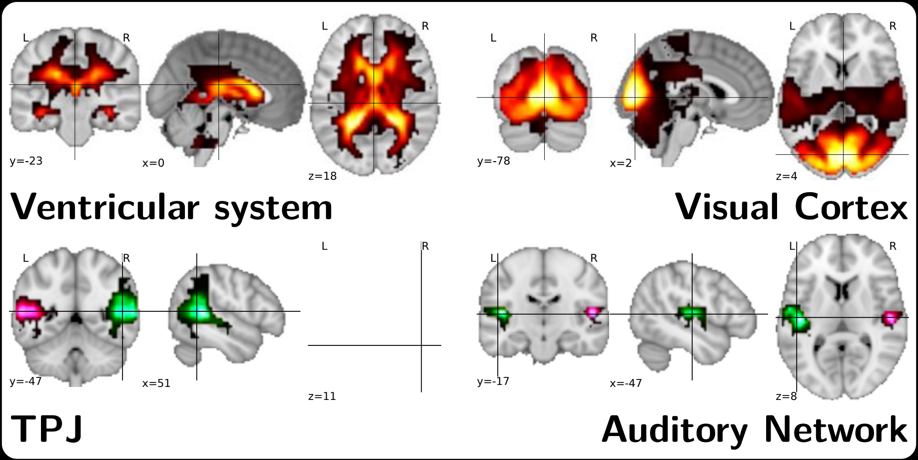
$$p(\mathbf{w}) = \ell_1(\nabla \mathbf{w})$$

Original

Clustering

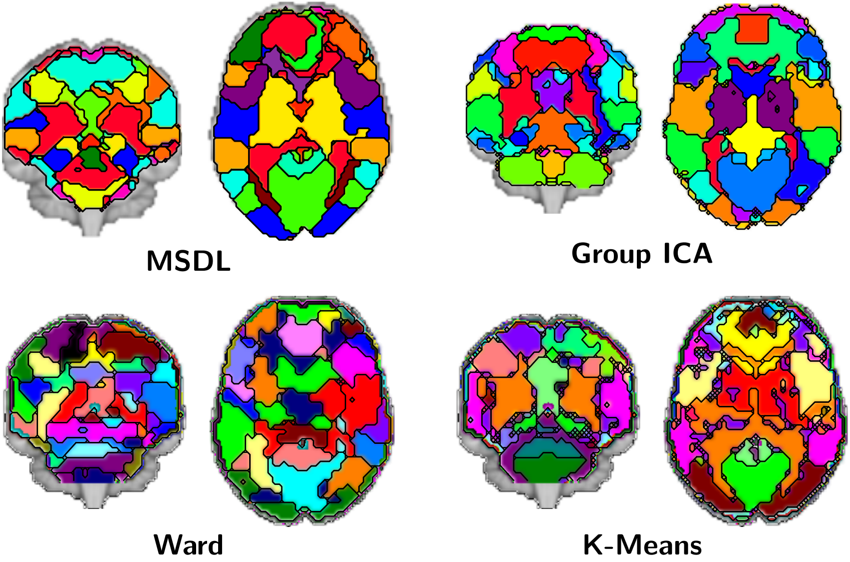
Total-variation

1 In dictionary learning: Total-variation MSDL



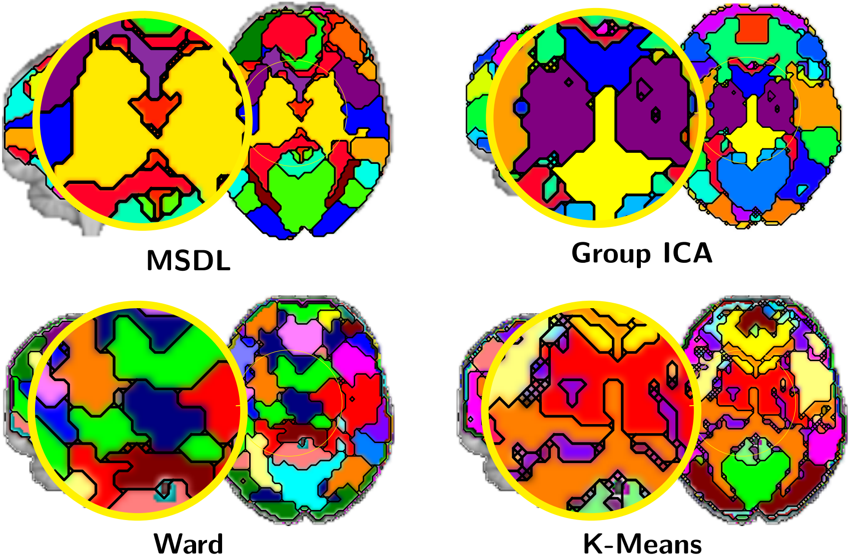
[Abraham... 2013]

Data-driven brain parcellations



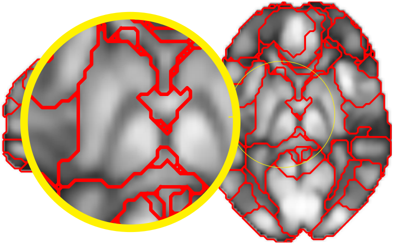
[Abraham... 2013]

Data-driven brain parcellations

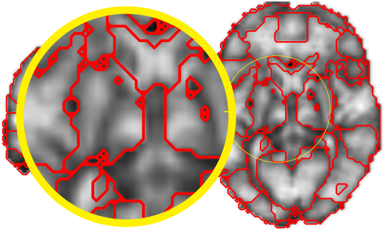


[Abraham... 2013]

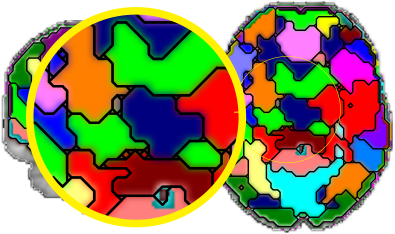
Data-driven brain parcellations



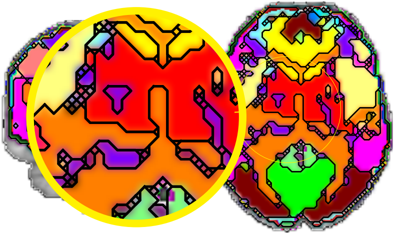
MSDL



Group ICA



Ward



K-Means

[Abraham... 2013]

Functional regions

AAL

Smith 2009 ICAs

Craddock 2011 Ncuts

Abraham 2013 TV-MSDL

Ward

Harvard-Oxford

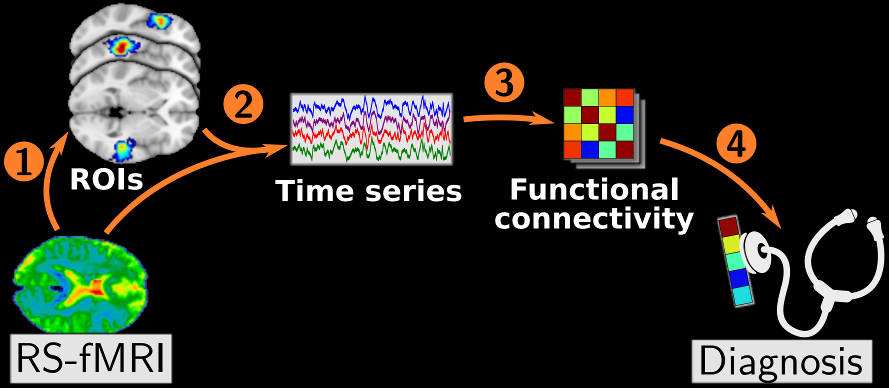
High model order ICA

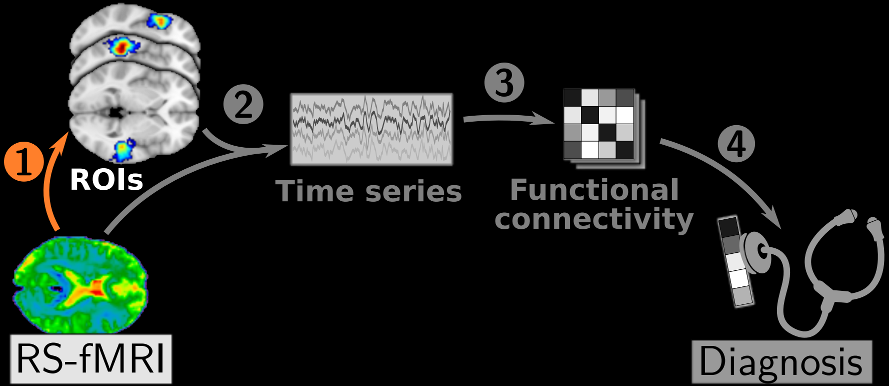
K-Means

Varoquaux 2011 Smooth-MSDL

Yeo 2011

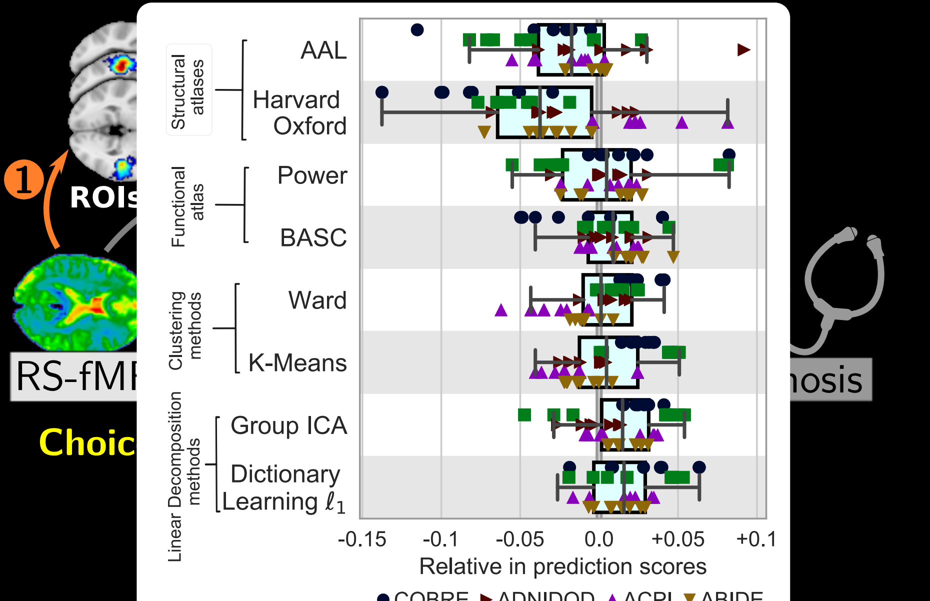
1 In connectome prediction settings





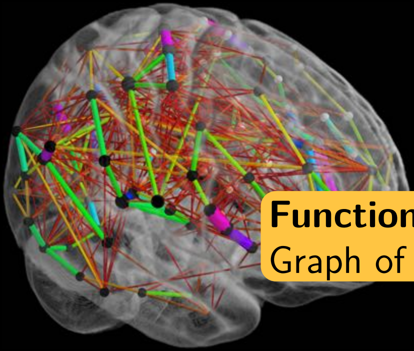
Choice of regions for best prediction?

1 In connectome prediction settings



[Reddy in prep]

2 Connectome: graph structure of brain activity



Functional connectome
Graph of interactions between regions

[Varoquaux and Craddock 2013]

Graphical model in cognitive neuroscience

Whish list

- Causal links
- Directed model:

$$IPS = V2 + MT$$
$$FEF = IPS + ACC$$

Whish list

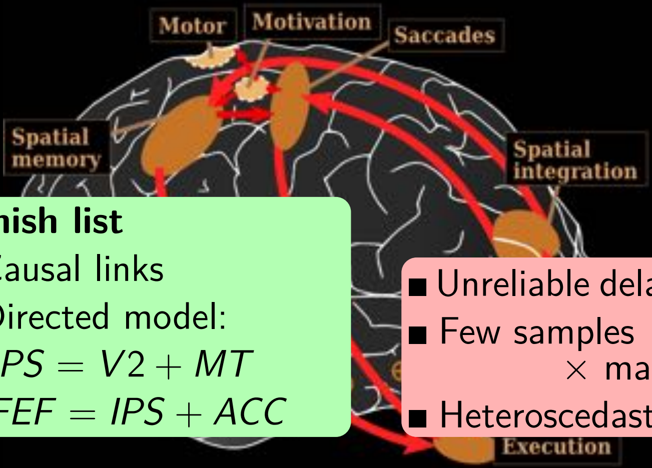
- Causal links
- Directed model:

  $IPS = V2 + MT$

  $FEF = IPS + ACC$

- Unreliable delays (HRF)
- Few samples $\times$ many signals
- Heteroscedastic noise

2 Graphical model in cognitive neuroscience



Whish list

- Causal links
- Directed model:

  $IPS = V2 + MT$

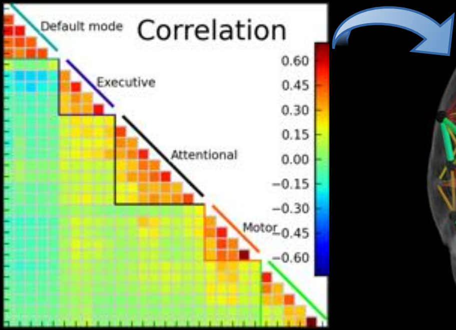
  $FEF = IPS + ACC$

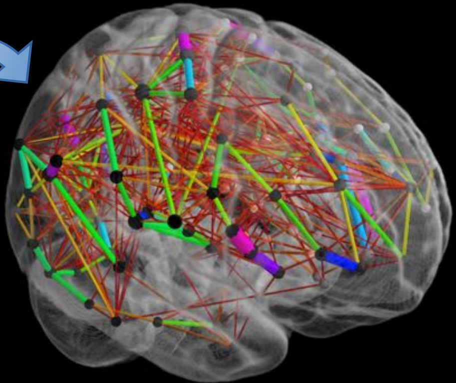
- Unreliable delays (HRF)
- Few samples $\times$ many signals
- Heteroscedastic noise

Independence structure

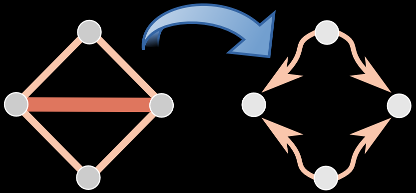
Knowing *IPS*, *FEF* is independent of *V2* and *MT*

2 From correlations to connectomes



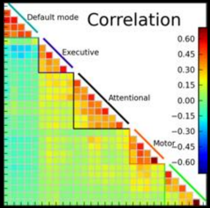


Conditional independence structure?



Probabilistic model for interactions

- Simplest **data generating process**
  = multivariate normal:



$$\mathcal{P}(\mathbf{X}) \propto \sqrt{|\mathbf{\Sigma}^{-1}|}e^{-\frac{1}{2}\mathbf{X}^T\mathbf{\Sigma}^{-1}\mathbf{X}}$$

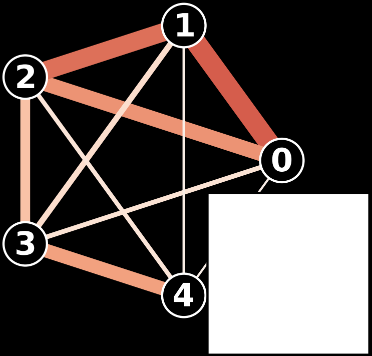
- Model parametrized by inverse covariance matrix,
  $\mathbf{K} = \mathbf{\Sigma}^{-1}$: ***conditional*** covariances

- Goodness of fit:
  likelihood of observed covariance $\hat{\mathbf{\Sigma}}$ in model $\mathbf{\Sigma}$

$$\mathcal{L}(\hat{\mathbf{\Sigma}}|\mathbf{K}) = \log|\mathbf{K}| - \text{trace}(\hat{\mathbf{\Sigma}}\,\mathbf{K})$$

Observations

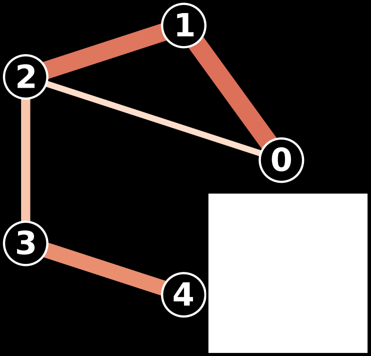
Covariance



Diagonal:
signal variance

Direct connections

Inverse covariance

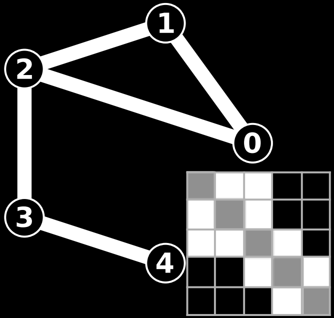


Diagonal:
node innovation

2 Independence structure (Markov graph)

Zeros in partial correlations
give **conditional independence**

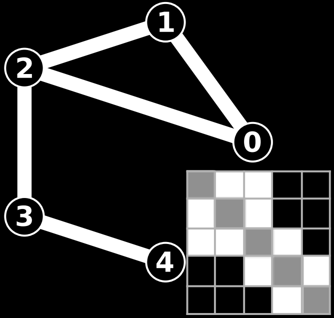
Reflects the large-scale brain interaction structure



2 Independence structure (Markov graph)

Zeros in partial correlations
give **conditional independence**

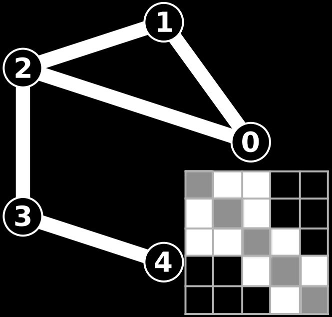
Ill-posed problem:
multi-collinearity
$\Rightarrow$ noisy partial correlations



Independence between nodes makes estimation of partial correlations well-conditionned.

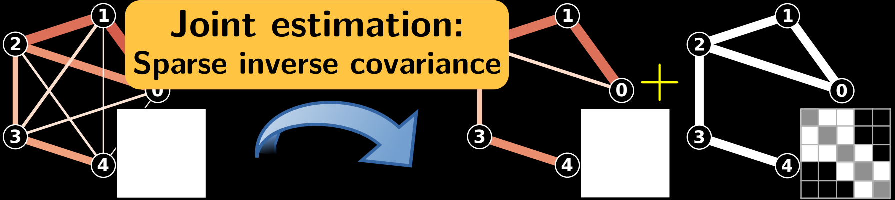
Chicken and egg problem

Zeros in partial correlations
give **conditional independence**



Ill-posed problem:
multi-collinearity
$\Rightarrow$ noisy partial correlations

Independence between nodes makes estimation of partial correlations well-conditionned.



Joint estimation:
Sparse inverse covariance

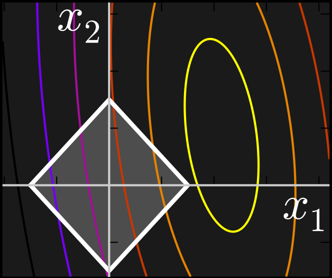
Maximum a posteriori:

- Fit models with a penalty
- Sparsity $\Rightarrow$ Lasso-like problem: ℓ_1 penalization

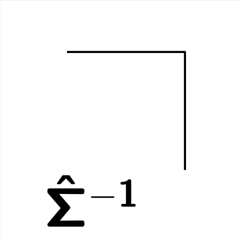
$$\mathbf{K} = \underset{\mathbf{K} \succ 0}{\operatorname{argmin}} \; \mathcal{L}(\hat{\mathbf{\Sigma}}|\mathbf{K}) + \lambda \, \ell_1(\mathbf{K})$$

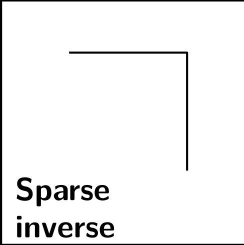
Data fit, Likelihood

Penalization,



 [Friedman... 2008, Varoquaux... 2010b, Smith... 2011]





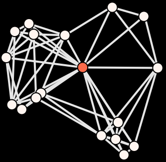
Likelihood of new data (cross-validation)

Subject data, Σ^{-1}	-57.1
Subject data, sparse inverse	43.0

[Varoquaux... 2010b]

Theoretical limitation to sparse recovery

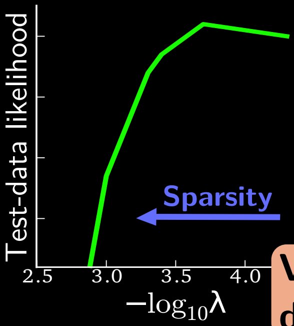
- Number of samples for s edges, p nodes:
  $n = \mathcal{O}\big((s+p)\log p\big)$ [Lam and Fan 2009]
- High-degree nodes fail [Ravikumar... 2011]

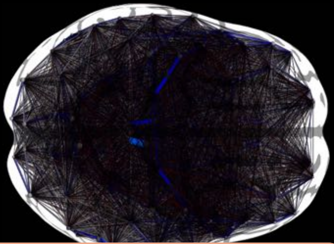


Empirically

- Optimal graph almost dense

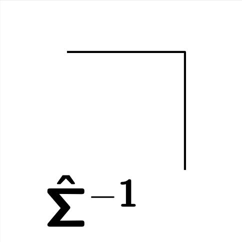
[Varoquaux... 2012]

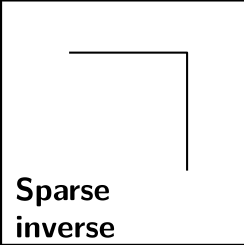


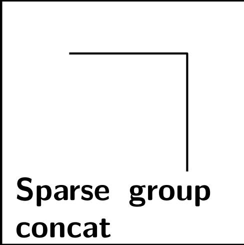


Very sparse graphs don't fit the data

Multi-subject to overcome subject data scarsity







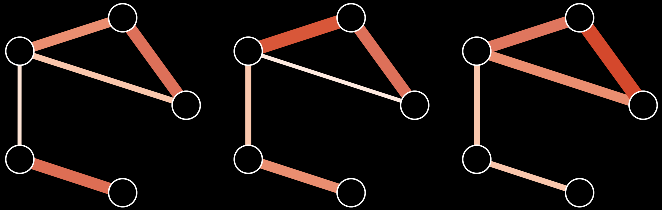
Likelihood of new data (cross-validation)

Subject data, Σ^{-1}	-57.1
Subject data, sparse inverse	43.0
Group concat data, Σ^{-1}	40.6
Group concat data, sparse inverse	41.8

Inter-subject variability

[Varoquaux... 2010b]

Common independence structure but different connection values



$$\{\mathbf{K}^s\} = \operatorname*{argmin}_{\{\mathbf{K}^s \succ 0\}} \sum_s \mathcal{L}(\hat{\mathbf{\Sigma}}^s | \mathbf{K}^s) + \lambda\, \ell_{21}(\{\mathbf{K}^s\})$$

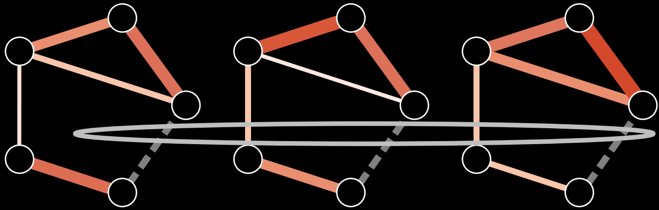
Multi-subject data fit, Likelihood

Group-lasso penalization

[Varoquaux... 2010b]

2 Multi-subject sparsity

Common independence structure but different connection values



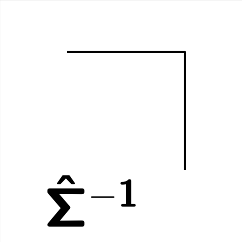
$$\{\mathbf{K}^s\} = \underset{\{\mathbf{K}^s \succ 0\}}{\operatorname{argmin}} \sum_s \mathcal{L}(\hat{\mathbf{\Sigma}}^s | \mathbf{K}^s) + \lambda\, \ell_{21}(\{\mathbf{K}^s\})$$

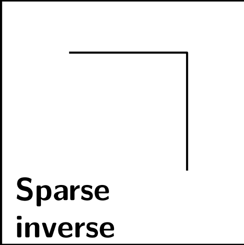
Multi-subject data fit, Likelihood

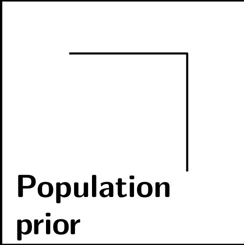
ℓ_1 on the connections of the ℓ_2 on the subjects

[Varoquaux... 2010b]

2 Multi-subject sparse graphs perform better



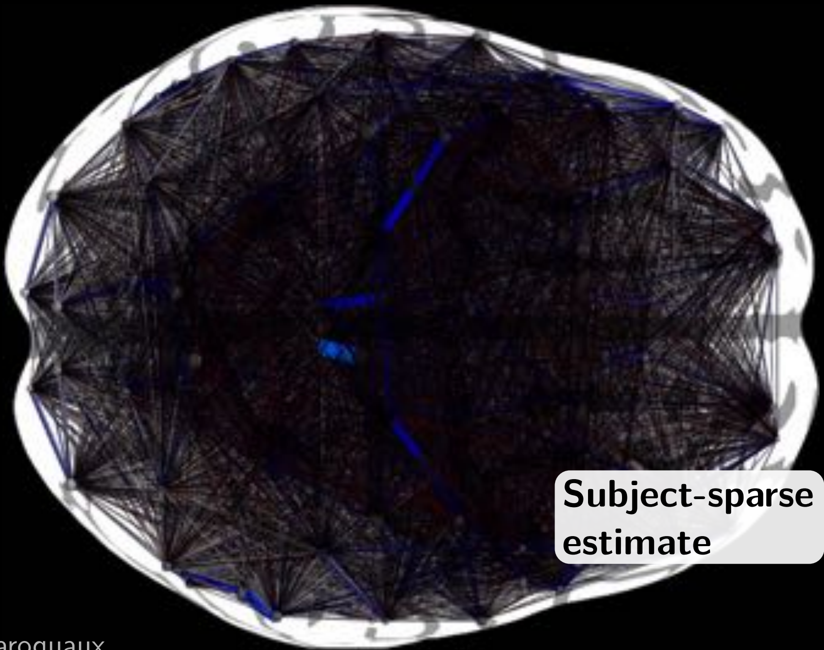




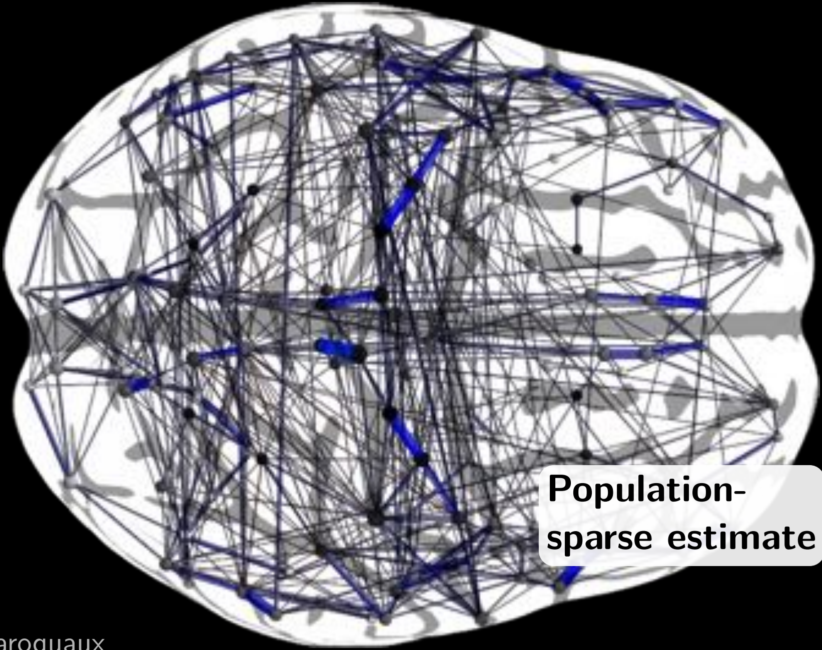
Likelihood of new data (cross-validation)		sparsity
Subject data, Σ^{-1}	-57.1	
Subject data, sparse inverse	43.0	60% full
Group concat data, Σ^{-1}	40.6	
Group concat data, sparse inverse	41.8	80% full
Group sparse model	**45.6**	20% full

[Varoquaux... 2010b]

2 Independence structure of brain activity

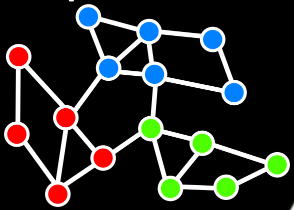


2 Independence structure of brain activity

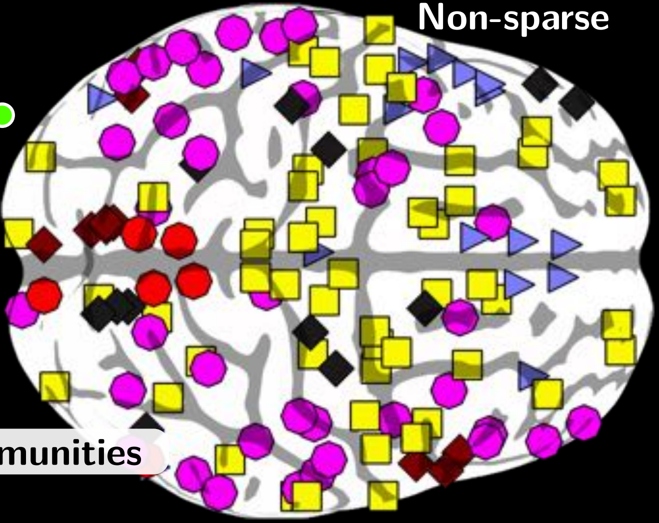


Population-sparse estimate

Graph communities

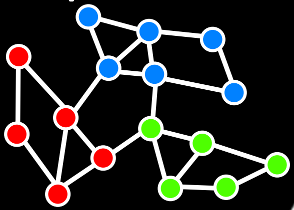


[Eguiluz... 2005]

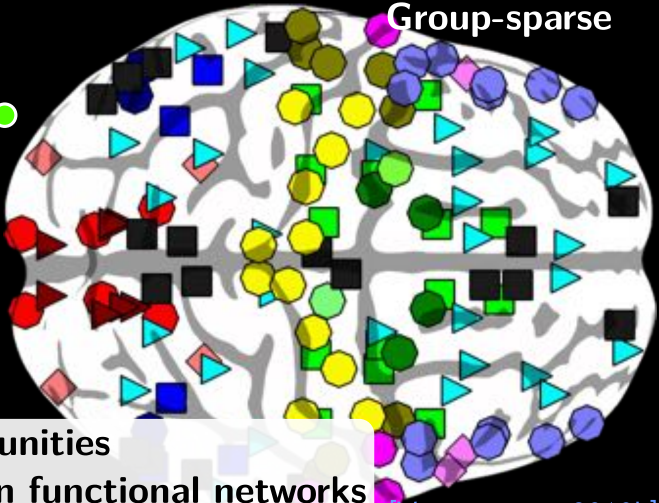


Neural communities

Graph communities



[Eguiluz... 2005]

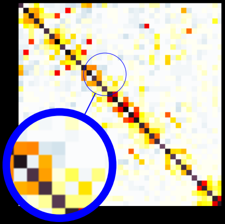


Neural communities
= large known functional networks

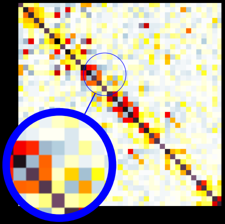
[Varoquaux... 2010b]

Sparsity is finicky

- Sensitive hyper-parameter
- Slow and unreliable convergence
- Unstable set of selected edges



Shrinkage



- Softly push partial correlations to zero

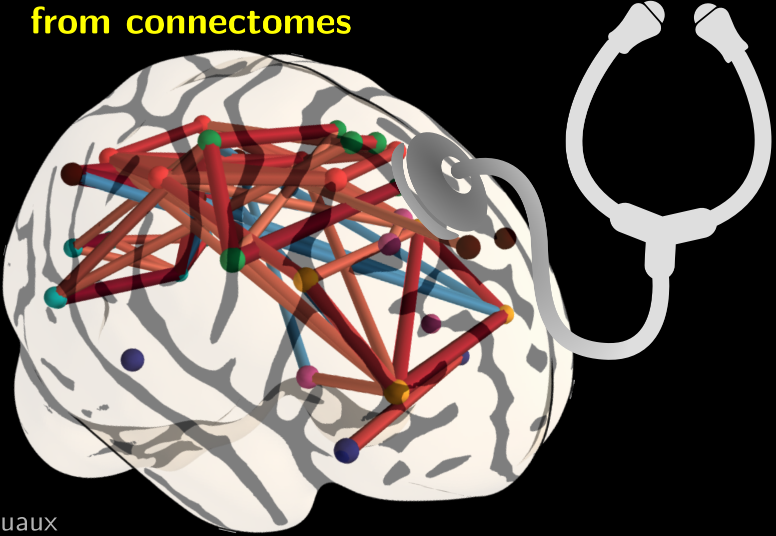
$$\mathbf{\Sigma}_{\text{Shrunk}} = (1-\lambda)\mathbf{\Sigma}_{\text{MLE}} + \lambda \mathbf{Id}$$

- Ledoit-Wolf oracle to set λ

[Ledoit and Wolf 2004]

3 Comparing connectomes

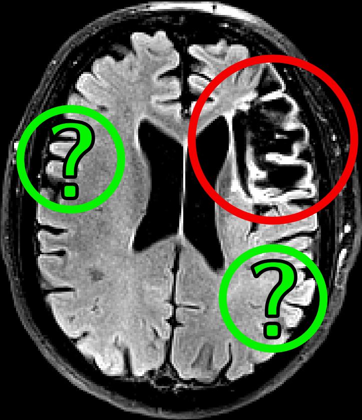
from connectomes



Detecting differences in connectivity

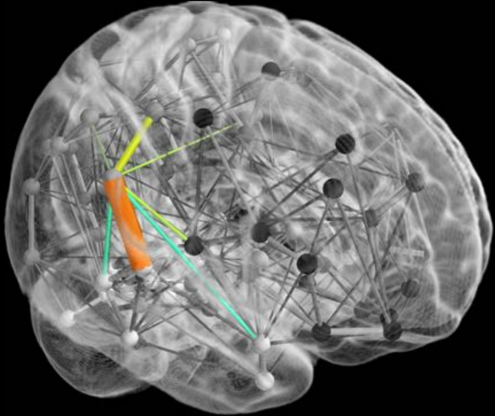
Functional markers on diminished patients?

Stroke outcome prognosis in ongoing activity



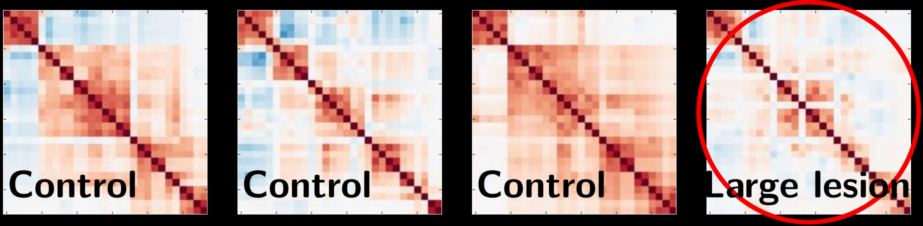
?

⇔



[Varoquaux... 2010a]

- Subject variability spread across correlation matrices

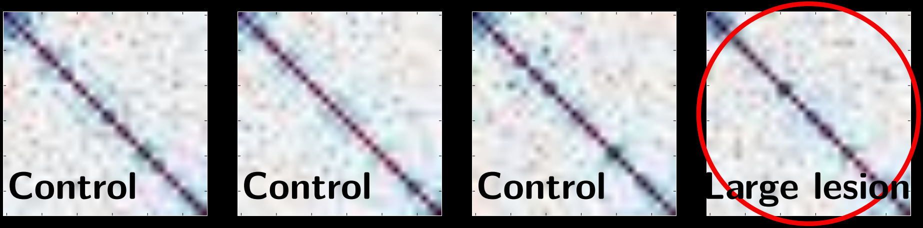


- $\mathbf{d\Sigma} = \mathbf{\Sigma}_2 - \mathbf{\Sigma}_1$ is not definite positive
  $\Rightarrow$ contradictory with Gaussian models

$\mathbf{\Sigma}$ does not live in a vector space

3 Inverse covariance very noisy

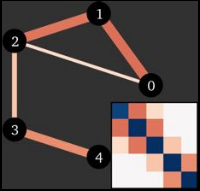
- Partial correlations are hard to estimate



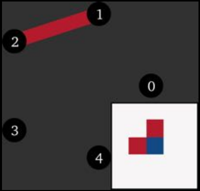
3 Simulation on a toy problem

- Simulate two processes with different inverse covariance

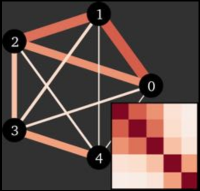
$\mathbf{K}_1$:



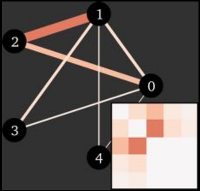
$\mathbf{K}_1 - \mathbf{K}_2$:



$\mathbf{\Sigma}_1$:

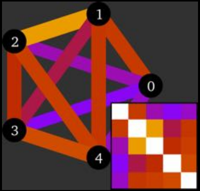


$\mathbf{\Sigma}_1 - \mathbf{\Sigma}_2$:

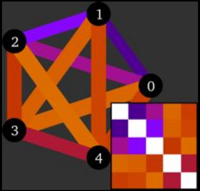


- Add jitter in observed covariance... sample

$\mathrm{MSE}(\mathbf{K}_1 - \mathbf{K}_2)$:



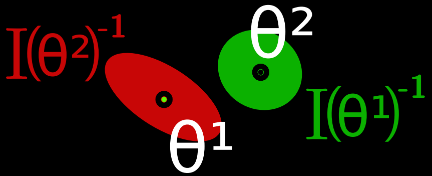
$\mathrm{MSE}(\mathbf{\Sigma}_1 - \mathbf{\Sigma}_2)$:



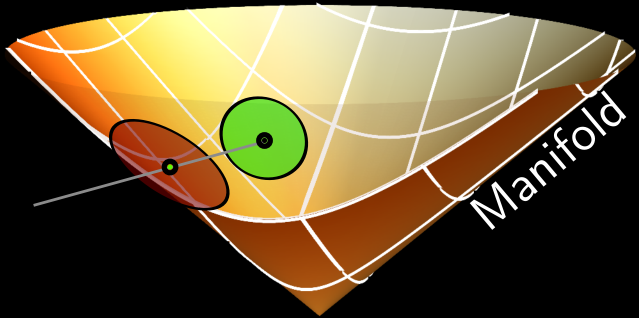
Non-local effects and non homogeneous noise

3 Theoretical settings: comparison of estimates

- Observations in 2 populations: $\mathbf{X}^1$ and $\mathbf{X}^2$
- Goal: comparing estimates: $\hat{\theta}(\mathbf{X}^1)$ and $\hat{\theta}(\mathbf{X}^1)$
- Asymptotic normality: $\hat{\theta}(\mathbf{X}^1) \sim \mathcal{N}\left(\theta^1, \mathbf{I}(\theta^1)^{-1}\right)$



- [Rao 1945] Fisher information **I** defines a metric on the manifold of models.
- We use it to choose a global parametrization for comparisons



3 Covariance manifold – $\mathcal{S}ym_n^+$

- Metric tensor (Fisher information) [Lenglet... 2006]

$$\langle \mathbf{d\Sigma}_1, \mathbf{d\Sigma}_2 \rangle_{\mathbf{\Sigma}} = \frac{1}{2}\text{trace}(\mathbf{\Sigma}^{-1}\, \mathbf{d\Sigma}_1\, \mathbf{\Sigma}^{-1}\, \mathbf{d\Sigma}_2)$$

- Nice properties of the $\mathcal{S}ym_n^+$ manifold (Lie group): metric can be fully integrated, gives rise to global mapping to a vector space (*Logarithmic map*).

- $\left\|\mathbf{\Sigma}_1, \mathbf{\Sigma}_2\right\|_{\mathbf{\Sigma}_1}^2 = \left\|\log\left(\mathbf{\Sigma_1}^{-\frac{1}{2}}\mathbf{\Sigma_2}\mathbf{\Sigma_1}^{-\frac{1}{2}}\right)\right\|^2,$

- Locally: $\left\|\mathbf{\Sigma}_1, \mathbf{\Sigma}_2\right\|_{\mathbf{\Sigma}_1} \propto \left|\text{trace}(\mathbf{\Sigma_1}^{-\frac{1}{2}}\mathbf{\Sigma_2}\mathbf{\Sigma_1}^{-\frac{1}{2}}) - p\right|$

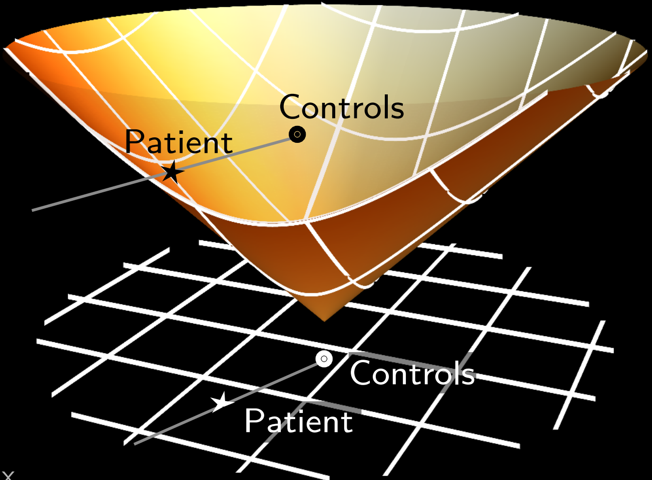
$$= \left\|\mathbf{d\Sigma}\right\|_{\text{Fro}}$$

where $\quad \mathbf{d\Sigma} = \mathbf{\Sigma}_1^{-1/2}\, \mathbf{\Sigma}_2\, \mathbf{\Sigma}_1^{-1/2}$

3 Reparametrization for uniform error geometry

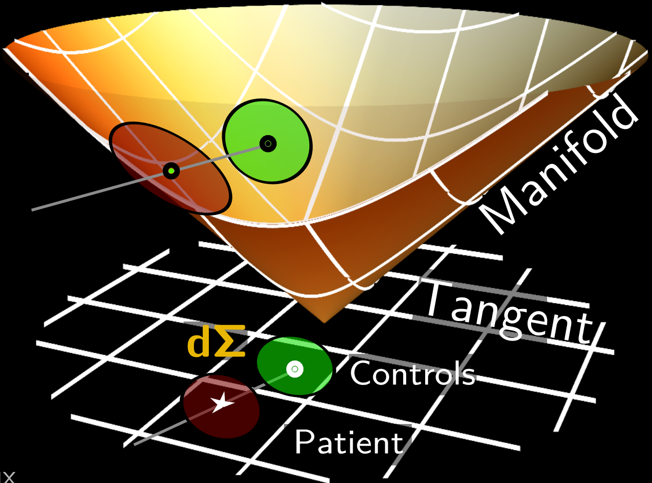
- Logarithmic map:

$$\boldsymbol{\Sigma}_1 \in \mathcal{S}ym_n^+ \; \boldsymbol{\Sigma}_2 \in \mathcal{S}ym_n^+ \rightarrow \overrightarrow{\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_2} \in \mathbb{R}^{\frac{1}{2}p\,(p-1)}$$



- Logarithmic map:

$$\mathbf{\Sigma}_1 \in \mathcal{S}ym_n^+ \; \mathbf{\Sigma}_2 \in \mathcal{S}ym_n^+ \rightarrow \overrightarrow{\mathbf{\Sigma}_1\mathbf{\Sigma}_2} \in \mathbb{R}^{\frac{1}{2}p\,(p-1)}$$
$$d(\mathbf{\Sigma}_1, \mathbf{\Sigma}_2) = \|\overrightarrow{\mathbf{\Sigma}_1\mathbf{\Sigma}_2}\|_2$$



Do *intrinsic* statistics on the parameterization:

- PDF
- Mean
- Parameter-level hypothesis testing

- Population covariance distribution: generalized normal

$$p(\mathbf{\Sigma}) \propto \exp\left(-\frac{1}{2\sigma^2}\|\mathbf{\Sigma}^{\star}\mathbf{\Sigma}\|^2_{\mathbf{\Sigma}^{\star}}\right) \qquad (1)$$

- Population mean: (Frechet mean)

$$\mathbf{\Sigma}^{\star} = \underset{\mathbf{\Sigma}}{\operatorname{argmin}} \sum_i \|\mathbf{\Sigma}\mathbf{\Sigma}_i\|^2_{\mathbf{\Sigma}} \qquad (2)$$

[Pennec 2006]

Edge-level statistics

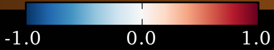
- H_0: subject $\in$ group model (1)

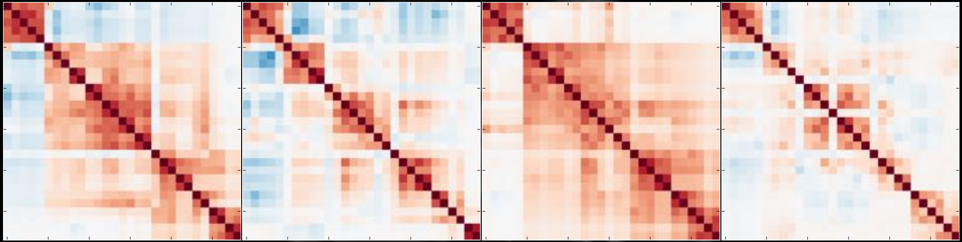
$\overrightarrow{\mathbf{d\Sigma}} \sim \mathcal{N}(0, \sigma\mathbf{I})$: Independant coefficients

$\Rightarrow$ **Univariate statistics on $\mathbf{d\Sigma}_{i,j}$**

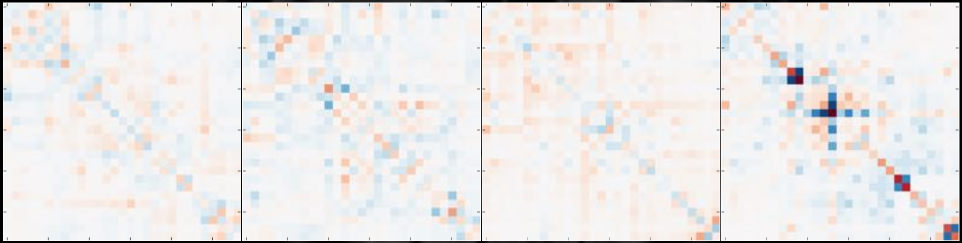
[Varoquaux... 2010a]

Correlation matrices: Σ



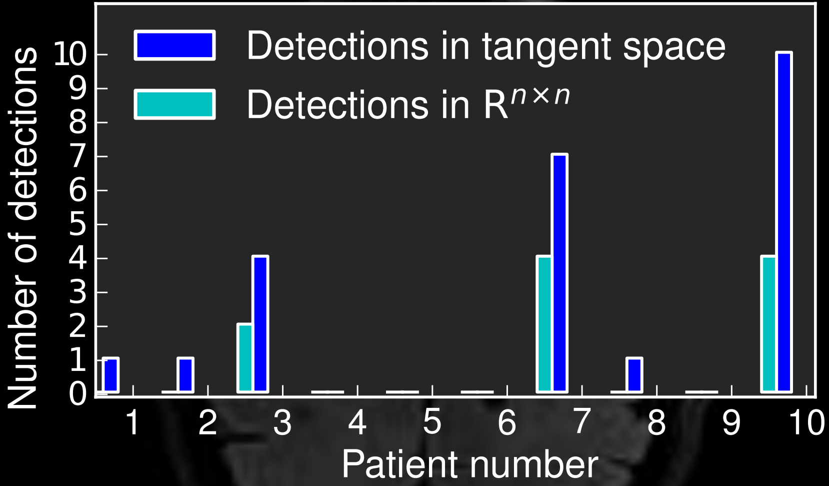


Residuals: $d\Sigma$



Control | Control | Control | Large lesion

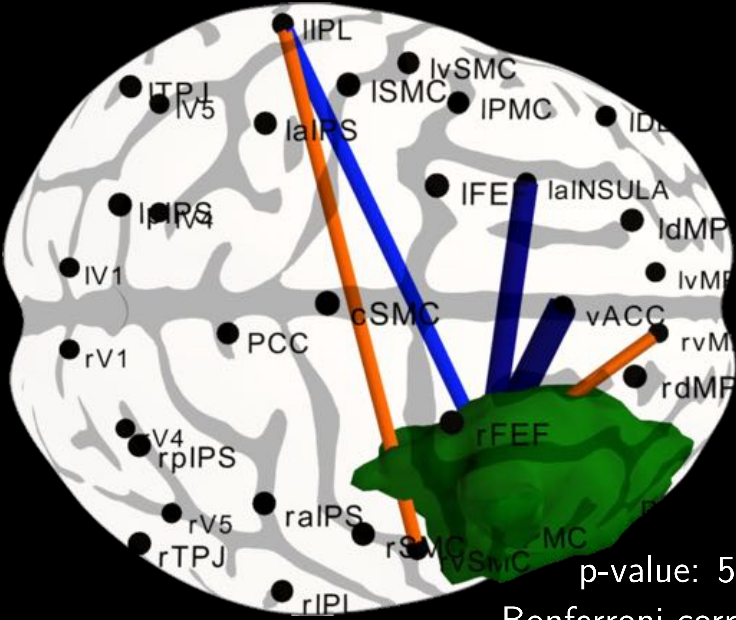
3 Number of edge-level differences detected



p-value: $5{\cdot}10^{-2}$

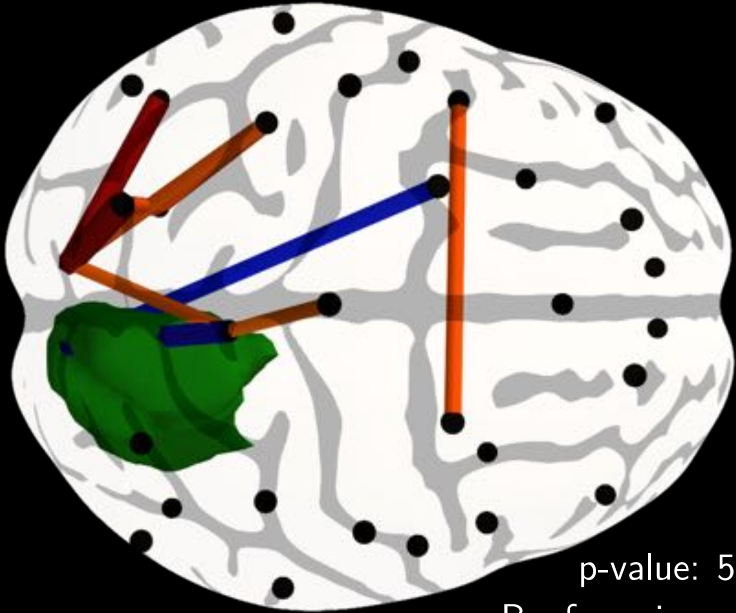
Bonferroni-corrected

3 Post-stroke covariance modifications

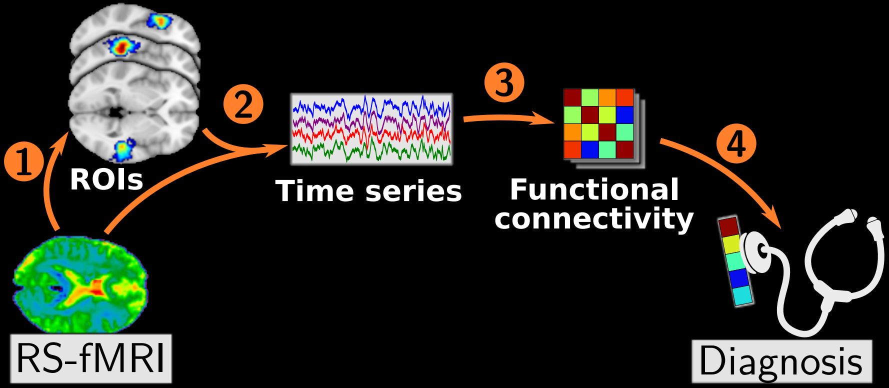


p-value: $5{\cdot}10^{-2}$

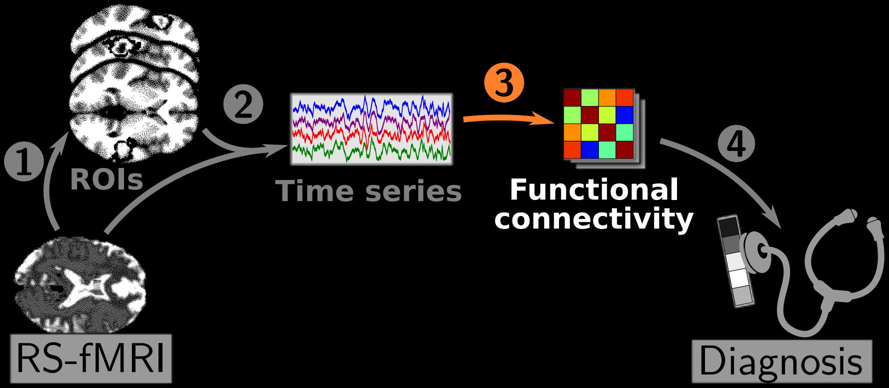
Bonferroni-corrected



p-value: $5{\cdot}10^{-2}$
Bonferroni-corrected

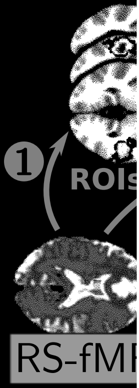

1
ROIs
2
Time series
3
Functional connectivity
4
RS-fMRI
Diagnosis

3 In connectome prediction settings

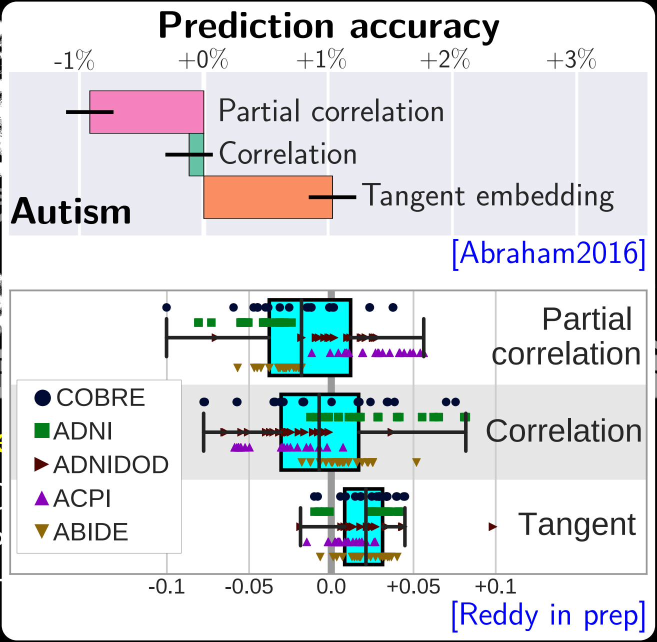


Connectivity matrix

- Correlation
- Partial correlations
- Tangent space


1
ROIs
RS-fMR


nosis

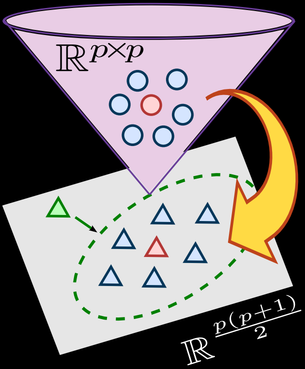

Prediction accuracy
-1%
+0%
+1%
+2%
+3%
Partial correlation
Correlation
Tangent embedding
Autism
[Abraham2016]
COBRE
ADNI
ADNIDOD
ACPI
ABIDE
Partial correlation
Correlation
Tangent
-0.1
-0.05
0.0
+0.05
+0.1
[Reddy in prep]

Conne

- C
- P
- T

- Dispersion of covariances in tangent space
- James-Stein shrinkage using this population model

$\Rightarrow$ Gives better biomarkers



Covariance space
- empirical covariance
- mean of covariances

Tangent space
- covariance embedding
- reference (mean)

Shrinkage
- population dispersion (covariance)
- shrinkage of a new estimate

[Rahim... 2017]

Statistics on covariance matrices

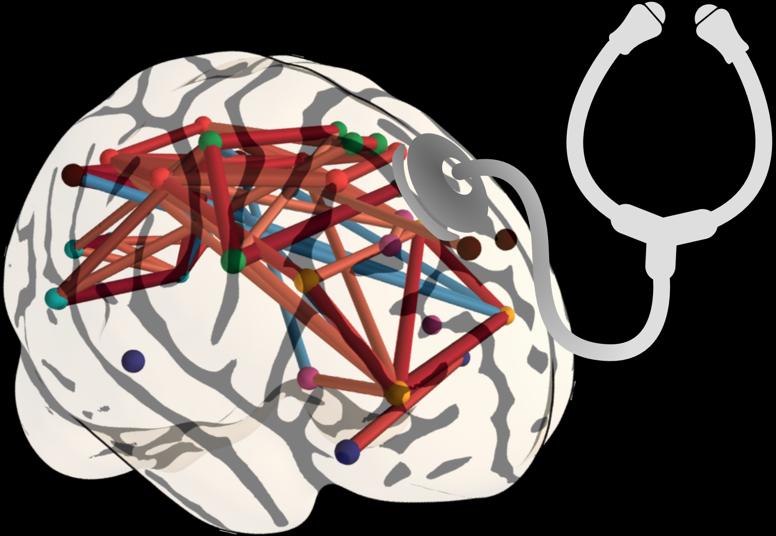
Do not live in vector space:

$\Rightarrow$ coefficients are not independent

Are a multivariate model

$\Rightarrow$ can be reparametrized with Cramer-Rao metric

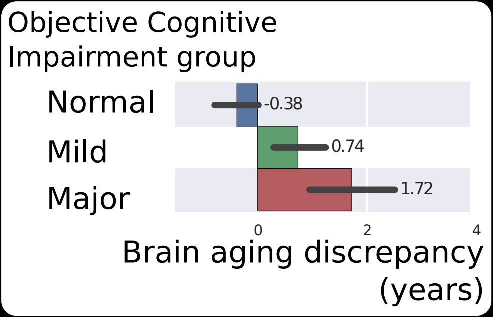
Population imaging and biomarkers



Predicting brain aging $\neq$ chronological age

- Combines brain connectivity and morphology
- Predicts age with a mean absolute error of **4.3 years**

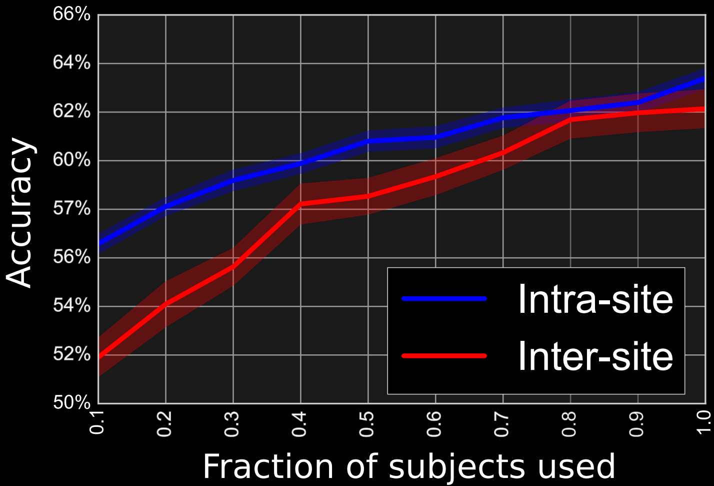
- Discrepency with chronological age correlates with cognitive impairment



[Liem... 2016]

Biomarker
surrogate,
but useful

Heterogeneity: predicting autism across sites



More data is better (up to 1000 subjects)

[Abraham... 2016]

Software

Nilearn: neuroimaging

`http://nilearn.github.io`

- Extracting signal in brain images
- Simple visualizations
- Extracting connectomes
- Learning networks and regions

Very easy to install and to script

Software

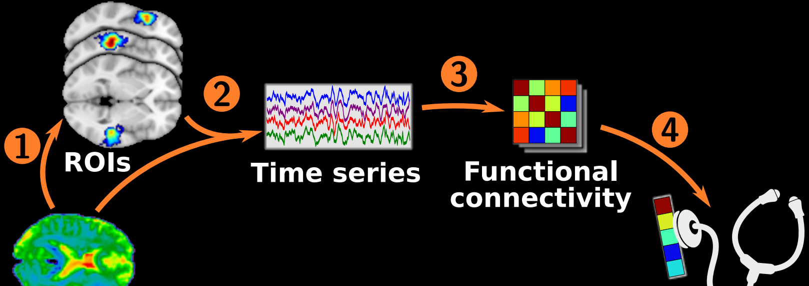
Scikit-learn: machine learning

http://scikit-learn.org

- Supervised & unsupervised learning
- $>$ 160 models
- Sparse models, random forests, clustering...
- Model selection, parallel computing

Excellent documentation

Connectomics: from mapping intrinsic activity to predicting phenotype



- Region extractions
  ⇒ Dictionary learning

- Comparing connectomes
  ⇒ Tangent space

@GaelVaroquaux

References I

A. Abraham, E. Dohmatob, B. Thirion, D. Samaras, and G. Varoquaux. Extracting brain regions from rest fMRI with total-variation constrained dictionary learning. In *MICCAI*, page 607. 2013.

A. Abraham, M. Milham, A. Di Martino, R. C. Craddock, D. Samaras, B. Thirion, and G. Varoquaux. Deriving reproducible biomarkers from multi-site resting-state data: An autism-based example. *NeuroImage*, 2016.

C. F. Beckmann and S. M. Smith. Probabilistic independent component analysis for functional magnetic resonance imaging. *Trans Med Im*, 23:137, 2004.

B. Biswal, F. Zerrin Yetkin, V. Haughton, and J. Hyde. Functional connectivity in the motor cortex of resting human brain using echo-planar MRI. *Magn Reson Med*, 34: 53719, 1995.

References II

R. C. Craddock, G. A. James, P. E. Holtzheimer, X. P. Hu, and H. S. Mayberg. A whole brain fMRI atlas generated via spatially constrained spectral clustering. *Human brain mapping*, 33:1914, 2012.

V. Doria, C. F. Beckmann, T. Arichi, N. Merchant, M. Groppo, F. E. Turkheimer, S. J. Counsell, M. Murgasova, P. Aljabar, R. G. Nunes, ... Emergence of resting state networks in the preterm human brain. *Proc Ntl Acad Sci*, 107:20015, 2010.

V. M. Eguiluz, D. R. Chialvo, G. A. Cecchi, M. Baliki, and A. V. Apkarian. Scale-free brain functional networks. *Physical review letters*, 94:018102, 2005.

J. Friedman, T. Hastie, and R. Tibshirani. Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, 9:432, 2008.

References III

M. Greicius, B. Krasnow, A. Reiss, and V. Menon. Functional connectivity in the resting brain: a network analysis of the default mode hypothesis. *Proceedings of the National Academy of Sciences*, 100:253, 2003.

T. Kenet, D. Bibitchkov, M. Tsodyks, A. Grinvald, and A. Arieli. Spontaneously emerging cortical representations of visual attributes. *Nature*, 425:954, 2003.

V. Kiviniemi, J. Kantola, J. Jauhiainen, A. Hyvärinen, and O. Tervonen. Independent component analysis of nondeterministic fmri signal sources. *Neuroimage*, 19:253, 2003.

V. Kiviniemi, T. Starck, J. Remes, X. Long, J. Nikkinen, M. Haapea, J. Veijola, ... Functional segmentation of the brain cortex using high model order group PICA. *Hum Brain Map*, 30:3865, 2009.

References IV

C. Lam and J. Fan. Sparsistency and rates of convergence in large covariance matrix estimation. *Annals of statistics*, 37: 4254, 2009.

O. Ledoit and M. Wolf. A well-conditioned estimator for large-dimensional covariance matrices. *J. Multivar. Anal.*, 88:365, 2004.

C. Lenglet, M. Rousson, R. Deriche, and O. Faugeras. Statistics on the manifold of multivariate normal distributions: Theory and application to diffusion tensor MRI processing. *Journal of Mathematical Imaging and Vision*, 25:423, 2006.

C. Lewis, A. Baldassarre, G. Committeri, G. Romani, and M. Corbetta. Learning sculpts the spontaneous activity of the resting human brain. *Proceedings of the National Academy of Sciences*, 106(41):17558, 2009.

References V

F. Liem, G. Varoquaux, J. Kynast, F. Beyer, S. K. Masouleh, J. M. Huntenburg, L. Lampe, M. Rahim, A. Abraham, R. C. Craddock, ... Predicting brain-age from multimodal imaging data captures cognitive impairment. *NeuroImage*, 2016.

J. Mairal, F. Bach, J. Ponce, and G. Sapiro. Online learning for matrix factorization and sparse coding. *Journal of Machine Learning Research*, 11:19, 2010.

X. Pennec. Intrinsic statistics on Riemannian manifolds: Basic tools for geometric measurements. *Journal of Mathematical Imaging and Vision*, 25:127, 2006.

M. Rahim, B. Thirion, and G. Varoquaux. Population-shrinkage of covariance to estimate better brain functional connectivity. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 460–468. Springer, 2017.

References VI

M. Raichle, A. MacLeod, A. Snyder, W. Powers, D. Gusnard, and G. Shulman. A default mode of brain function. *Proceedings of the National Academy of Sciences*, 98:676, 2001.

M. E. Raichle and M. A. Mintun. Brain work and brain imaging. *Annu. Rev. Neurosci.*, 29:449–476, 2006.

C. Rao. Information and accuracy attainable in the estimation of statistical parameters. *Bull. Calcutta Math. Soc.*, 37:81, 1945.

P. Ravikumar, M. J. Wainwright, G. Raskutti, B. Yu, ... High-dimensional covariance estimation by minimizing ℓ_1-penalized log-determinant divergence. *Electronic Journal of Statistics*, 5:935–980, 2011.

References VII

S. Smith, K. Miller, G. Salimi-Khorshidi, M. Webster, C. Beckmann, T. Nichols, J. Ramsey, and M. Woolrich. Network modelling methods for fMRI. *Neuroimage*, 54:875, 2011.

E. A. Stamatakis, R. M. Adapa, A. R. Absalom, and D. K. Menon. Changes in resting neural connectivity during propofol sedation. *PLoS One*, 5:e14224, 2010.

B. Thirion, G. Varoquaux, E. Dohmatob, and J. Poline. Which fMRI clustering gives good brain parcellations? *Name: Frontiers in Neuroscience*, 8:167, 2014.

G. Varoquaux and R. C. Craddock. Learning and comparing functional connectomes across subjects. *NeuroImage*, 80: 405, 2013.

References VIII

G. Varoquaux, F. Baronnet, A. Kleinschmidt, P. Fillard, and B. Thirion. Detection of brain functional-connectivity difference in post-stroke patients using group-level covariance modeling. In *MICCAI*. 2010a.

G. Varoquaux, A. Gramfort, J. B. Poline, and B. Thirion. Brain covariance selection: better individual functional connectivity models using population prior. In *NIPS*. 2010b.

G. Varoquaux, S. Sadaghiani, P. Pinel, A. Kleinschmidt, J. B. Poline, and B. Thirion. A group model for stable multi-subject ICA on fMRI datasets. *NeuroImage*, 51:288, 2010c.

G. Varoquaux, A. Gramfort, F. Pedregosa, V. Michel, and B. Thirion. Multi-subject dictionary learning to segment an atlas of brain spontaneous activity. In *Inf Proc Med Imag*, page 562, 2011.

References IX

G. Varoquaux, A. Gramfort, J. B. Poline, and B. Thirion. Markov models for fMRI correlation structure: is brain functional connectivity small world, or decomposable into networks? *Journal of Physiology - Paris*, 106:212, 2012.