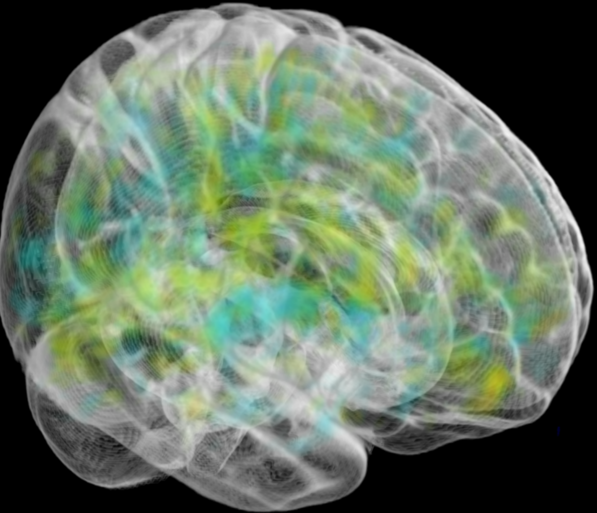
Graphical models of brain function across individuals



Gaël Varoquaux

INSERM U992, Unicog
INRIA, Parietal

Joint work with:

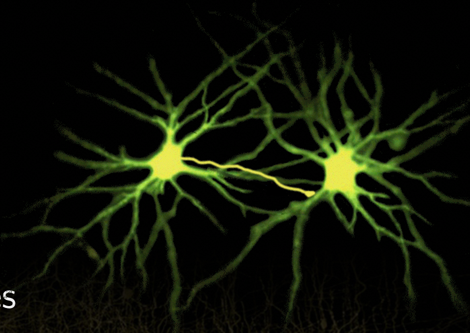
Bertrand Thirion
Andreas Kleinschmidt
Alexandre Gramfort
Pierre Fillard

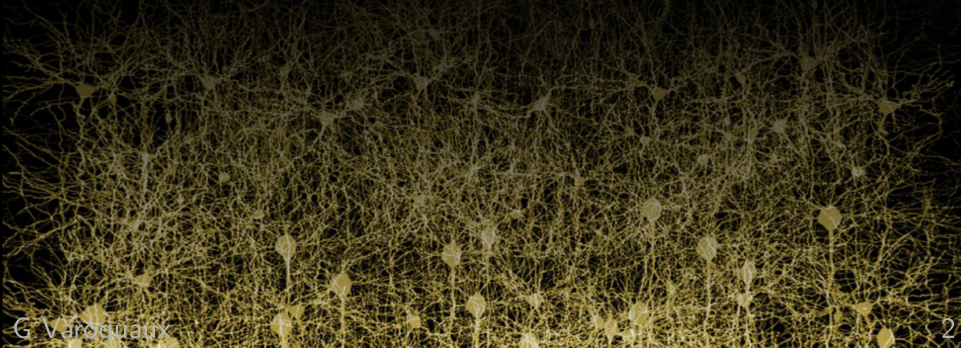
NeuroSpin INRIA Inserm

Neural networks

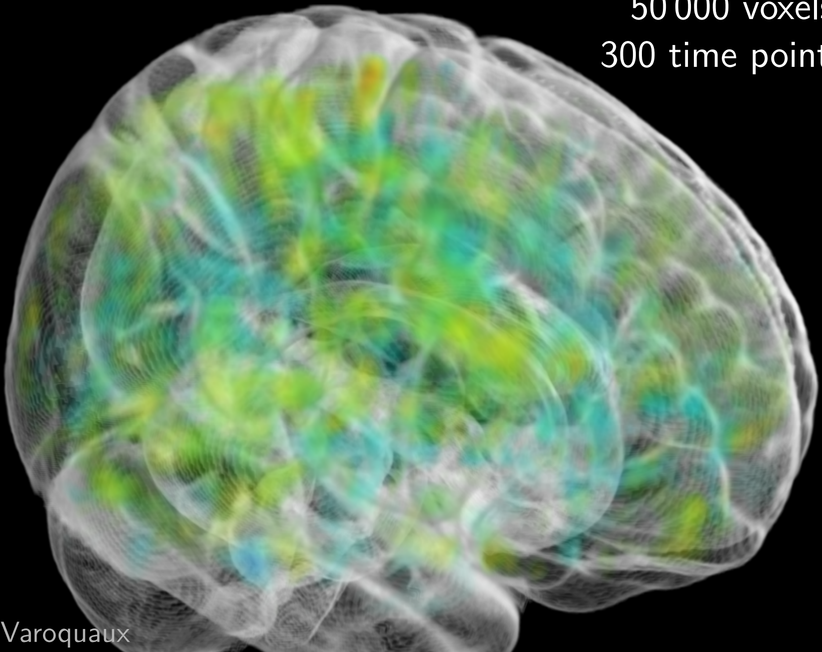
- [Watts and Strogatz 1998]
  Small world properties in
  C. Elegans neural network

- Human brain:
  10^{11} neurons, 10^{16} synapses



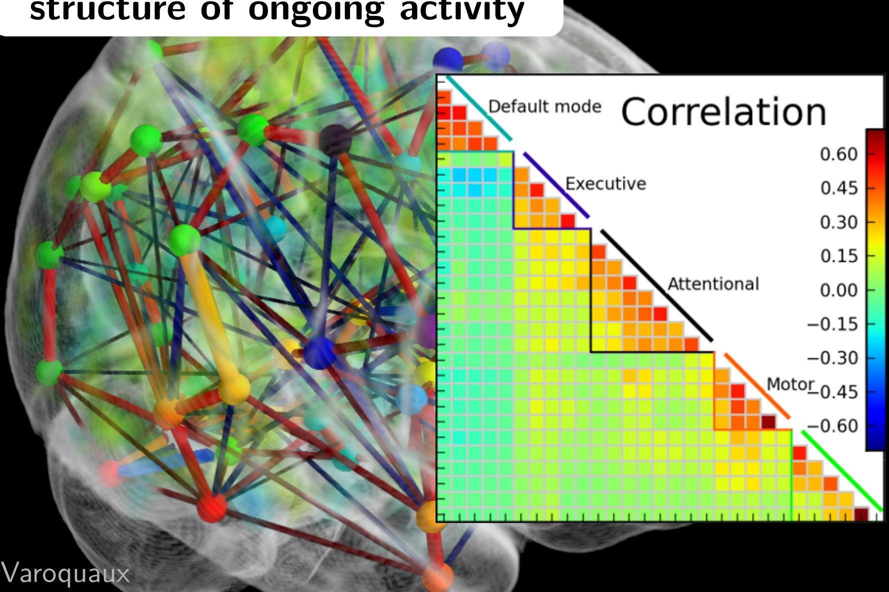


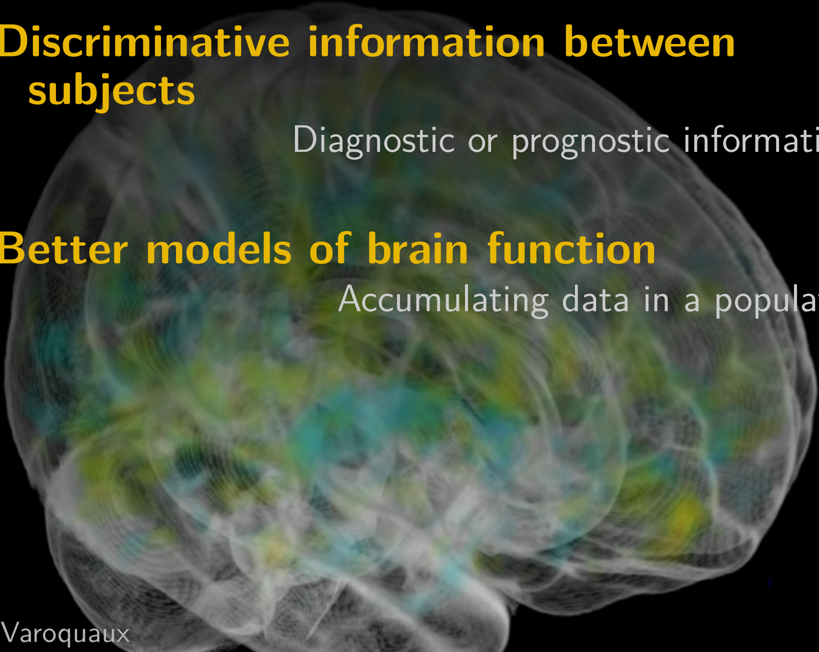
Functional brain imaging



50 000 voxels,
300 time points

Modeling the correlation structure of ongoing activity





Discriminative information between subjects

Diagnostic or prognostic information?

Better models of brain function

Accumulating data in a population

Presentation outline

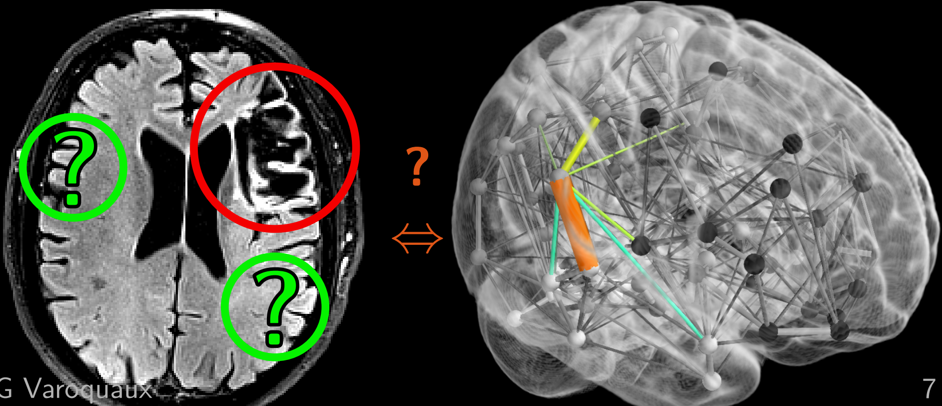
1. **Detecting differences in connectivity**

2. **Individual models with population data**

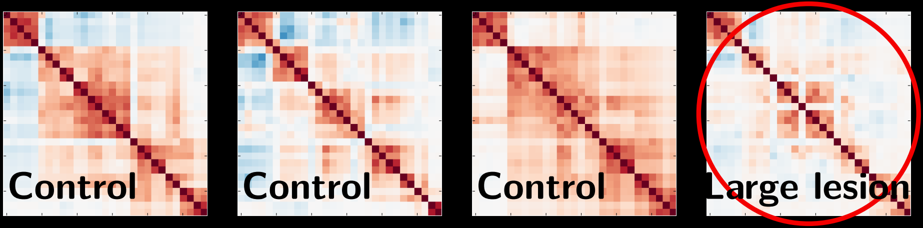
1 Detecting differences in connectivity

Functional markers on diminished patients?

Stroke outcome prognosis in ongoing activity



- Subject variability spread across correlation matrices



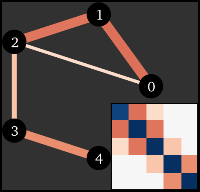
- $\mathbf{d\Sigma} = \mathbf{\Sigma}_2 - \mathbf{\Sigma}_1$ is not definite positive
  $\Rightarrow$ contradictory with Gaussian models

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

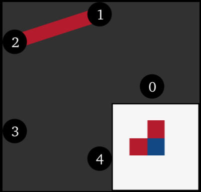
1 Simulation on a toy problem

- Simulate two processes with different inverse covariance

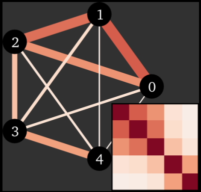
$\mathbf{K}_1$:



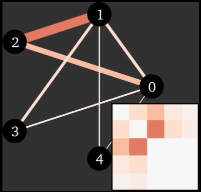
$\mathbf{K}_1 - \mathbf{K}_2$:



$\mathbf{\Sigma}_1$:

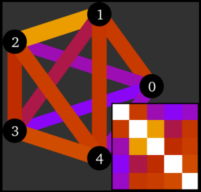


$\mathbf{\Sigma}_1 - \mathbf{\Sigma}_2$:

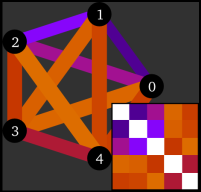


- Add jitter in observed covariance... sample

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



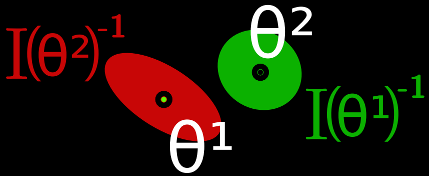
MSE($\mathbf{\Sigma}_1 - \mathbf{\Sigma}_2$):



Non-local effects and non homogeneous noise

1 Theoretical settings: comparison of estimates

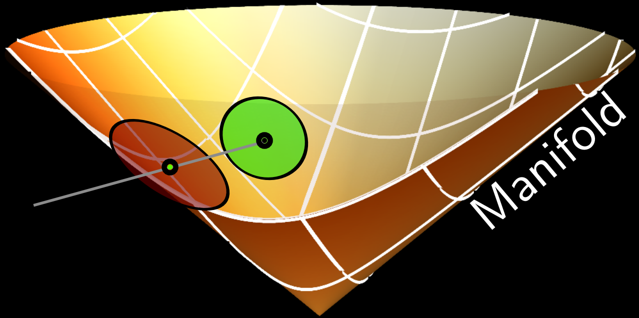
- Observations in 2 populations: $\mathbf{X}^1$ and $\mathbf{X}^2$
- Goal: comparing estimates: $\hat{\theta}(\mathbf{X}^1)$ and $\hat{\theta}(\mathbf{X}^1)$
- Asymptotic normality: $\hat{\theta}(\mathbf{X}^1) \sim \mathcal{N}\left(\theta^1, \mathbf{I}(\theta^1)^{-1}\right)$



1 Theoretical settings: comparison of estimates

- [Rao 1945] Fisher information **I** defines a metric on the manifold of models.

- We use it to choose a global parametrization for comparisons



1 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*).

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

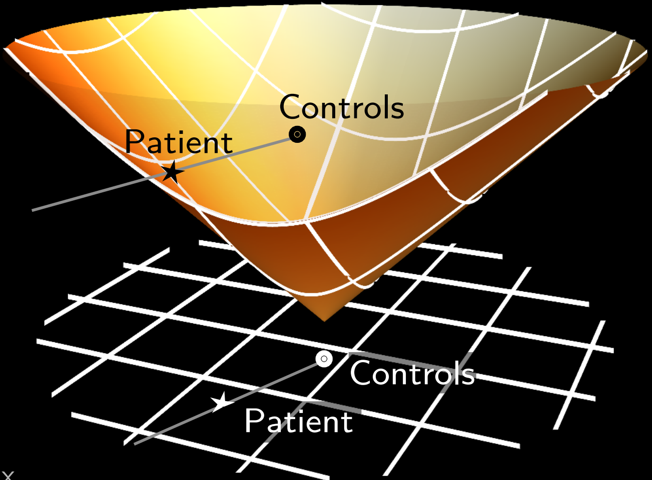
- Locally: $\left\|\mathbf{\Sigma}_1, \mathbf{\Sigma}_2\right\|_{\mathbf{\Sigma}_1} \propto \left|\text{trace}(\mathbf{\Sigma_1}^{-\frac{1}{2}}\mathbf{\Sigma_2}\mathbf{\Sigma_1}^{-\frac{1}{2}}) - p\right|$
$= \left\|\mathbf{d\Sigma}\right\|_{\text{Fro}}$

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

1 Reparametrization for uniform error geometry

- Logarithmic mapping:

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

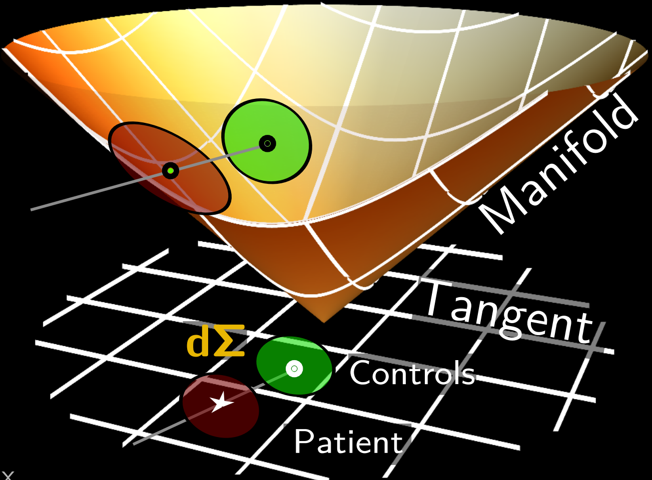


1 Reparametrization for uniform error geometry

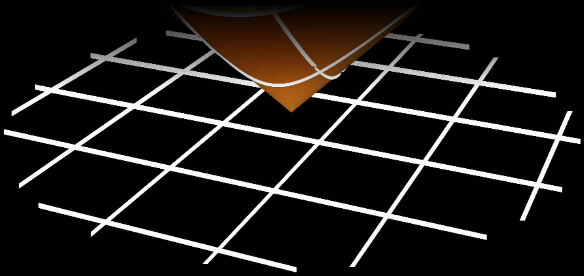
- Logarithmic mapping:

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

$$d(\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2) = \|\overrightarrow{\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_2}\|_2$$



1 Statistics...



Do *intrinsic* statistics on the parameterization:

- Mean (Frechet mean)
- PDF
- Parameter-level hypothesis testing

Population-level covariance distribution

- Generalized isotropic normal distribution:

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

- Population mean:

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

Efficient gradient descent algorithm

Principled computation of:

- **group mean $\boldsymbol{\Sigma}^{\star}$ and spread σ**
- **likelihood of new data**

Population-level covariance distribution

- Generalized isotropic normal distribution:

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

Edge-level statistics

- Under null hypothesis: subject $\in$ group model (1)

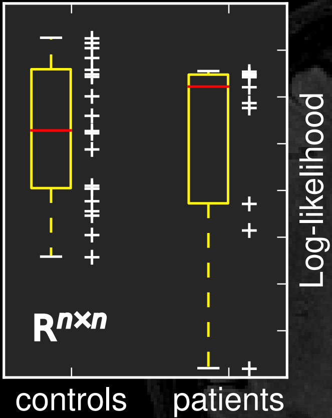
$$\overrightarrow{\mathbf{d\Sigma}} \sim \mathcal{N}(0, \sigma \mathbf{I}) \text{ : Independant coefficients}$$

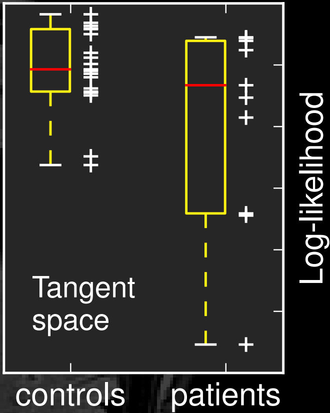
$\Rightarrow$ **Univariate statistics on $\mathbf{d\Sigma}_{i,j}$**

[Varoquaux 2010]

1 Discriminating strokes patients from controls

Leave one out likelihood





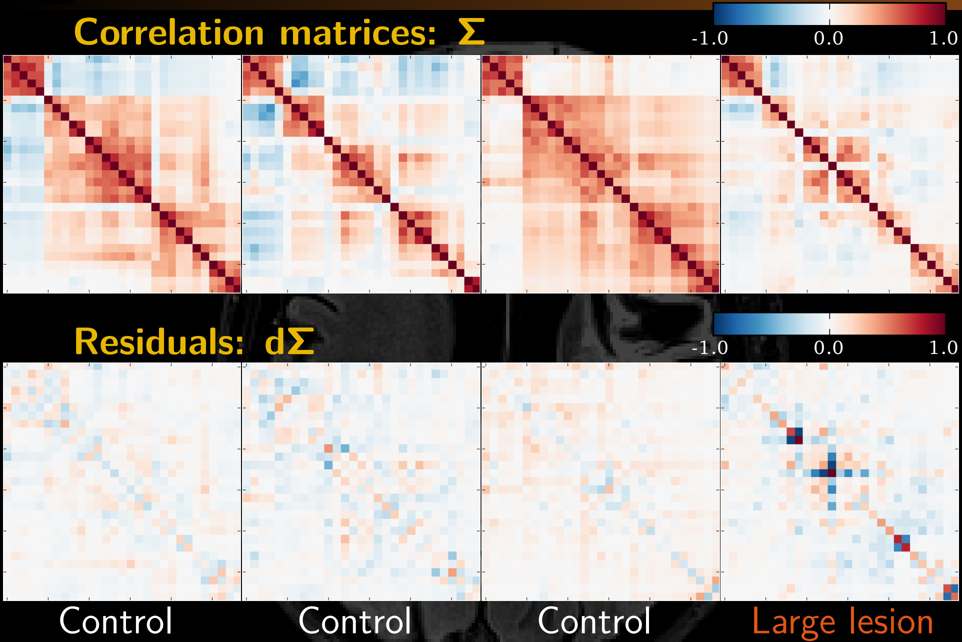
Probabilistic model on manifold discriminates patients better

20 controls, 10 patients

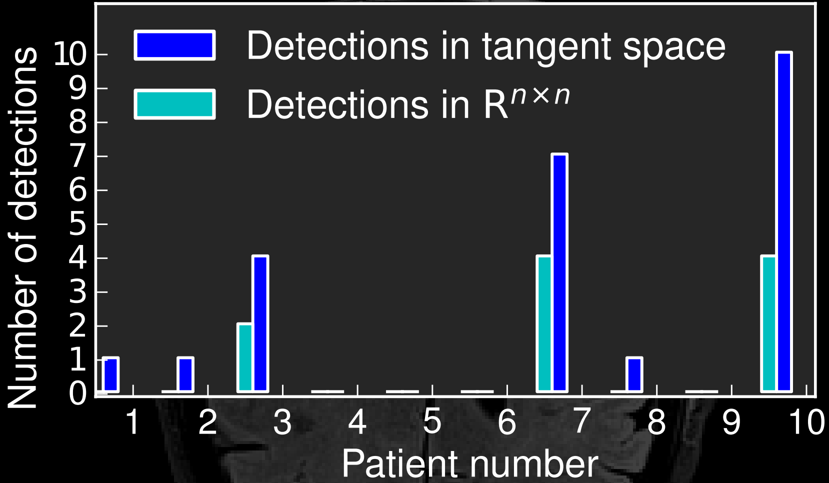
1 Residuals

Correlation matrices: Σ

Residuals: $d\Sigma$



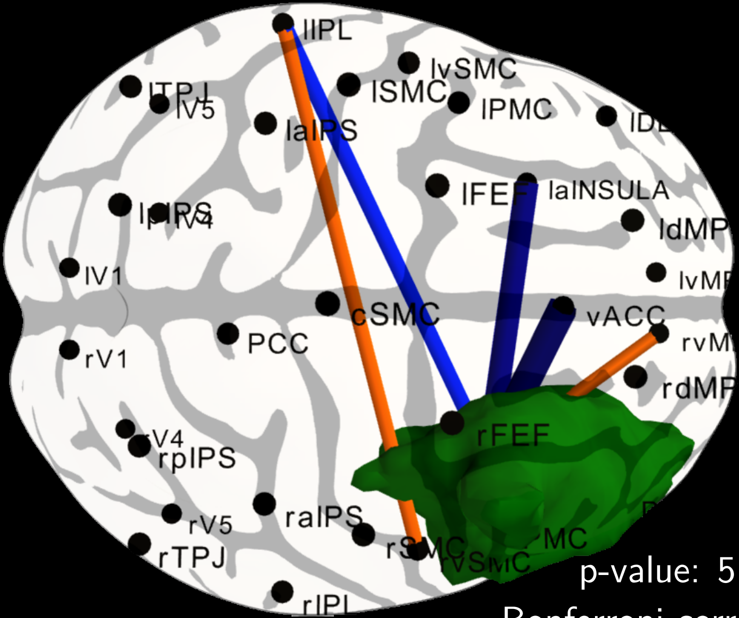
1 Number of edge-level differences detected



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

Bonferroni-corrected

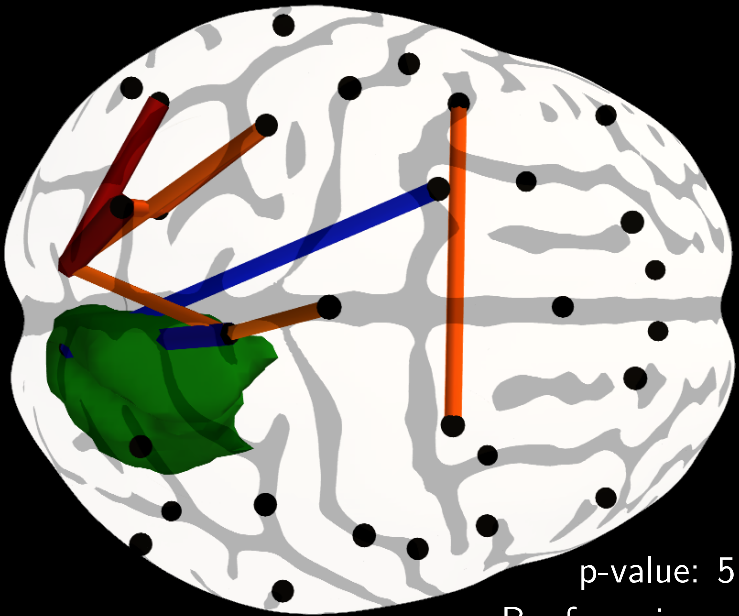
1 Post-stroke covariance modifications



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

Bonferroni-corrected

1 Post-stroke covariance modifications



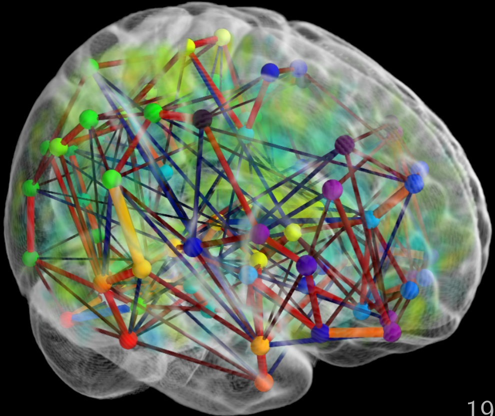
p-value: $5{\cdot}10^{-2}$

Bonferroni-corrected

2 Individual models with population data

Sample complexity of covariance estimation: p^2

$\Rightarrow$ for large graphs, need to use multi-subject datasets to improve covariance estimation



Penalized sparse inverse covariance estimation

- Maximum a posteriori: fit models with a prior

$$\mathbf{K} = \underset{\mathbf{K} \succ 0}{\operatorname{argmax}} \; \mathcal{L}(\hat{\mathbf{\Sigma}}|\mathbf{K}) + f(\mathbf{K})$$

- Standard sparse inverse-covariance estimation:
  Prior: *many pairs of regions are not connected*

  Lasso-like problem:

  ℓ_1 penalization $\qquad f(\mathbf{K}) = \sum_{i \neq j} |\mathbf{K}_{i,j}|$

- Maximum a posteriori: fit models with a prior

$$\mathbf{K} = \underset{\mathbf{K} \succ 0}{\operatorname{argmax}} \; \mathcal{L}(\hat{\mathbf{\Sigma}}|\mathbf{K}) + f(\mathbf{K})$$

- Our contribution: **Population prior**:
  same independence structure across subjects

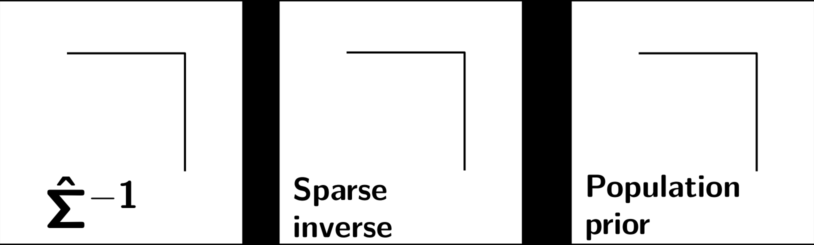
  $\Rightarrow$ Estimate together all $\{\mathbf{K}^s\}$ from $\{\hat{\mathbf{\Sigma}}^s\}$

  Group-lasso (mixed norms):

  ℓ_{21} penalization $\quad f\left(\{\mathbf{K}^s\}\right) = \lambda \sum_{i \neq j} \sqrt{\sum_s (\mathbf{K}^s_{i,j})^2}$

Convex optimization problem

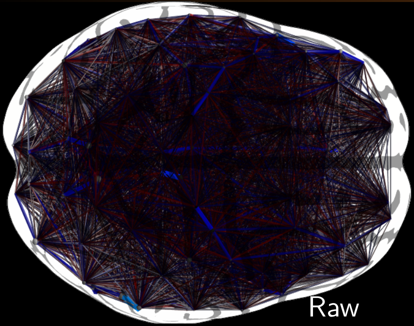
2 Population-sparse graph perform better



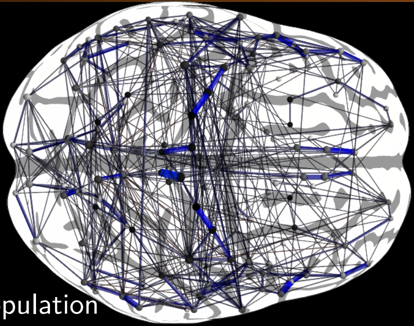
Likelihood of new data (nested cross-validation)

Subject data, Σ^{-1}	-57.1
Subject data, sparse inverse	43.0
Group average data, Σ^{-1}	40.6
Group average data, sparse inverse	41.8
Population prior	**45.6**

[Varoquaux 2010b]



Raw correlations



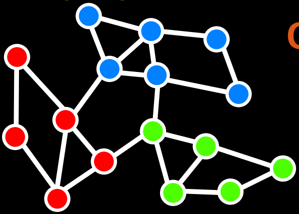
Population prior

2 Graphs of brain function?

Cognitive function arises from the interplay of specialized brain regions:

The functional segregation of local areas [...] contrasts sharply with their global integration during perception and behavior [Tononi 1994]

A proposed measure of functional segregation



Graph modularity =
divide in *communities* to
maximize intra-class connections
versus extra-class

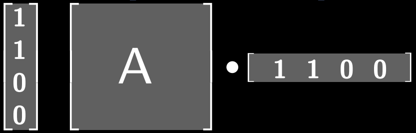
2 Graph cuts to isolate functional communities

- Find communities to maximize modularity:

$$Q = \sum_{c=1}^{k} \left(\frac{\mathcal{A}(V_c, V_c)}{\mathcal{A}(V, V)} - \left(\frac{\mathcal{A}(V, V_c)}{\mathcal{A}(V, V)} \right)^2 \right)$$

$\mathcal{A}(V_a, V_b)$ is the sum of edges going from V_a to V_b

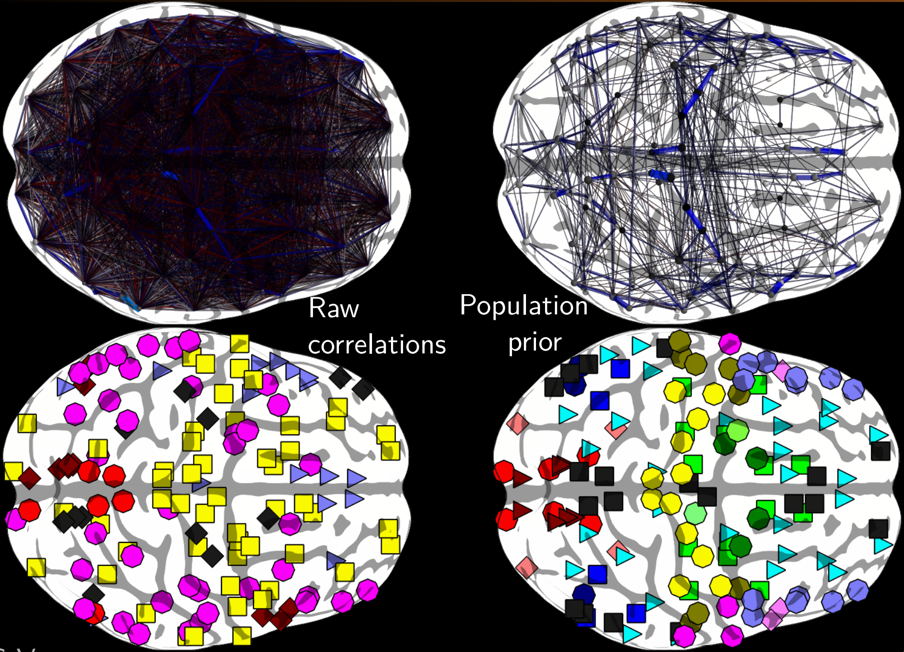
- Rewrite as an eigenvalue problem [White 2005]



⇒ Spectral clustering = spectral embedding + k-means

- Similar to normalized graph cuts

2 Brain graphs and communities



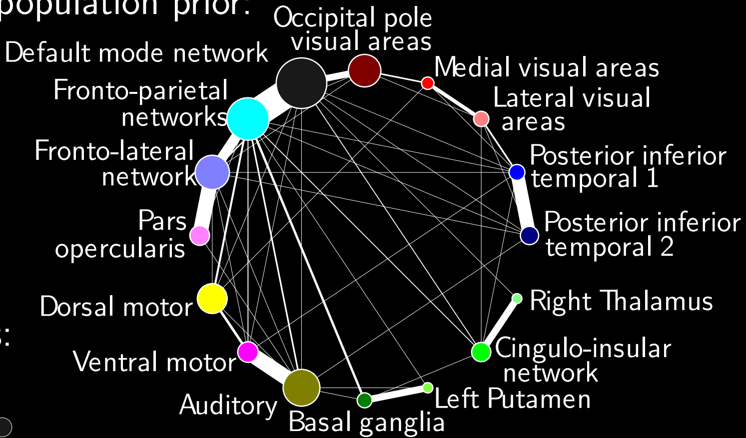
Proposed measure for functional integration: mutual information (Tononi)

$$\text{Integration:} \quad I_{c_1} = \frac{1}{2} \log \det(\mathbf{K}_{c_1})$$

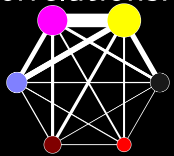
$$\text{Mutual information:} \quad M_{c_1,c_2} = I_{c_1 \cup c_2} - I_{c_1} - I_{s_2}$$

Proposed measure for functional integration: mutual information (Tononi)

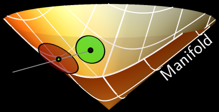
With population prior:



Raw correlations:

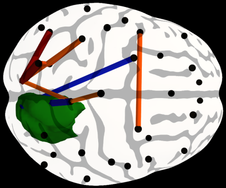


Estimating a family of brain functional graphs



Random model for covariance

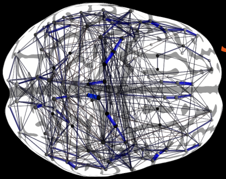
Suitable for inference: account for geometry of errors



Defines a discriminant **likelihood**

$$\mathbf{d\Sigma} \sim \mathbf{\Sigma}_{\text{Group}}^{-\frac{1}{2}}\, \mathbf{\Sigma}\, \mathbf{\Sigma}_{\text{Group}}^{-\frac{1}{2}}$$

Univariate statistics in **tangent space** identifies edges



Joint estimation of a family of graphs

Multi-subjects mapping of functional connectivity

Bibliography

- [Watts and Strogatz 1998]: D. Watts and S. Strogatz, *Collective dynamics of "small-world" networks*, Nature **393** p. 440 (1998)
- [Rao 1945]: C.R. Rao, *Information and accuracy attainable in the estimation of statistical parameters*, Bull. Calcutta Math. Soc. **37** p. 81 (1945)
- [Lenglet 2006] 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** p. 423 (2006)
- [Varoquaux 2010] 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*, MICCAI (2010)

Bibliography

- [Varoquaux 2010b] G. Varoquaux, A. Gramfort, J.B. Poline and B. Thirion, *Brain covariance selection: better individual functional connectivity models using population prior*, NIPS (2010)
- [Tononi 1994] G. Tononi, O. Sporns, G. Edelman, *A measure for brain complexity: relating functional segregation and integration in the nervous system*, PNAS, **91** p. 5033 (1994)
- [White 2005] S. White, P. and Smyth, *A spectral clustering approach to finding communities in graphs*, 5th SIAM international conference on data mining, p. 274 (2005)