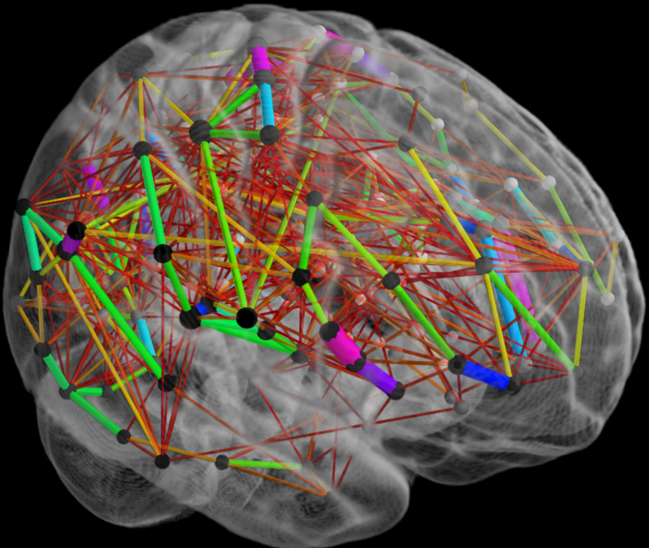
Learning graphical models of the brain

Gaël Varoquaux



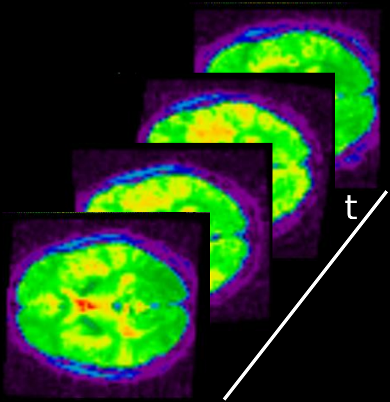
PARIETAL

Inria

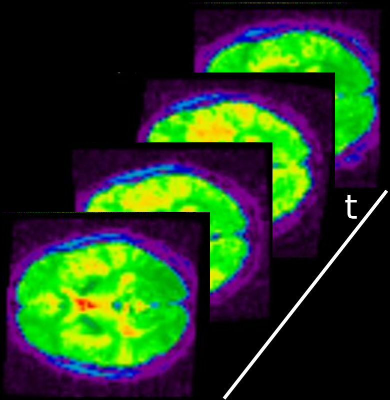
NeuroSpin

functional MRI (fMRI)

Recordings of brain activity



functional MRI (fMRI)



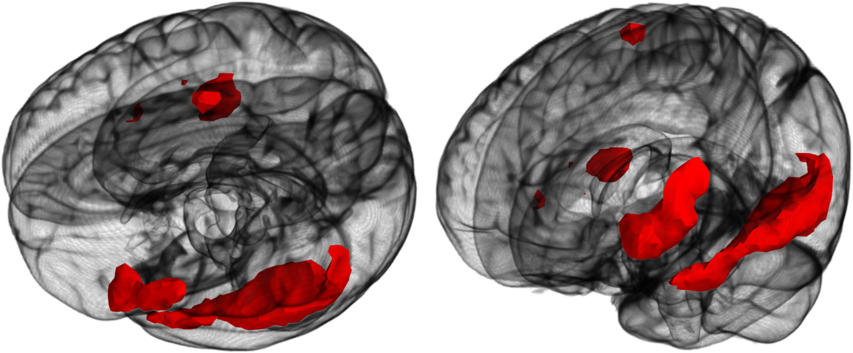
Recordings of brain activity

Brain mapping:

- the motor system: “move the right hand”
- the language system: “say three names of animals”

Brain mapping:

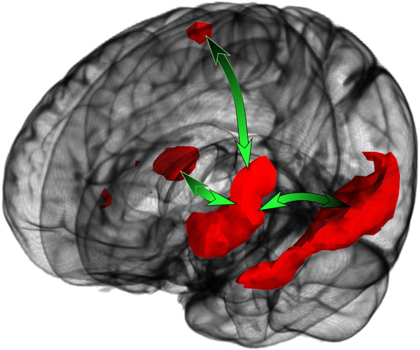
The language network



- the language system: “say three names of animals”

Brain mapping:

The language network



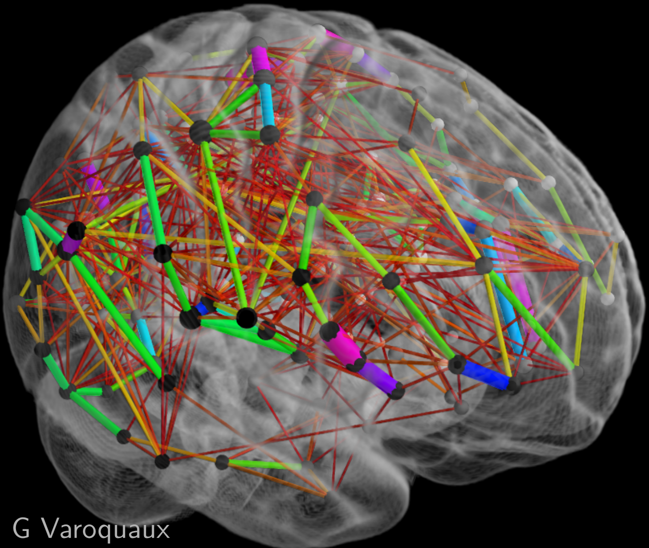
Interacting sub-systems:

- Sounds
- Lexical access
- Syntax

- the language system: "say three names of animals"

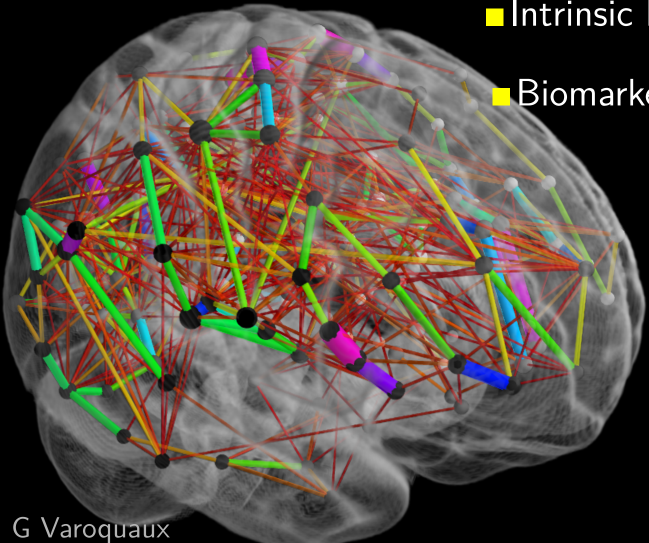
The functional connectome

View of the brain as a set of regions and their interactions



The functional connectome

View of the brain as a set of regions and their interactions

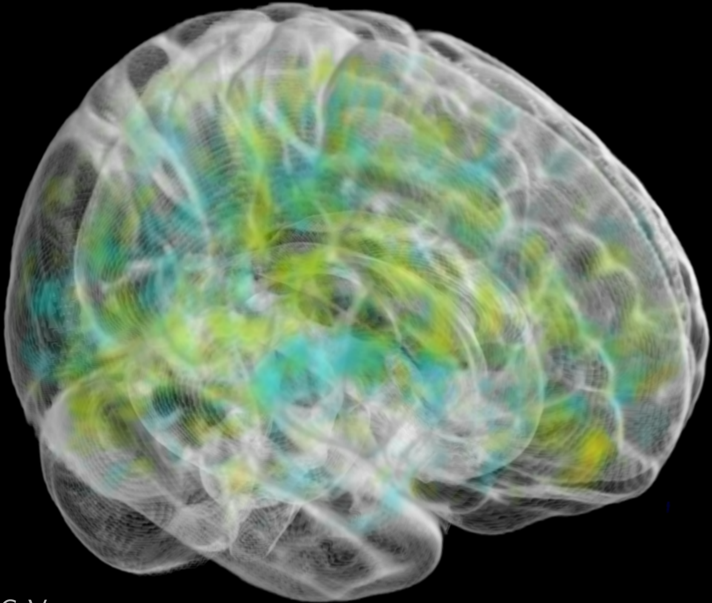


- Intrinsic brain architecture
- Biomarkers of pathologies

Learn a graphical model

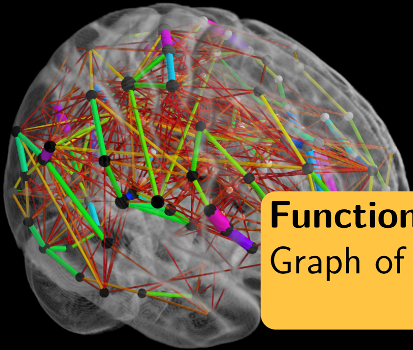
Human Connectome Project: 30M$

Resting-state fMRI



Outline

1 Graphical structures of brain activity

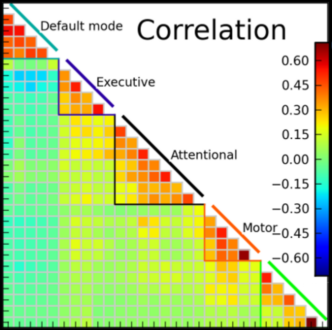


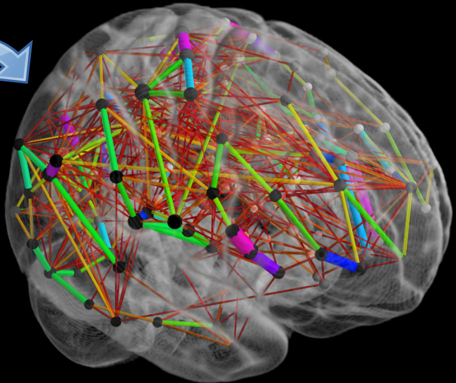
Functional connectome

Graph of interactions between regions

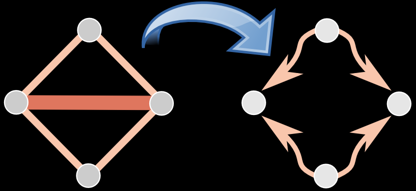
[Varoquaux & Craddock 2013]

1 From correlations to connectomes





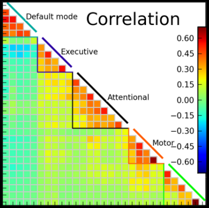
Conditional independence structure?



1 Probabilistic model for interactions

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

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



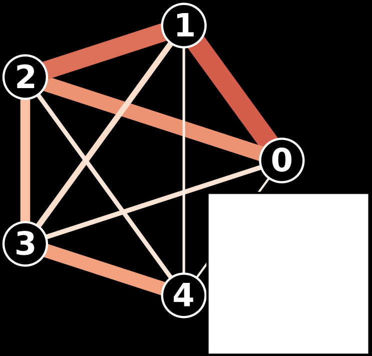
- Model parametrized by inverse covariance matrix, $\mathbf{K} = \mathbf{\Sigma}^{-1}$: ***conditional*** covariances

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

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

Observations

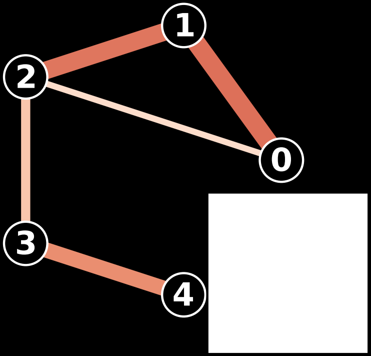
Covariance



Diagonal:
signal variance

Direct connections

Inverse covariance

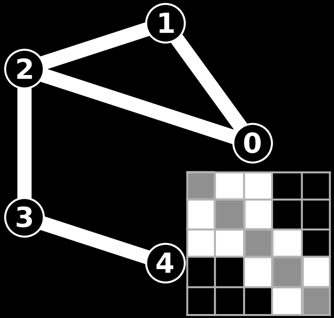


Diagonal:
node innovation

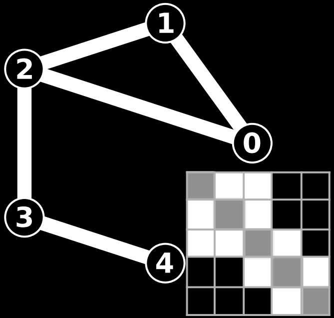
1 Independence structure (Markov graph)

Zeros in partial correlations
give **conditional independence**

Reflects the large-scale brain interaction structure



Zeros in partial correlations
give **conditional independence**



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

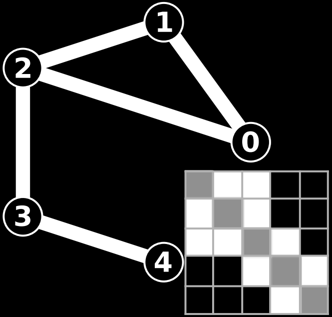
Independence between nodes makes estimation of partial correlations well-conditionned.

Chicken and egg problem

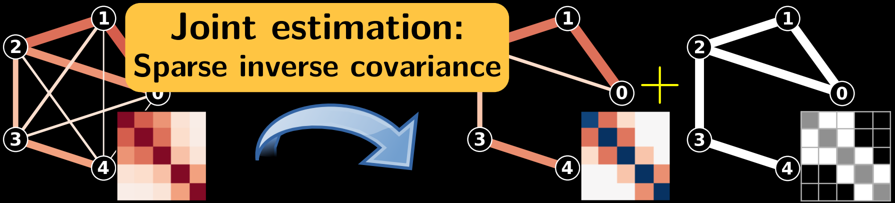
1 Independence structure (Markov graph)

Zeros in partial correlations
give **conditional independence**

Ill-posed problem:
multi-collinearity
⇒ noisy partial correlations



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



1 Sparse inverse covariance estimation: penalized

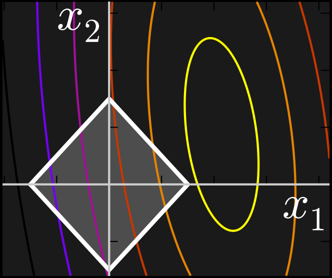
Maximum a posteriori:

- 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,



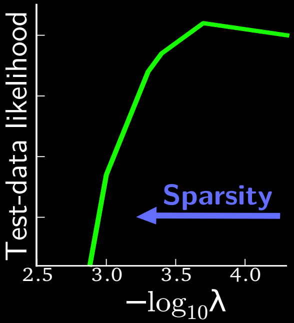
[Varoquaux NIPS 2010] [Smith 2011]

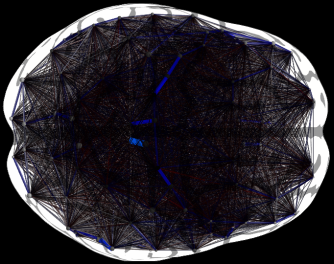
Maximum a posteriori:

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

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

- Optimal graph almost dense





Maxi

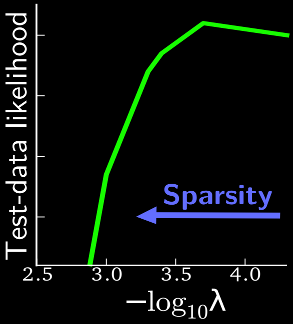
- Fit r

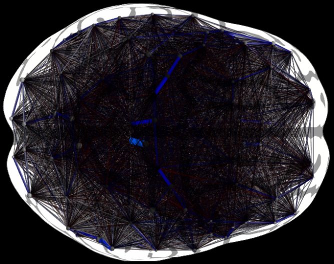
- **Bias of ℓ_1:**
  very sparse graphs don't fit the data

- Sparsity $\Rightarrow$ Lasso-like problem: ℓ_1 penalization

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

- Optimal graph almost dense





Maxi

- **Bias of ℓ_1:**
  very sparse graphs don't fit the data

- Fit r

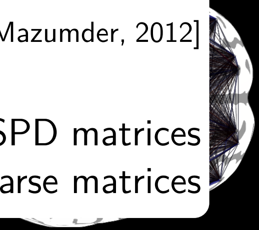
- Sparsity $\Rightarrow$ Lasso-like problem: ℓ_1 penalization

K

Algorithmic considerations:

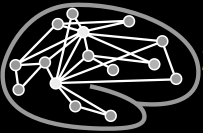
- **Very ill-conditionned input matrices**
- **Graph-lasso** [Friedman 2008] **doesn't work well**
  primal-dual algorithm with approximation
  when switching from dual to primal [Mazumder, 2012]
- **Good success with ADMM**
  split optimization: loss solved with SPD matrices
  penalty solved with sparse matrices

a



$-\log_{10}\lambda$

1 Very sparse graphs: greedy construction



Sparse inverse covariance algorithm: PC-DAG [Rutimann & Buhlmann 2009]

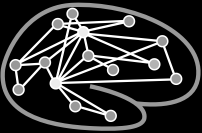
Greedy approach

1. **PC-alg**: fill graph by independence tests conditioning on neighbors
2. Learn covariance on resulting structure

Good for very sparse graphs

[Varoquaux J. Physio Paris, 2012]

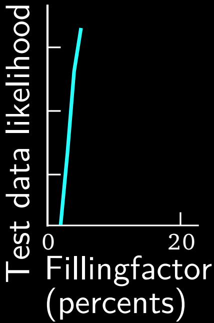
1 Sparse graphs: greedy construction

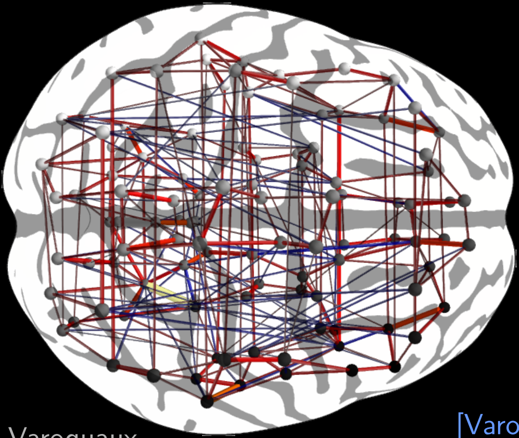


Iterate construction alg.

High-degree nodes appear very quickly

complexity $\propto$ exp degree



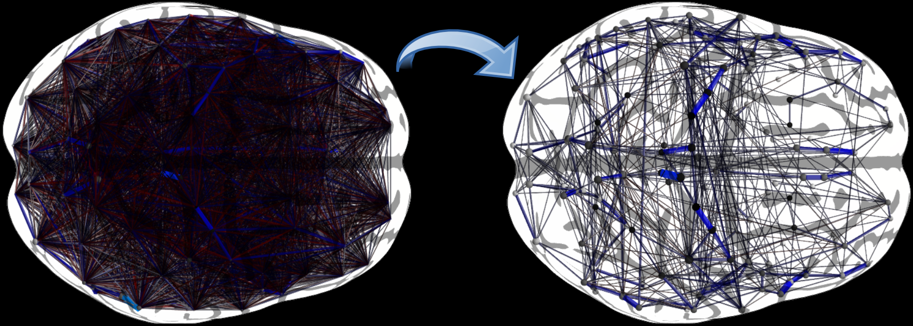


Lattice-like structure with hubs

[Varoquaux J. Physio Paris, 2012]

2 Multi-subject graph learning

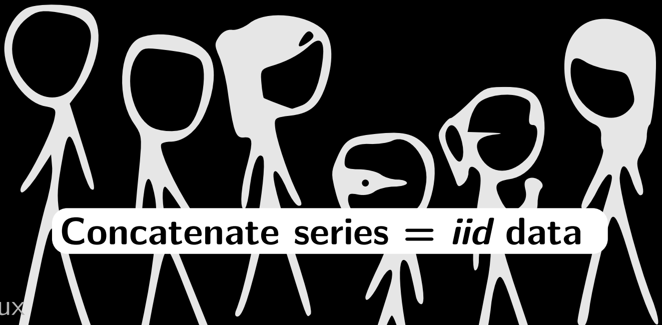
Not enough data per subject to recover structure

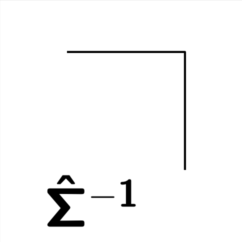


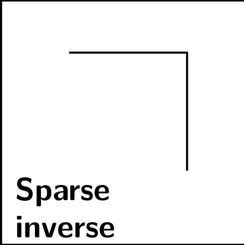
Sparse recovery for Gaussian graphs

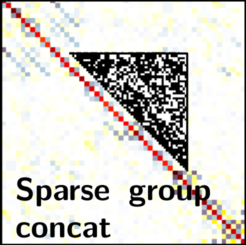
- ℓ_1 structure recovery has phase-transitions behaviors
- For Gaussian graphs with s edges, p nodes:
  $n = \mathcal{O}\big((s+p)\log p\big), \quad s = o\big(\sqrt{p}\big)$ [Lam & Fan 2009]

Need to accumulate data across subjects









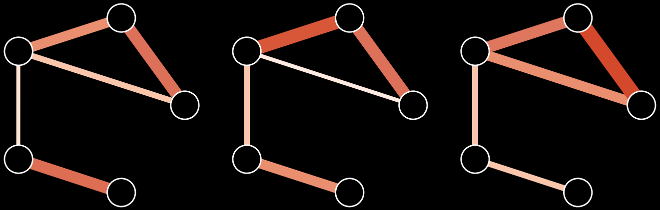
Likelihood of new data (cross-validation)

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-subect variability

[Varoquaux NIPS 2010]

Common independence structure but different connection values



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

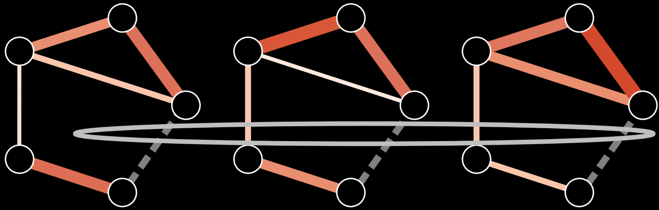
Multi-subject data fit, Likelihood

Group-lasso penalization

[Varoquaux NIPS 2010]

2 Multi-subject modeling

Common independence structure but different connection values



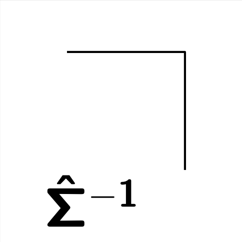
$$\{\mathbf{K}^s\} = \underset{\{\mathbf{K}^s \succ 0\}}{\operatorname{argmin}} \sum_s \mathcal{L}(\hat{\mathbf{\Sigma}}^s | \mathbf{K}^s) + \lambda\, \ell_{21}(\{\mathbf{K}^s\})$$

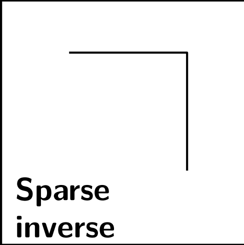
Multi-subject data fit, Likelihood

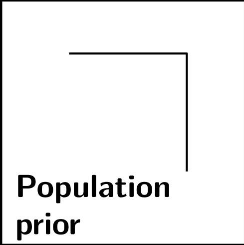
ℓ_1 on the connections of the ℓ_2 on the subjects

[Varoquaux NIPS 2010]

Population-sparse graph perform better



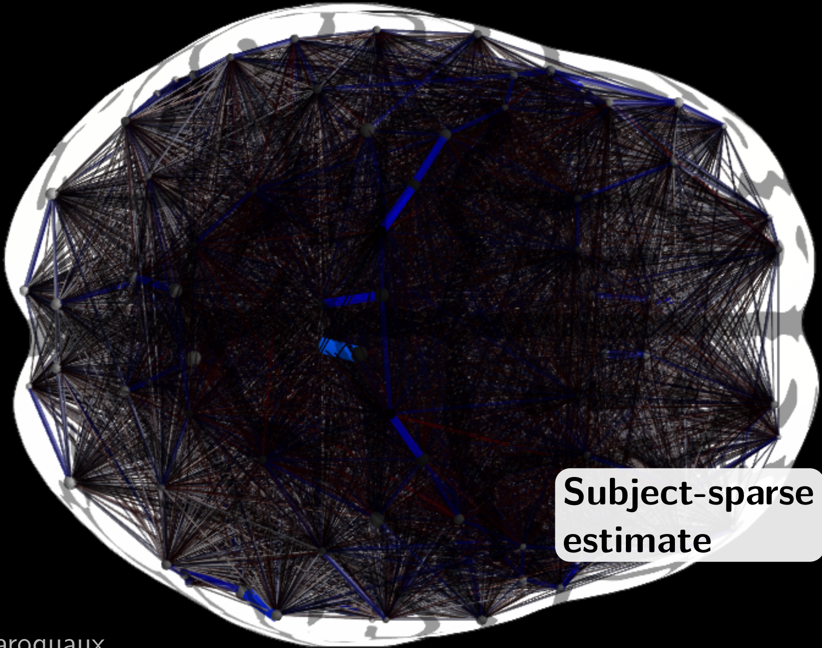




Likelihood of new data (cross-validation)		sparsity
Subject data, Σ^{-1}	-57.1	
Subject data, sparse inverse	43.0	60% full
Group concat data, Σ^{-1}	40.6	
Group concat data, sparse inverse	41.8	80% full
Group sparse model	**45.6**	20% full

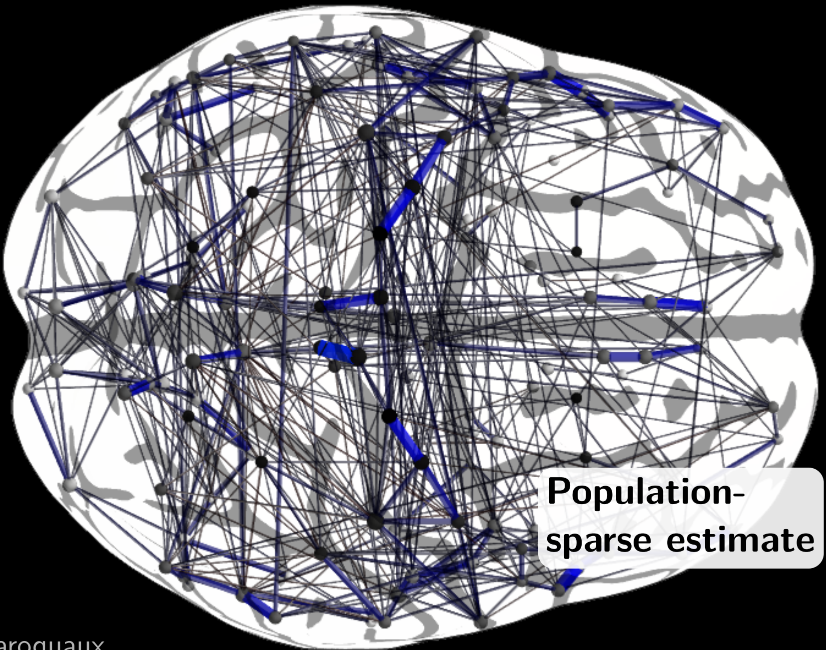
[Varoquaux NIPS 2010]

2 Independence structure of brain activity



Subject-sparse estimate

2 Independence structure of brain activity

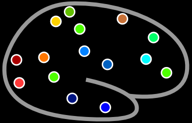


2 Large scale organization

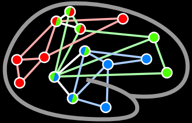
High-level cognitive function arises from the interplay of specialized brain regions:

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

Functional segregation: nodes of connectome
atomic functions – *tonotopy*



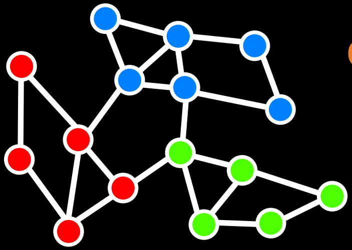
Global integration: functional networks
high-level functions – *language*



High-level 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]

Scale-free hierarchical integration / segregation

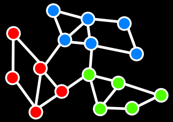


Graph modularity =
divide in *communities* to maximize intra-class connections versus extra-class

[Eguiluz 2005]

2 Graph cuts to isolate functional communities

- Find communities to maximize modularity:



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

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

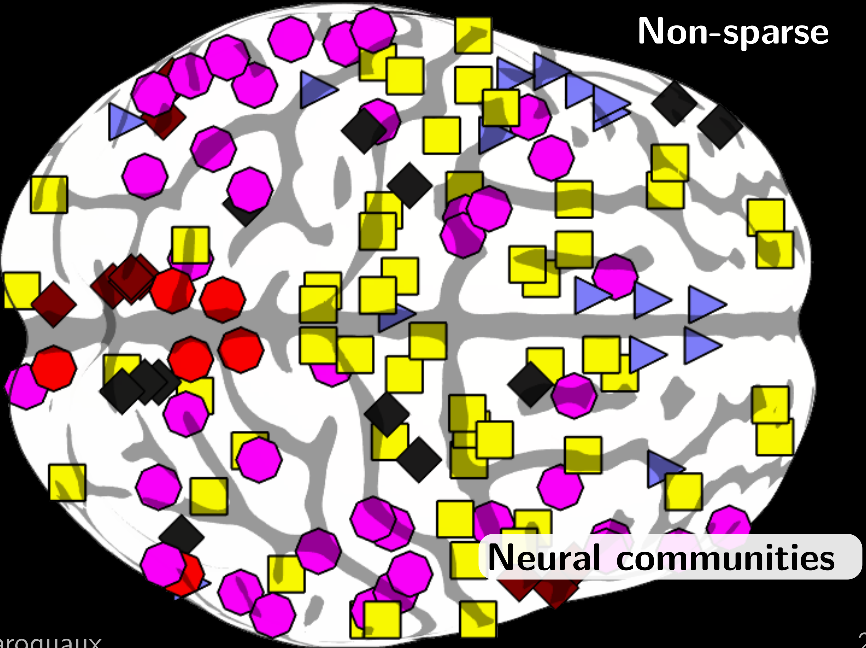
- Rewrite as an eigenvalue problem [White 2005]

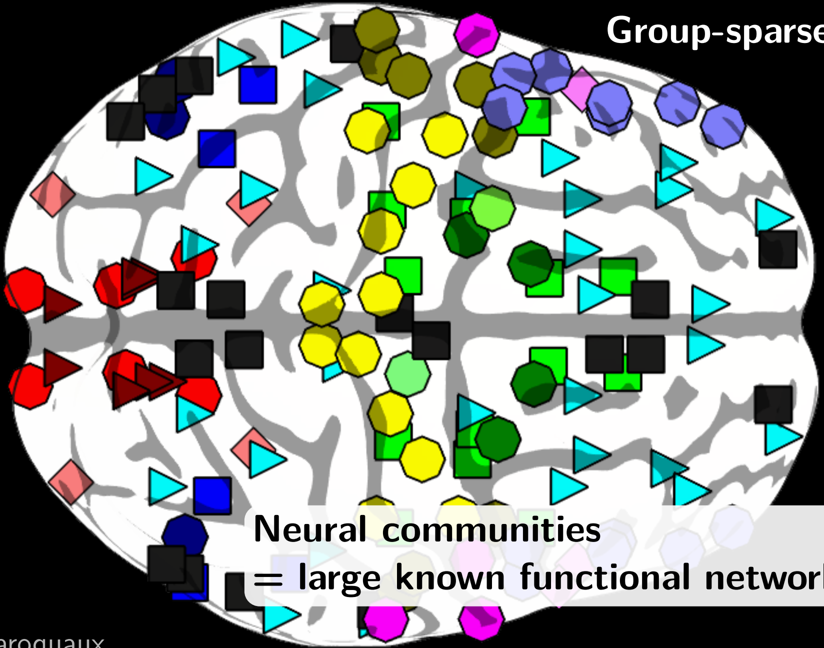
$$\begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix} \begin{bmatrix} \quad A \quad \end{bmatrix} \bullet \begin{bmatrix} 1 & 1 & 0 & 0 \end{bmatrix}$$

$\Rightarrow$ Spectral clustering = spectral embedding + k-means

- Similar to normalized graph cuts

2 Large scale organization

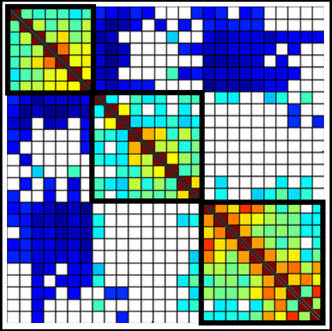




Neural communities
= large known functional networks

Proposed measure for functional integration: mutual information (Tononi)

[Marrelec 2008, Varoquaux & Craddock 2013]



- Integration:

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

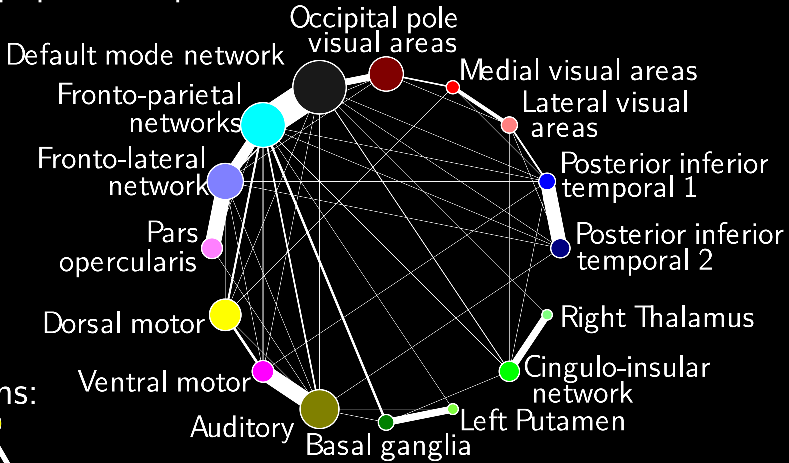
"energy" in network

- Mutual information:

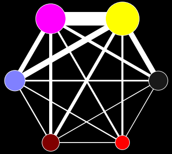
$$M_{c_1, c_2} = I_{c_1 \cup c_2} - I_{c_1} - I_{s_2}$$

"cross-talks" between networks

With population prior:

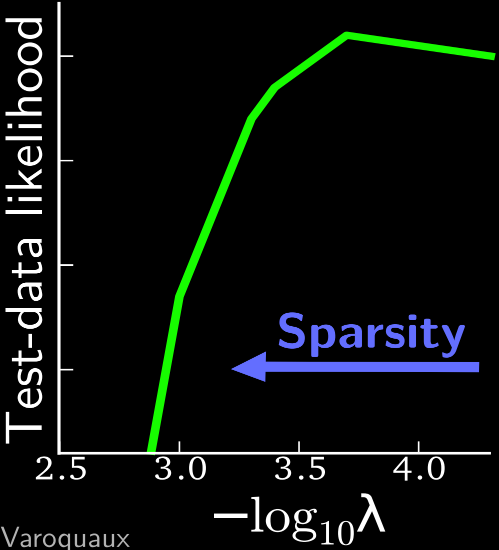


Raw correlations:



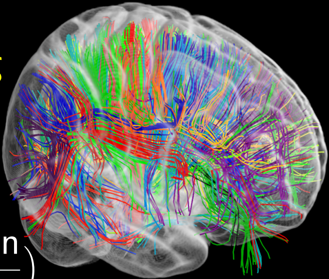
[Varoquaux NIPS 2010]

3 Beyond ℓ_1 models



3 Weighted-ℓ_1: incorporating additional prior

- Not all connections are as likely



Tractography: the physical wiring

Noisy estimate of likelihood of functional connection

$\Rightarrow$ Provides a soft prior:

$$\mathcal{P}(\text{func conn}) \propto \exp\left(-\frac{\text{anat conn}}{\sigma}\right)$$

Graph MAP estimate:

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

$$\lambda_{i,j} = \lambda_0 \exp\left(-\frac{\text{anat conn}}{\sigma}\right)$$

[Ng MICCAI 2012]

3 Weighted-ℓ_1: incorporating additional prior

- Not all connections are as likely

Tractography: the physical wiring

Noisy estimate of likelihood of functional connection

$\Rightarrow$ Provides a soft prior:

$\mathcal{P}(\text{func conn}) \propto \exp\left(-\frac{\text{anat conn}}{\sigma}\right)$

Limitation:

Tractography estimates unreliable

Graph MAP estimate:

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

$$\lambda_{i,j} = \lambda_0 \exp\left(-\frac{\text{anat conn}}{\sigma}\right)$$

[Ng MICCAI 2012]

3 Reweighted-ℓ_1: learning inhomogenous penalty

Ideas

- As in regression reweighted ℓ_1 [Candes 2008]:
  First ℓ_1 estimates gives rescaling for penalties
  $\Rightarrow$ Support recovery in heteroschedastic settings
  Equivalent to non-convex ℓ_0 approximation
  But we have no edge-level residual

- As in stability selection [Meinshausen 2010]:
  Edges stable to perturbations most likely

[Phlypo MICCAI 2014]

Perturbations

- We have many subjects:
  run an ℓ_1 model per subject

  $\Rightarrow$ Posterior probability of edge presence: $\mathcal{P}_{ij}$

  fit a binomial

Reweighting

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

$$\lambda_{i,j} = \lambda_0 \, \mathcal{P}_{ij}$$

[Phlypo MICCAI 2014]

Statistical learning for functional connectomes

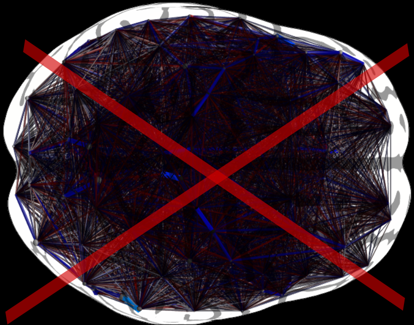
fMRI: scarsity of data + low SNR

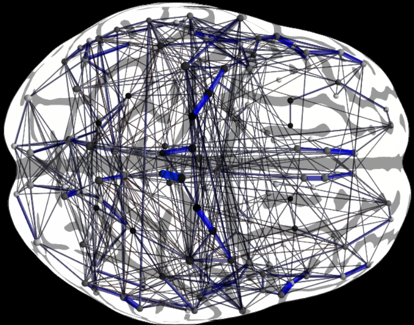
Graphical Gaussian models: sparse inverse covariance

- ℓ_1/ℓ_{21} penalty
- Iterative non convexity

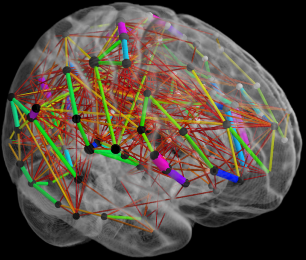
Software: Python, open source

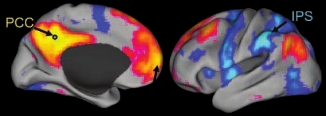
- http://scikit-learn.org
- http://nilearn.github.io



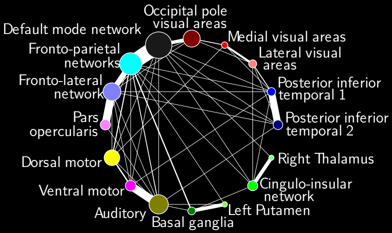


Statistical learning for **functional connectomes**

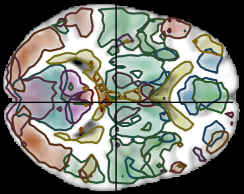




- Complex graph with a modular structure



- The communities are cognitive networks that link to behavior



- Requires the definition of regions

[Abraham 2013]

@GaelVaroquaux