# Graph inference via multiple testing

## Applications in neuroscience

Sophie Achard

CNRS, GIPSA-lab, Grenoble Image Parole Signal Automatique (UMR5216)

joint work with P. Borgnat, I. Gannaz and M. Roux

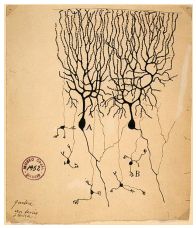
sophie.achard@gipsa-lab.fr

2 April 2019

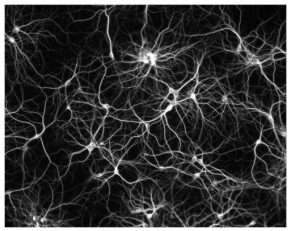
gipsa-lab


# The brain, an anatomical and functional network

## Microscale

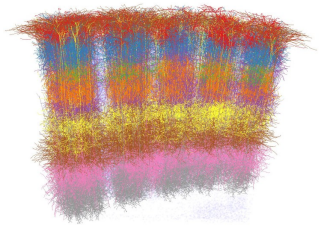


[Cajal, 1890]



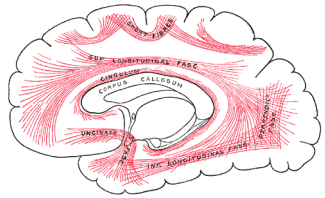
[Yuste, 2015]

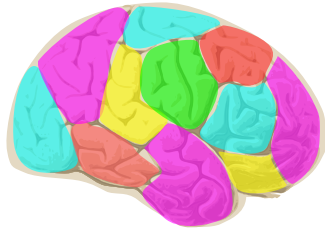
## Mesoscale



[Oberlaender *et al.*, 2013]

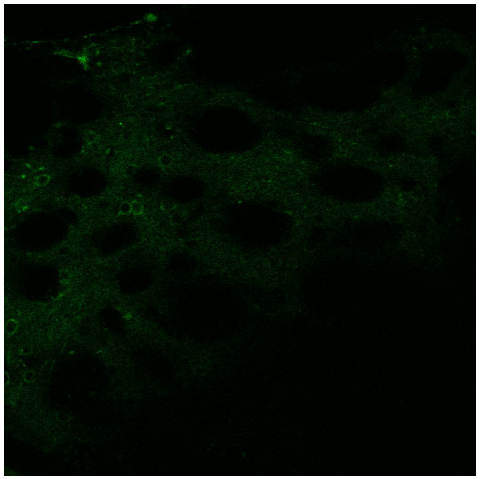
## Macroscale





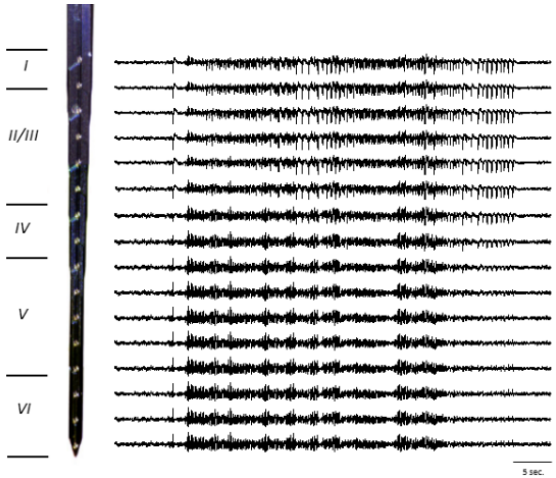
## Neuroimaging

## Optical imaging



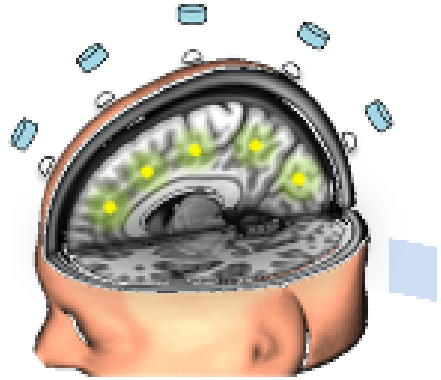
[Badreddine PhD]

## Microelectrodes




Thanks to Florian Studer

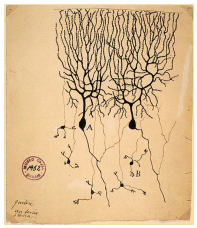
## fMRI/EEG/MEG



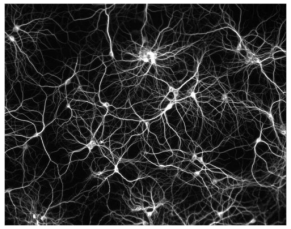
[De Vico Fallani *et. al*, 2016]

# The brain, an anatomical and functional network

Microscale

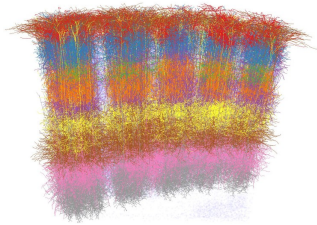


[Cajal, 1890]



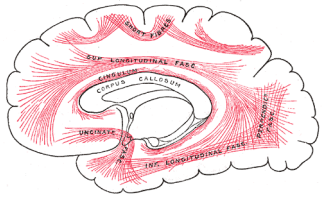
[Yuste, 2015]

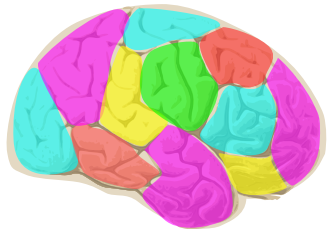
Mesoscale



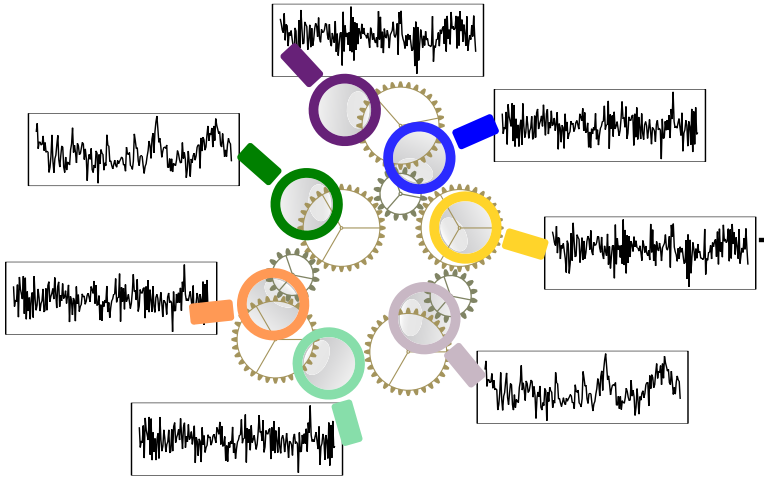
[Oberlaender *et al.*, 2013]

Macroscale

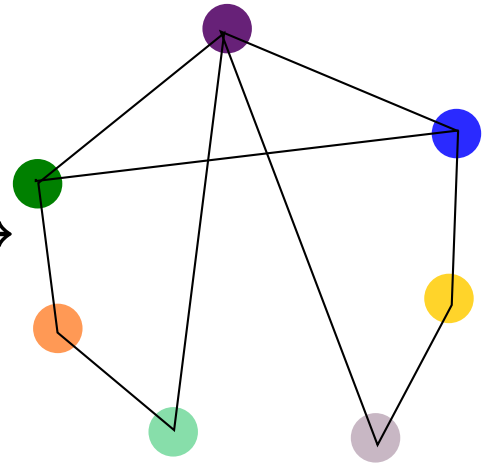


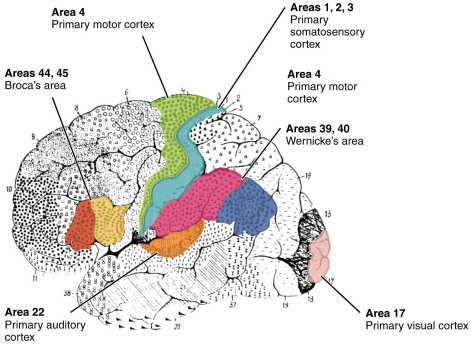
Mathematical modeling

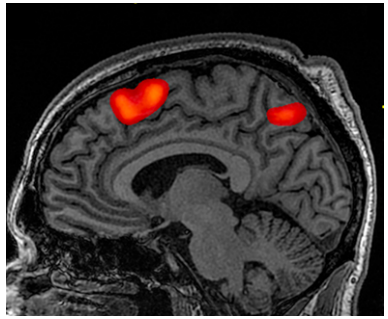


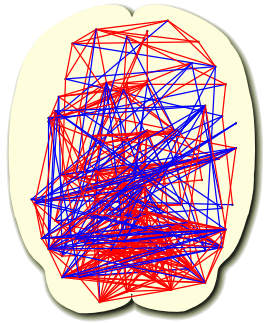
Network

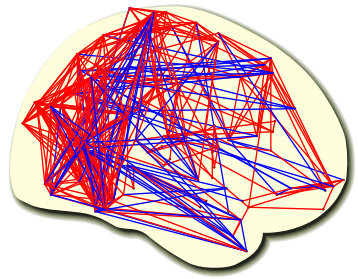


# The brain, an anatomical and functional network





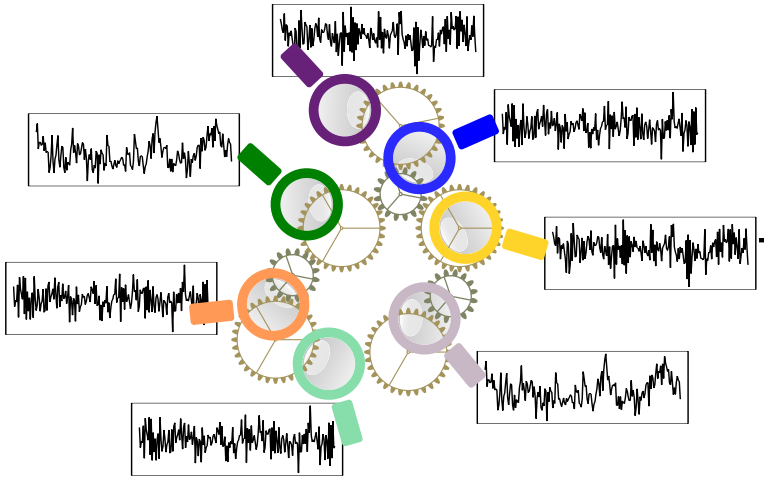


Phrenology view

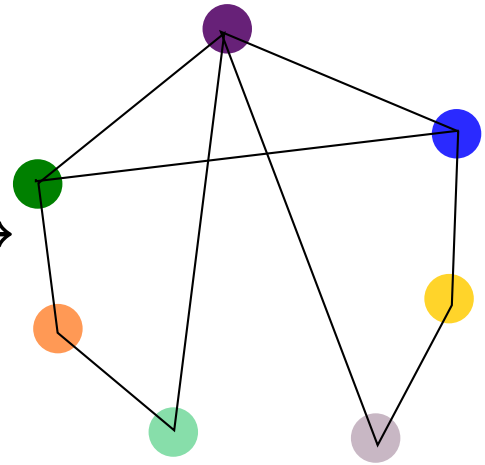
Neuroimagery

Connectivist view

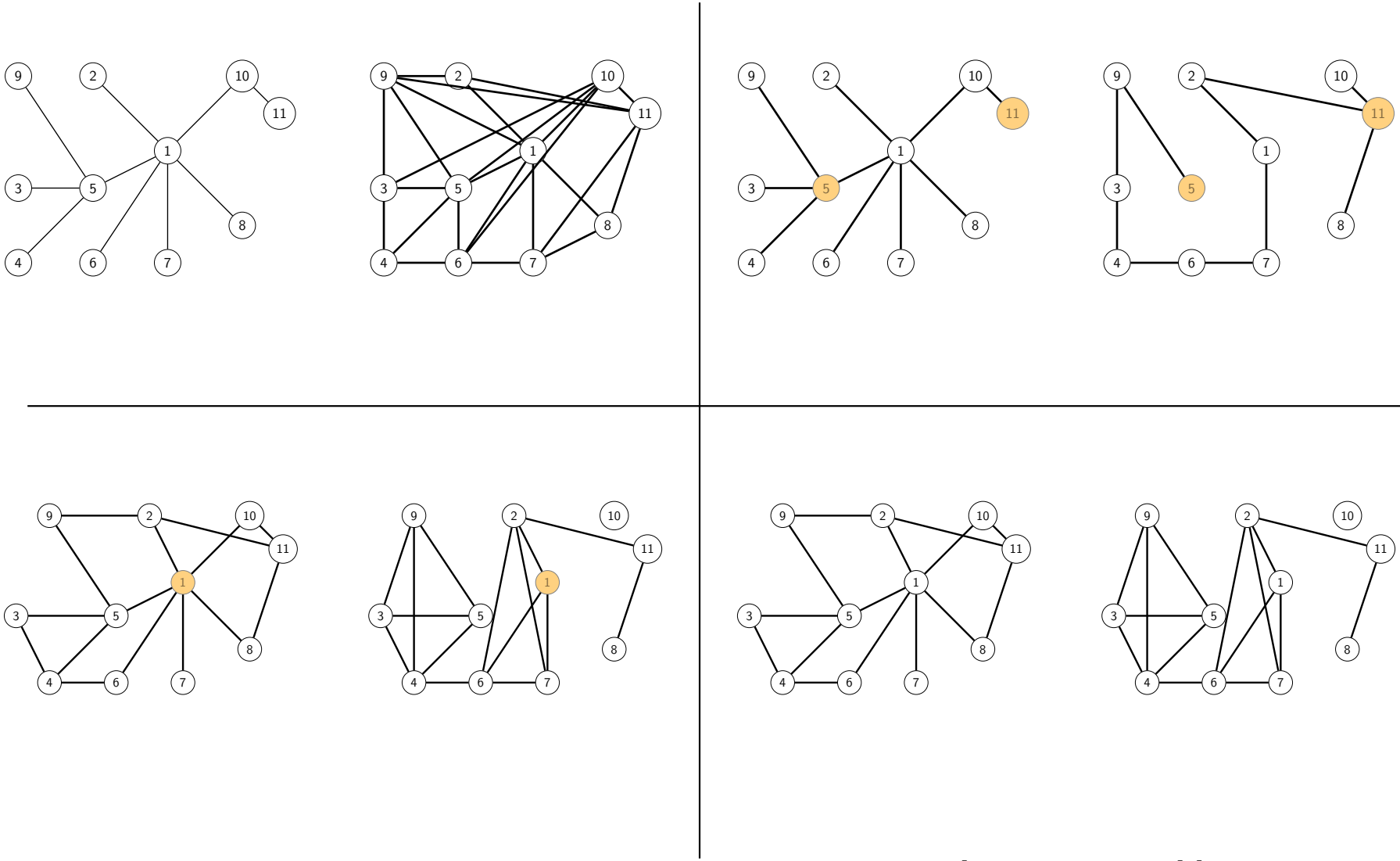
Mathematical modeling



Network →



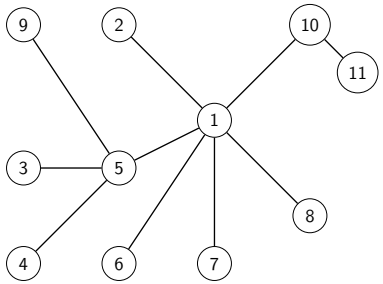
# Interpretation of graph metrics



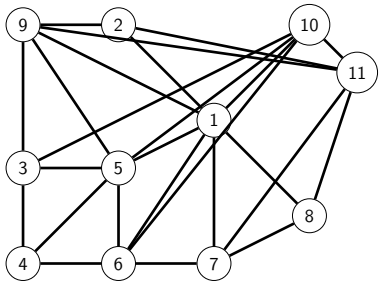
[**Latora** *et al.* 2001] [**Bullmore** *et al.* 2009]
[**Csárdi** *et al.* 2006]

# Interpretation of graph metrics

## Cost

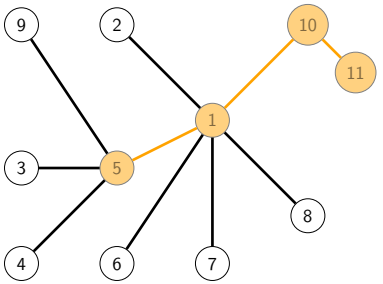



Low Cost

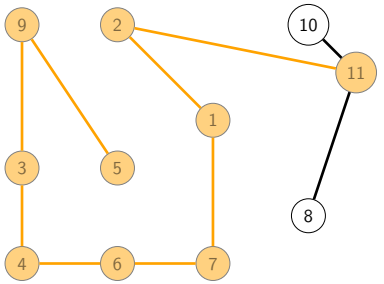



High Cost

## Efficiency

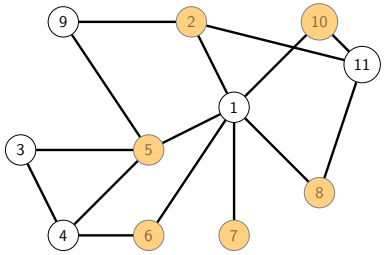



High Eglob

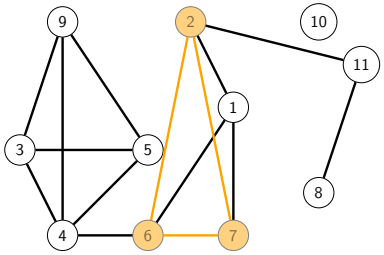



Low Eglob

## Clustering

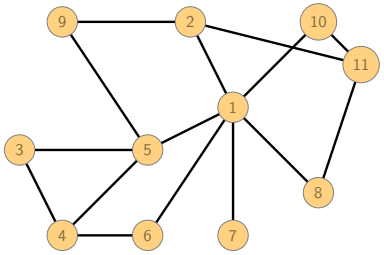



Low Clust

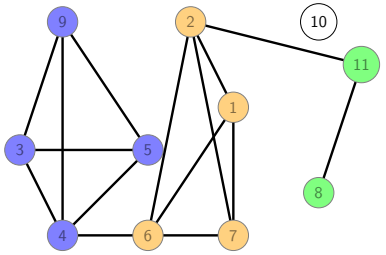



High Clust

## Modularity



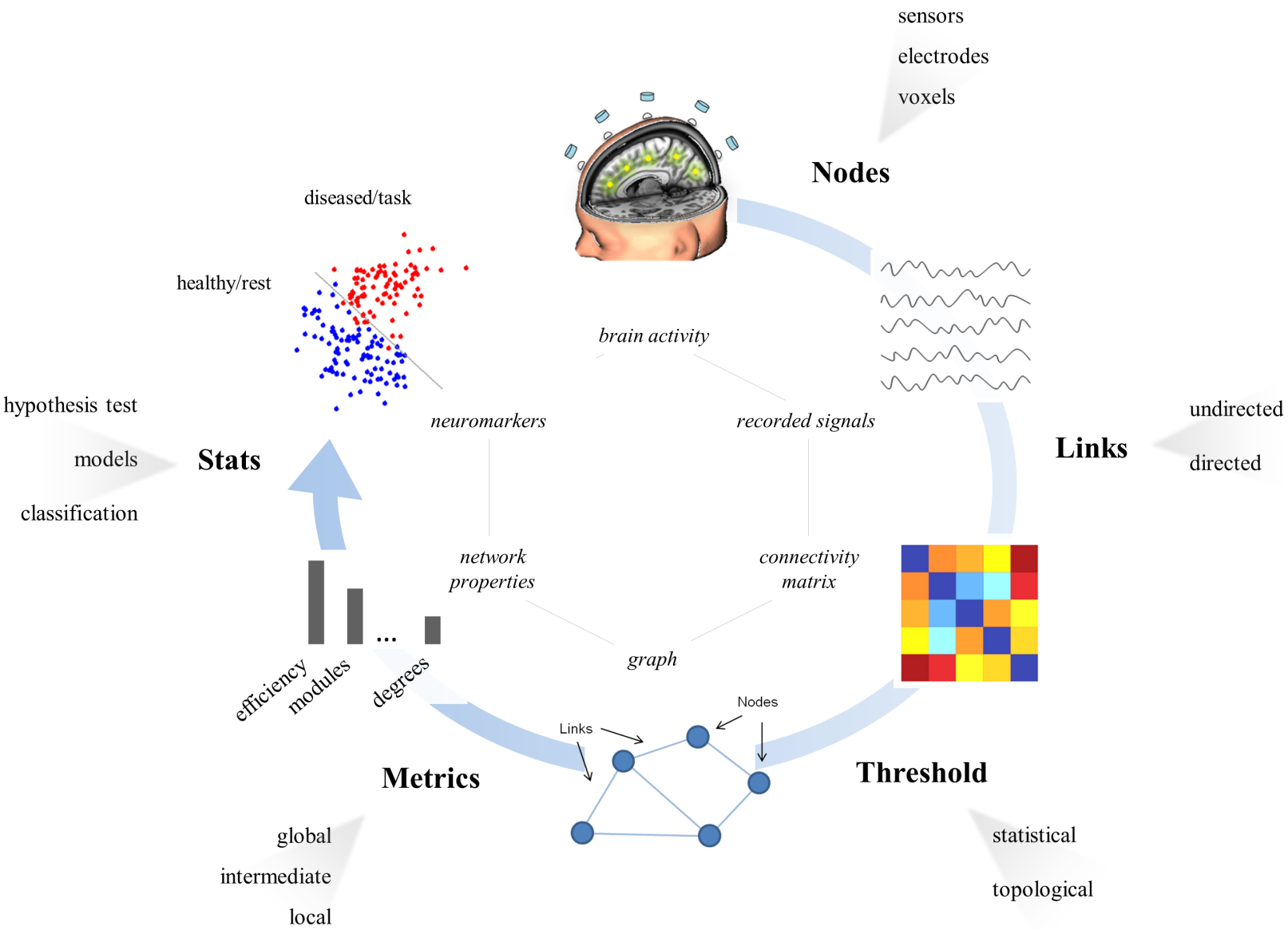

Low Modularity




High Modularity

[**Latora** *et al.* 2001] [**Bullmore** *et al.* 2009]
[**Csárdi** *et al.* 2006]

# Exploring the brain using networks analysis: summary




[De Vico Fallani *et al.*2014]

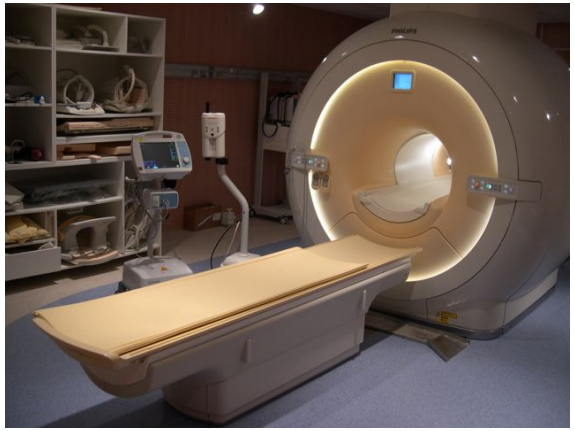
# From dead salmon to dead rat

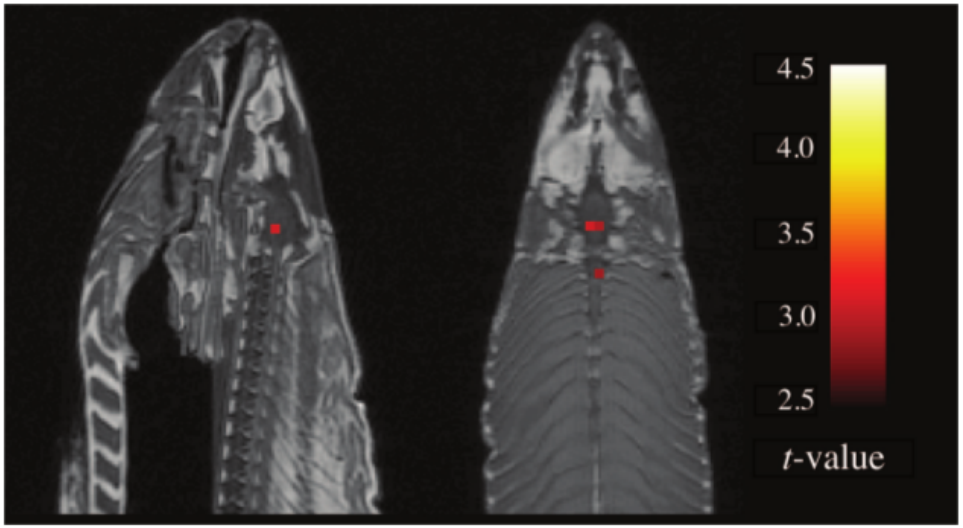
Journal of Serendipitous and Unexpected Results

**Neural Correlates of Interspecies Perspective Taking in the Post-Mortem Atlantic Salmon: An Argument For Proper Multiple Comparisons Correction**

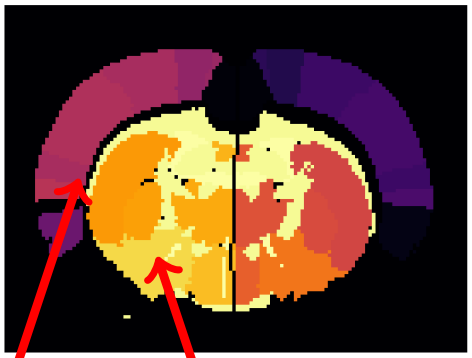
Craig M. Bennett [1*], Abigail A. Baird [2], Michael B. Miller [1] and George L. Wolford [3]

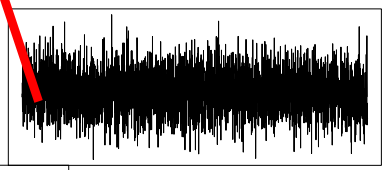
(2011)

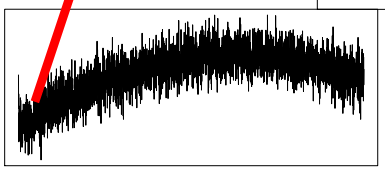




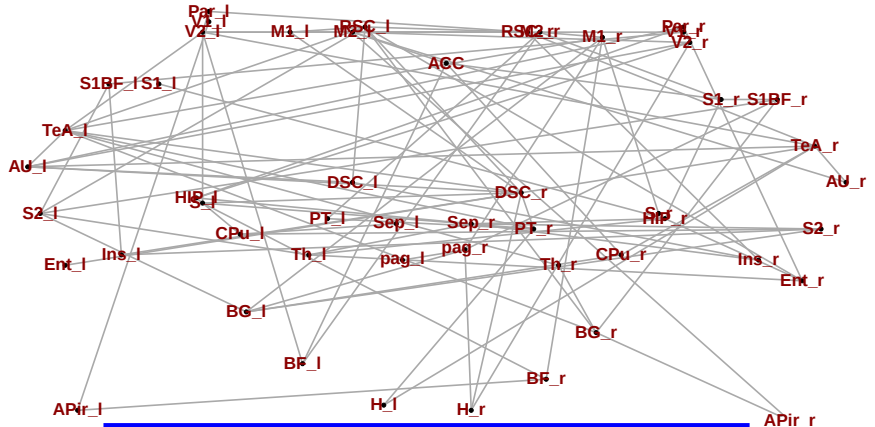
# Dead rat results



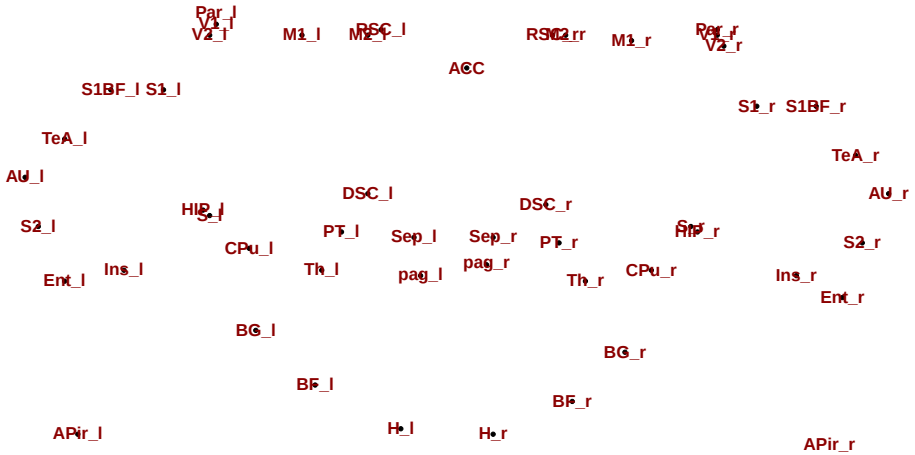



Without multiple correction




With multiple correction



# Reproducibility of global efficiency

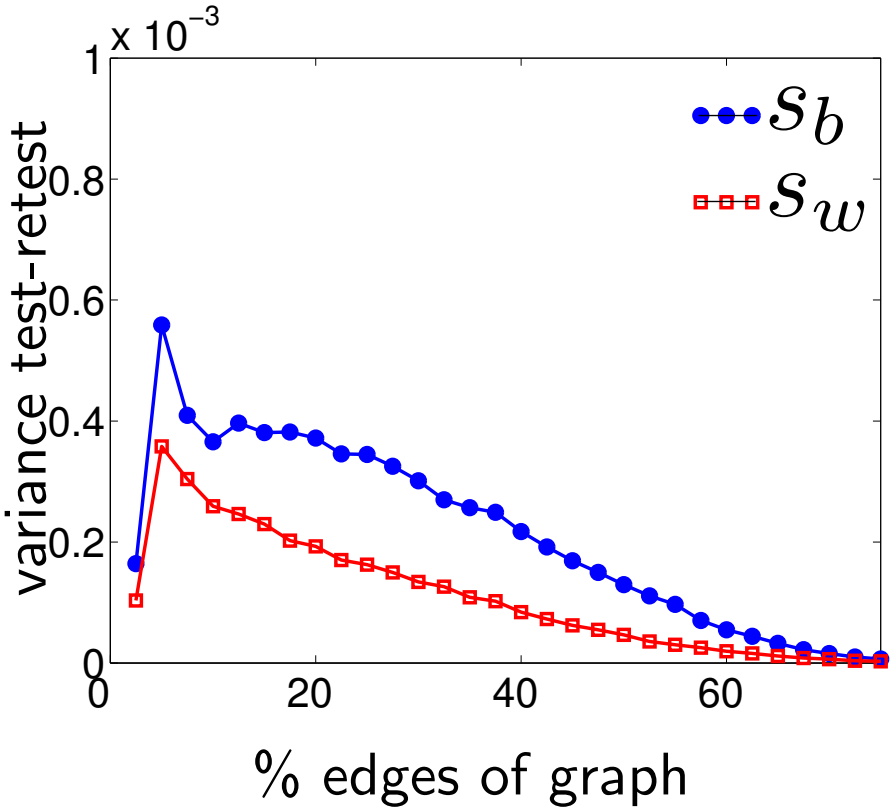
HUMAN Connectome PROJECT

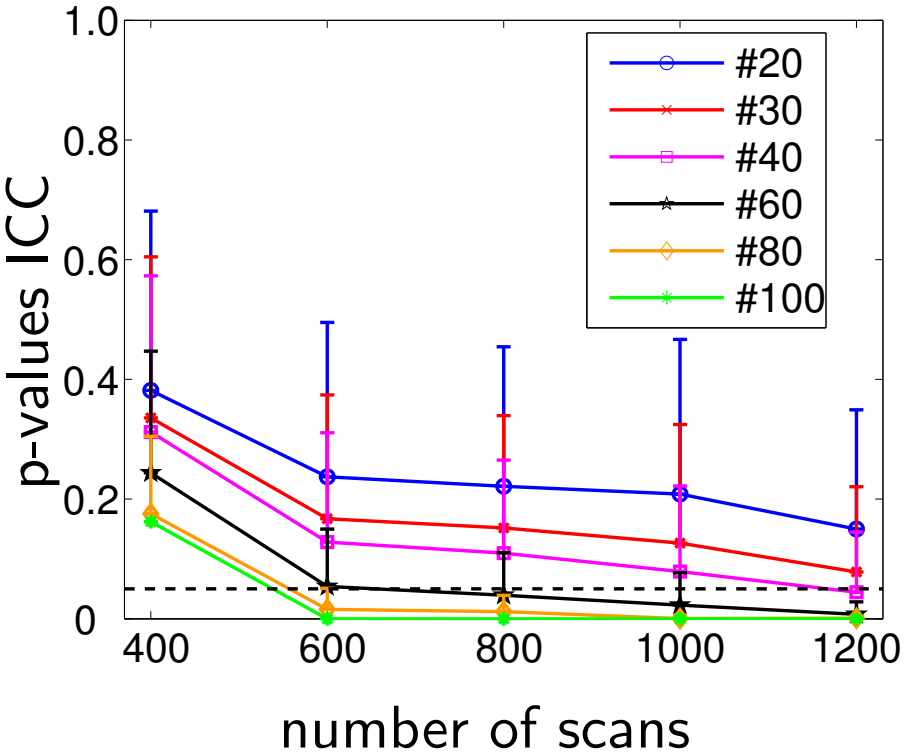
100 subjects rs-fMRI
2 sessions
14 min. TR=0.7s

Intra class correlation

$$ICC = \frac{s_b - s_w}{s_b + s_w}$$

$s_b$ variance between subj.
$s_w$ variance within session




[**Termenon** *et al.* 2016]

# Reproducibility of global efficiency

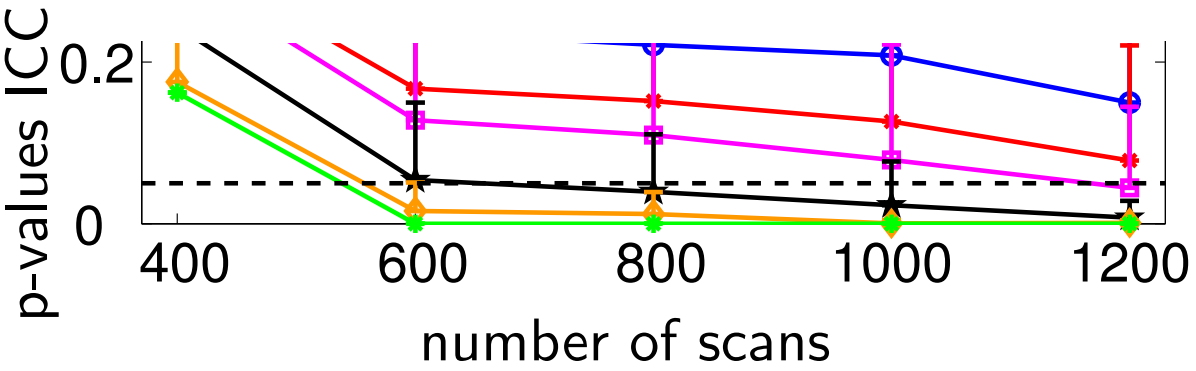
HUMAN Connectome PROJECT

100 subjects rs-fMRI
2 sessions
14 min. TR=0.7s

Intra class correlation

$$ICC = \frac{s_b - s_w}{s_b + s_w}$$

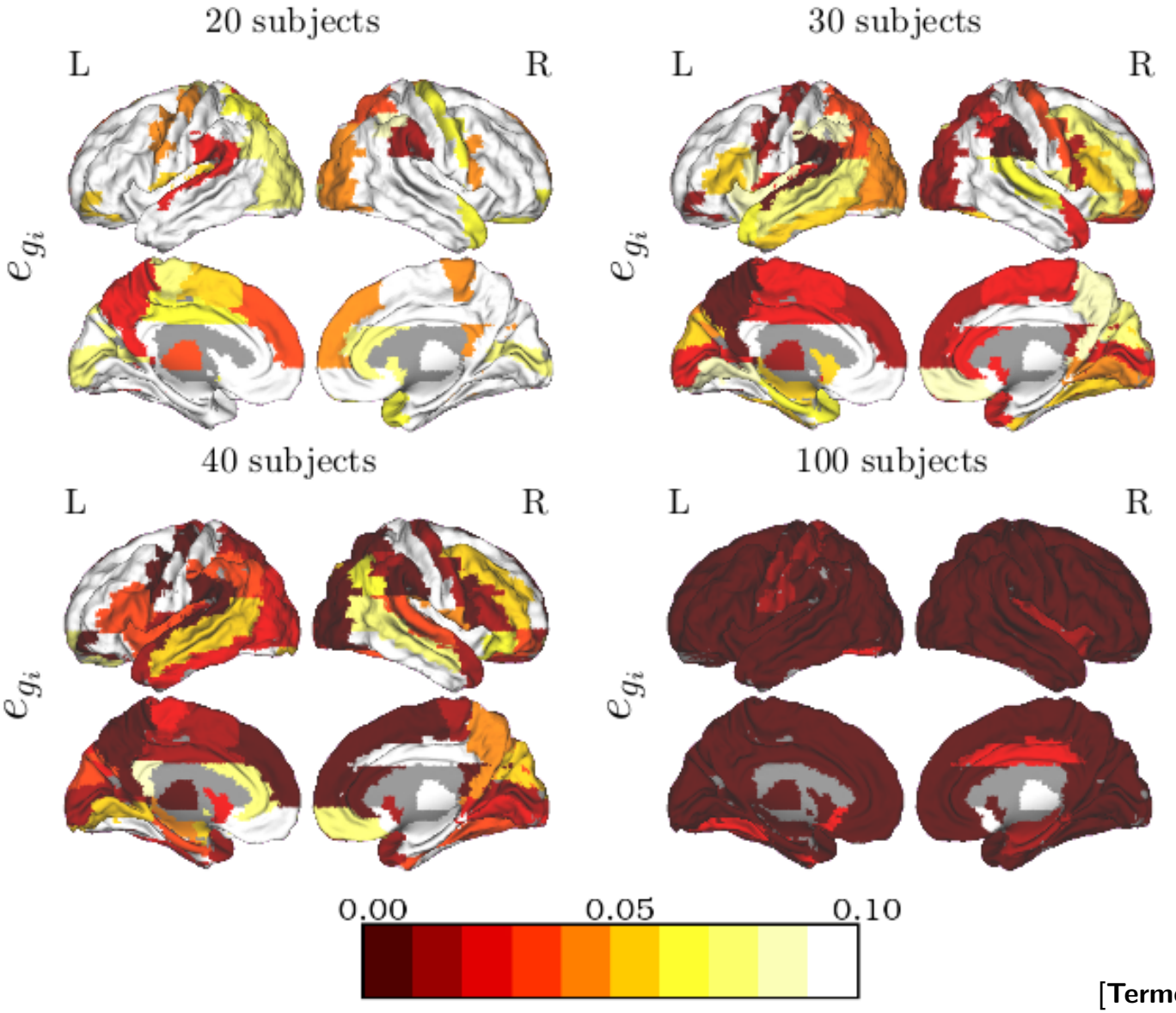
$s_b$ variance between subj.
$s_w$ variance within session




Data are available to download from my website!

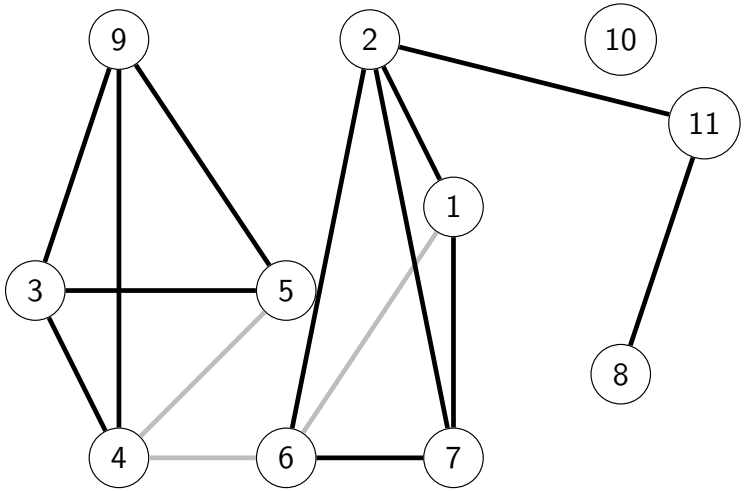
[**Termenon** *et al.* 2016]

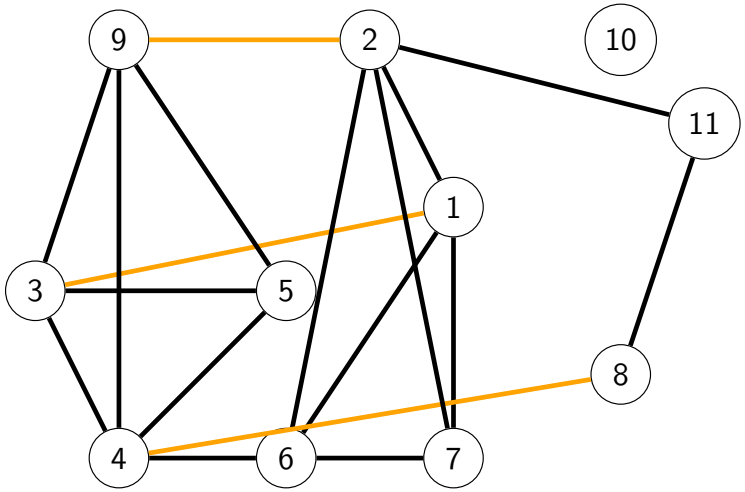
# Spatial pattern of reproducible local metrics

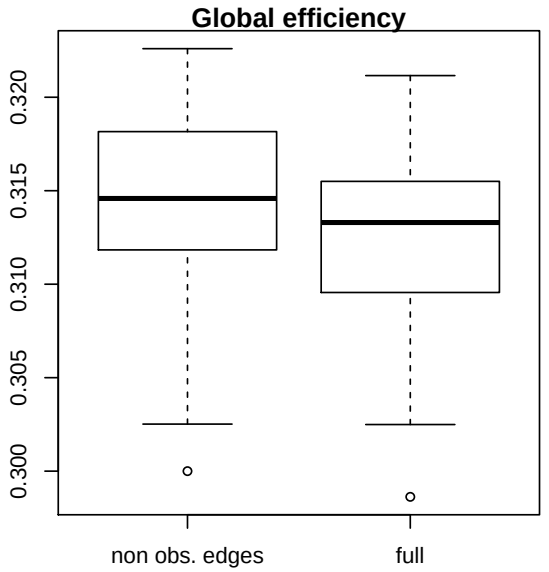


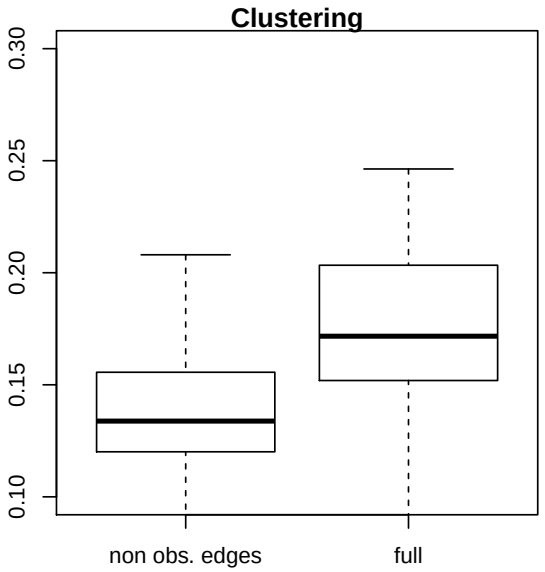

[**Termenon** *et al.* 2016]

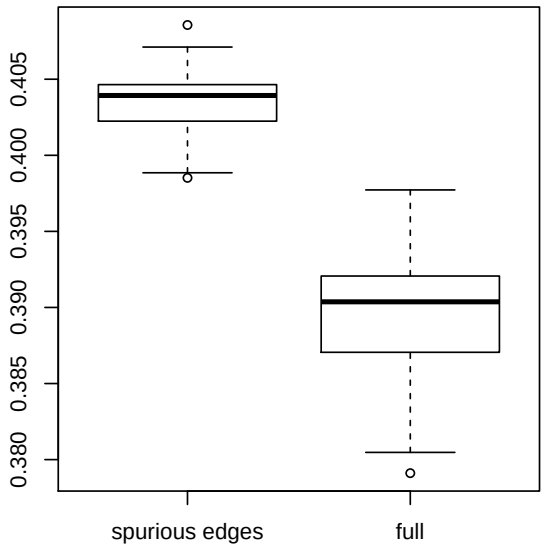
# Robustness to wrong identification of edges

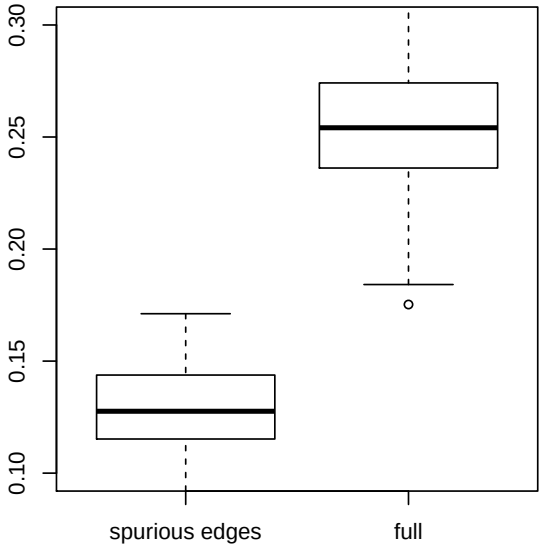


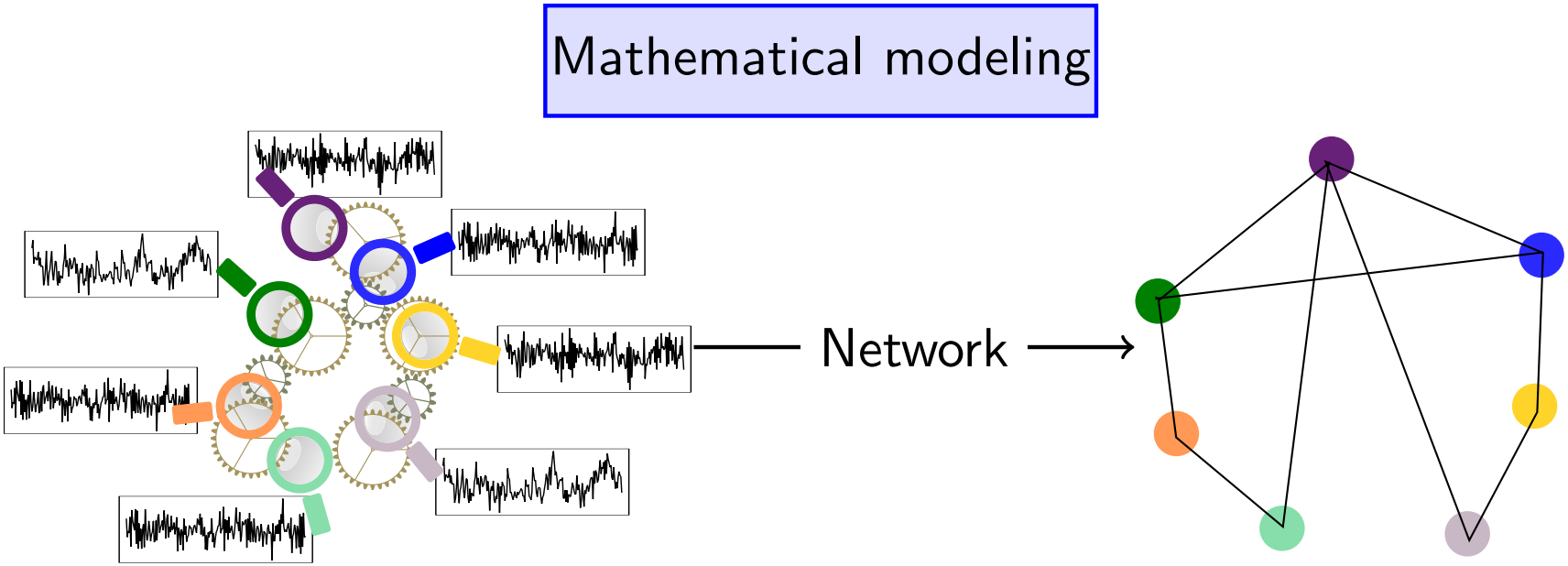



# Objectives




- Find the largest size of correct edges
- Control the number of non correct edges
- Optimize the procedures based on the size of data

# Main questions

- What are the keys parameters to be robust?
- How to quantify the robustness of graph metrics?
- How to reveal alterations of networks?

Not large graphs, around hundreds of nodes.

However complex multivariate time series.

# Existing methods to estimate covariance matrices

- Hard thresholding [**Bickel** *et al.* 2008]
  Threshold of covariance matrices; cross validation.

- Shrinkage [**Ledoit** *et al.* 2004]
  Optimisation towards identity; forcing zeros in the covariance matrices.

- Regularisation [**Bien** *et al.* 2011]
  Optimisation using lasso penalisation; non convex problem.

- Graph Signal Processing [**Pasdeloup** *et al.* 2017]
  Optimisation through smoothness of signal on graphs; eigenvalues decomposition.

# Graph inference with multiple testing

Objectives of our method:

- Simple application to any graph measures
- Control the false positive
- Control the false negative
- Unweighted graphs

# Simple test

- $(X_i, Y_i)_{1 \leq i \leq n}$ i.i.d $\mathcal{N}(0, \Sigma)$.
- "$H_0 : r = 0$" against "$H_1 : r \neq 0$".
- Reject of $H_0$ if $T(X) = |\hat{r}| / \left(\sqrt{(n-2)(1-\hat{r}^2)}\right) > t_\alpha$.
- Type I error: $\mathbb{P}_{H_0}\left(T(X) > t_\alpha\right)$ (false positive).
- $t_\alpha$ type I error of level $\leq \alpha$.
  Here, $t_\alpha$ = quantile of order $(1 - \alpha/2)$ of $\mathcal{S}t(n-2)$ (law of $T(X)$ under $H_0$).

# $p$-value

A $p$-value, $p(X)$, is a transformation of $T(X)$ such that :

- Under $H_0$, $p(X) \sim \mathcal{U}(0,1)$,
- Under $H_1$, $p(X)$ "close to" 0.

Then,

$$\text{reject } H_0 \text{ if } p(X) \leq \alpha \Rightarrow \text{type I error} \leq \alpha.$$

# Multiple testing

Approach in order to reject part of null hypothesis

- Simultaneous tests of $m$ mull hypotheses such that $m_0 = |\mathcal{H}_0|$ are true.
- Computation of the $m$ $p$-values.
- Define the set $\mathcal{R} \subseteq \{1, \ldots, m\}$ of the rejected null hypotheses.

# Definition: graph inference via multiple testing

- $\mathbb{X} = \left(X^{(1)}, \ldots, X^{(n)}\right)$ i.i.d sample such that for all $j = 1, \ldots, n$, $X^{(j)} \sim \mathcal{N}_d(0, \Sigma)$ and $\Sigma = (\rho_{ii'})_{1 \leq i, i' \leq d}$, $\rho_{ii} = 1$ for all $i = 1, \ldots, d$.

- $\mathcal{H} = \left\{(i, i'), i < i'\right\}$. For all $h \in \mathcal{H}$, do $m = d(d-1)/2$ tests

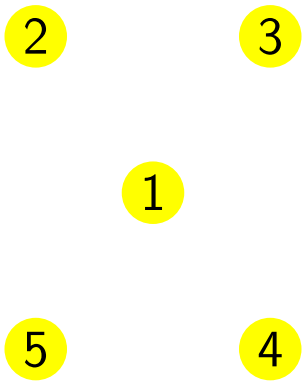
$$H_{0,h} \ : \ \text{“}\rho_h = 0\text{”} \quad \text{against} \quad H_{1,h} \ : \ \text{“}\rho_h \neq 0\text{”},$$

via the test statistic $T_{n,h}$ and p values $p_{n,h}$.
For instance, with empirical correlation,

$$T_{n,h}\left(\mathbb{X}\right) = \sqrt{n} \frac{\frac{1}{n}\sum_{j=1}^{n} X_i^{(j)} X_{i'}^{(j)} - \bar{X}_i \bar{X}_{i'}}{\hat{\sigma}(X_i)\hat{\sigma}(X_{i'})} = \sqrt{n}\hat{\rho}_{n,h}.$$

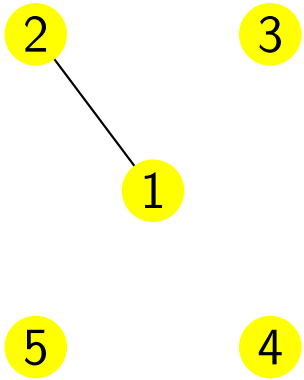
# Exemple

- $d = 5$.
- $\mathcal{H} = \{(1,2);(1,3);(1,4);(1,5);(2,3);(2,4);(2,5);(3,4);(3,5);(4,5)\}$.
- $$H_{0,h} \ : \ \text{“}\rho_h = 0\text{”} \quad \text{contre} \quad H_{1,h} \ : \ \text{“}\rho_h \neq 0\text{”}.$$
- Construction of the graph:



# Exemple

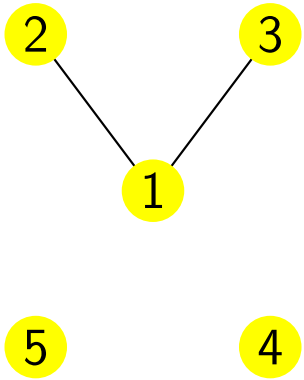
- $d = 5$.
- $\mathcal{H} = \{(1,2);(1,3);(1,4);(1,5);(2,3);(2,4);(2,5);(3,4);(3,5);(4,5)\}$.
- $$H_{0,h} \ : \ \text{“}\rho_h = 0\text{”} \quad \text{contre} \quad H_{1,h} \ : \ \text{“}\rho_h \neq 0\text{”}.$$
- Construction of the graph:




$$\mathcal{R} = \{(1,2)$$

# Exemple

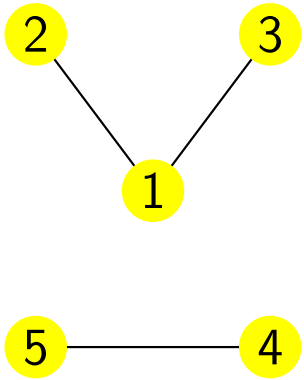
- $d = 5$.
- $\mathcal{H} = \{(1,2);(1,3);(1,4);(1,5);(2,3);(2,4);(2,5);(3,4);(3,5);(4,5)\}$.
- $$H_{0,h} \ : \ \text{“}\rho_h = 0\text{”} \quad \text{contre} \quad H_{1,h} \ : \ \text{“}\rho_h \neq 0\text{”}.$$
- Construction of the graph:




$$\mathcal{R} = \{(1,2);(1,3)$$

# Exemple

- $d = 5$.
- $\mathcal{H} = \{(1,2);(1,3);(1,4);(1,5);(2,3);(2,4);(2,5);(3,4);(3,5);(4,5)\}$.
- $$H_{0,h} \ : \ \text{“}\rho_h = 0\text{”} \quad \text{contre} \quad H_{1,h} \ : \ \text{“}\rho_h \neq 0\text{”}.$$
- Construction of the graph:




$$\mathcal{R} = \{(1,2);(1,3);(4,5)\}.$$

# Why it is important to take into account the multiplicity in the context of graphs?

- 51 brain regions: 1275 tests.
- If we test **true** null hypotheses **indepently** simultaneously at level of significance $\alpha = 5\%$.
- What is the probability to get at least one significant results by chance (*i.e.* reject a null hypothesis)?

$$\begin{aligned}\mathbb{P}(\text{“at least one significant result”}) &= 1 - \mathbb{P}(\text{“no significant results”}) \\ &= 1 - (1 - 0.05)^{1275} \\ &\simeq 1\end{aligned}$$

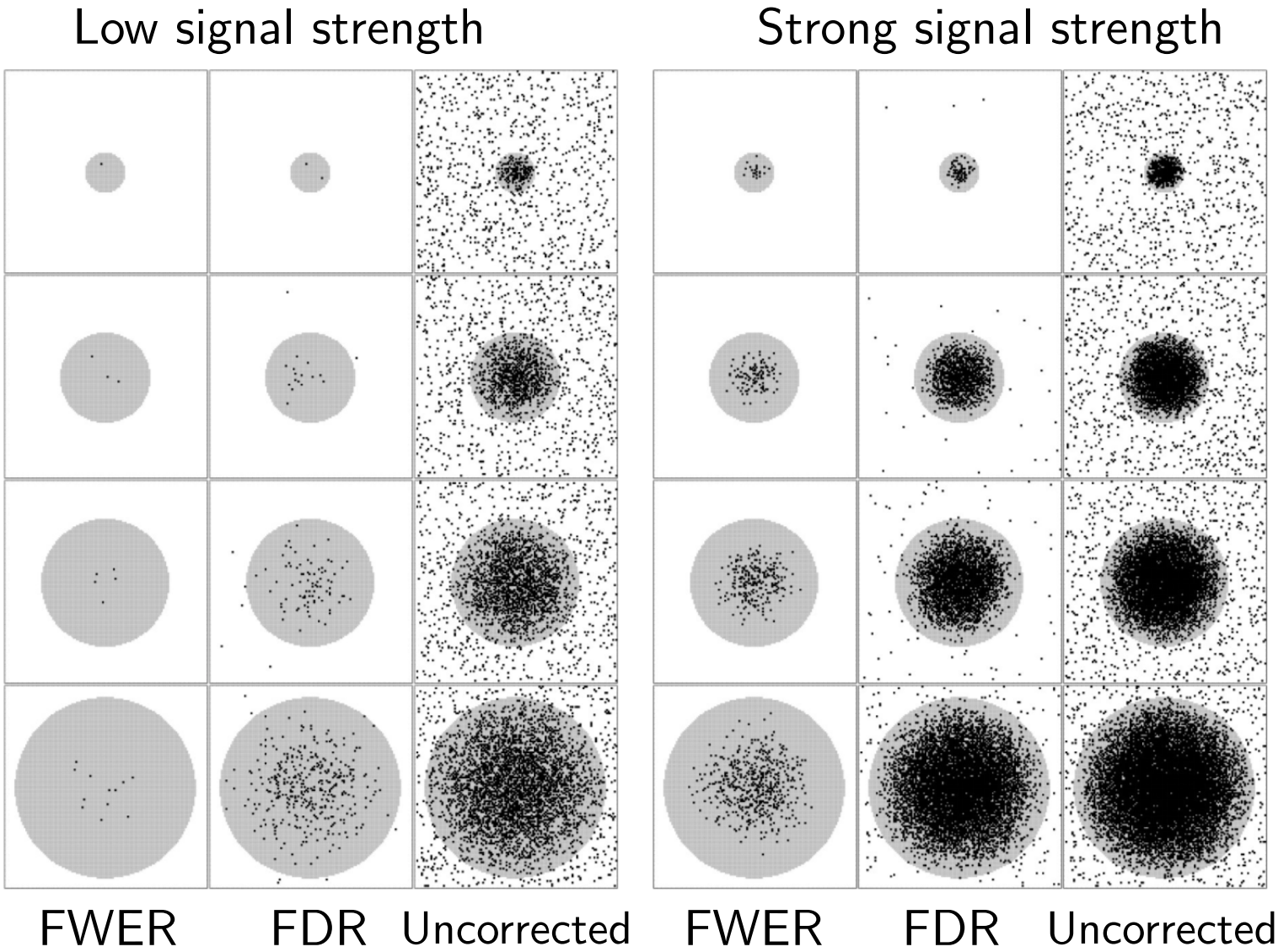
# Control of the error due to multiplicity

Two main criteria are usually used based on $\mathcal{Q}$, the proportion of false discovery

$$\mathcal{Q} = \frac{|\mathcal{R} \cap \mathcal{H}_0|}{|\mathcal{R}| \vee 1}.$$

- $\text{FWER}(\mathcal{R}) = \mathbb{P}(\mathcal{Q} > 0)$
- $\text{FDR}(\mathcal{R}) = \mathbb{E}[\mathcal{Q}]$

# Illustration FWER vs FDR




[**Roquain** 2007]

# asymptotic control of FWER

Our objective:

$$\limsup_{n \to +\infty} \text{FWER}(\mathcal{R}) \leq \alpha.$$

- Asymptotically in $n$
- Valid as soon as the statistics are converging
- Valid given any graph structures
- Can we evaluate the power?
- Can we extend to FDR?

# asymptotic control of FWER for correlations

For all $h \in \{1, \ldots, m\}$, we consider the test statistic $T_{n,h}\left(\mathbb{X}^{(n)}\right)$.
In the case of correlations,

$$T_{n,h}\left(\mathbb{X}^{(n)}\right) = \sqrt{n}\,\widehat{\rho}_{n,h}\left(\mathbb{X}^{(n)}\right).$$

Assume that,

$$\sqrt{n}\left(\,\hat{\rho}_{n,\cdot}(\mathbb{X}^{(n)}) - \rho(P)\right) \xrightarrow{d} \mathcal{N}_m(0, \Sigma), \quad \text{when } n \to +\infty,$$

where $\xrightarrow{d}$ denotes the convergence in distribution. We assume that $\Sigma$ is invertible.
Let $\left(p_{n,h}\left(\mathbb{X}^{(n)}\right)\right)_{1 \le h \le m}$ be a family of $p$-values resulting from each $m$ individual test:

$$\forall h \in \{1, \ldots, m\},\ p_{n,h}\left(\mathbb{X}^{(n)}\right) = 2\left[1 - \Phi\left(\frac{\left|\sqrt{n}\,\widehat{\rho}_{n,h}\left(\mathbb{X}^{(n)}\right)\right|}{\sqrt{1 - \widehat{\rho}^2_{n,h}\left(\mathbb{X}^{(n)}\right)}}\right)\right],$$

where $\Phi$ is the standard Gaussian cumulative distributive function.

# Bonferroni

- $\mathcal{R}^{bonf} = \left\{ h \in \mathcal{H} : p_{n,h} \leqslant \alpha/2m \right\}.$

**Theorem: [Dunn, 1961]**

The procedure $\mathcal{R}^{bonf}$ controls asymtotically the FWER at significance level $\alpha$,

$$\text{FWER}(\mathcal{R}^{bonf}) \leq \alpha.$$

# Šidák Inequality

**Proposition :**

Let $X \sim \mathcal{N}_m(0, \Sigma)$ with $\Sigma$ invertible, then for all $b \in \mathbb{R}_+^m$,

$$\mathbb{P}(|X| \leq b) \geq \prod_{i=1}^{m} \mathbb{P}(|X_i| \leq b_i).$$

# Šidák

- $\mathcal{R}^s = \left\{ h \in \mathcal{H} : |T_{n,h}| > \Phi^{-1}\left( \frac{1}{2}(1-\alpha)^{(1/m)} + \frac{1}{2} \right) \right\}.$

**Theorem: [Westfall, Young 1993 and Drton, Perlmann 2004]**

The procedure $\mathcal{R}^s$ controls asymtotically the FWER at significance level $\alpha$,

$$\limsup_{n \to +\infty} \text{FWER}(\mathcal{R}^s) \leq \alpha.$$

# Remarks

- Procedure valid for all types of dependence (Gaussian case).
- FWER controlled such as in the simple case!

# Summary of statistical framework

$\mathbf{X} = \{\mathbf{X}(k), k \in \mathbb{Z}\}$ long memory process, $1 \leqslant \ell, m \leqslant d$,

## Statistical Hypotheses

$\mathcal{H}_0^{(h)} : \rho_{\ell,m} = 0 \quad \mathcal{H}_1^{(h)} : \rho_{\ell,m} \neq 0$

$M = \frac{d(d-1)}{2}$ hypotheses number

$\mathcal{R}$ = set of rejected null hypotheses

Tests are dependent!

## Choice of Statistics

$\widehat{\rho}_{n,h}$ s.t. when $n \to +\infty$,

$$\sqrt{n}\left(\hat{\rho}_{n,\cdot}(\mathbb{X}^{(n)}) - \rho(P)\right) \longrightarrow \mathcal{N}_m(0, \Sigma),$$

evaluation of $p$-values $= p_{n,h}$

## Controlling procedure

$$p_{n,h} < \alpha/M$$

$$\hat{h} = \max\{h \in \{0, \ldots, m\} : p_{(n,h)} \leq \alpha h/m\}$$

TestCorr

## Choice of criterion

$$\mathrm{FWER}(\mathcal{R}, P) = \mathbb{P}(|\mathcal{R} \cap \mathcal{H}_0(P)| \geq 1)$$

$$\mathrm{FDR}(\mathcal{R}, P) = \mathbb{E}\left[\frac{|\mathcal{R} \cap \mathcal{H}_0(P)|}{|\mathcal{R}| \vee 1}\right]$$

[**Roux** *et al.* 2018] [**Hero** *et al.* 2013]
[**Drton** *et al.* 2004]

# Simulations

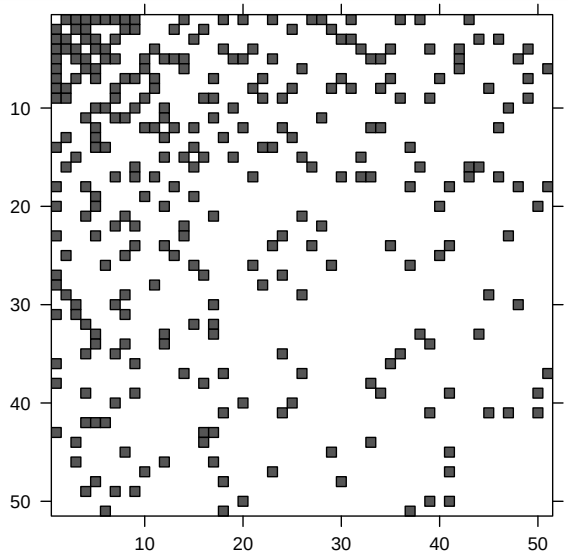
Simulate $\mathbf{M}$ such that $\mathbf{M} = I + \rho A$, where $\mathbf{I}$ is the identity matrix, and $\rho$ is a real.

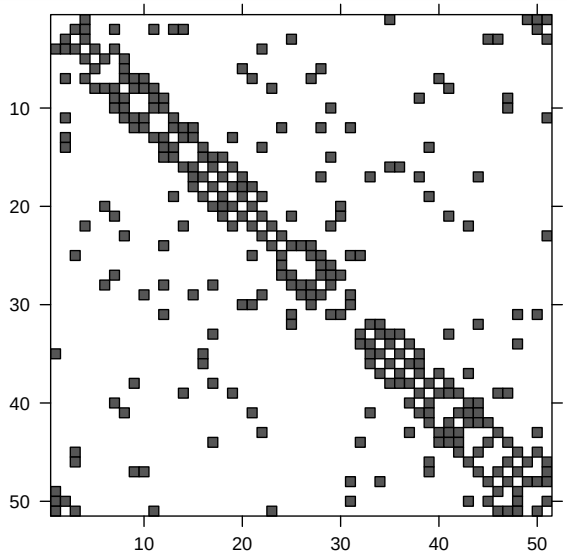
$\mathbf{M}$ is positive definite if and only if $|\rho| < \frac{1}{|\lambda_{min}|}$, where $\lambda_{min}$ is the smallest eigenvalue.

Let us take $d = 51$, $m = 150$. This is true for $\rho = 0.2$ for 4 different structures: Erdos-Renyi, Stochastic Block Model, Small-World and Scale-free.

# Network structures used in simulations

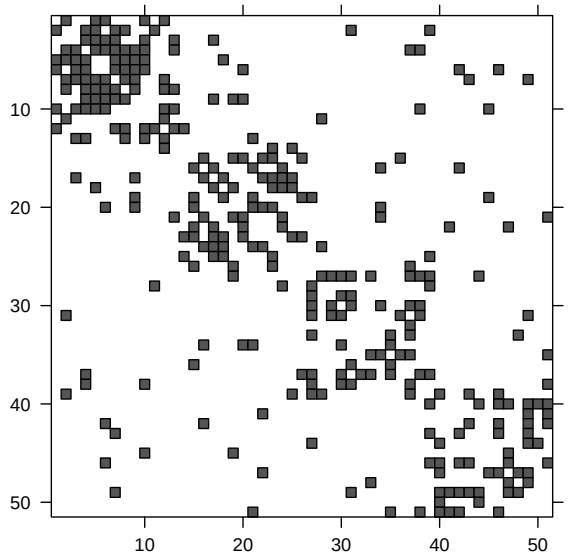
Scale-free

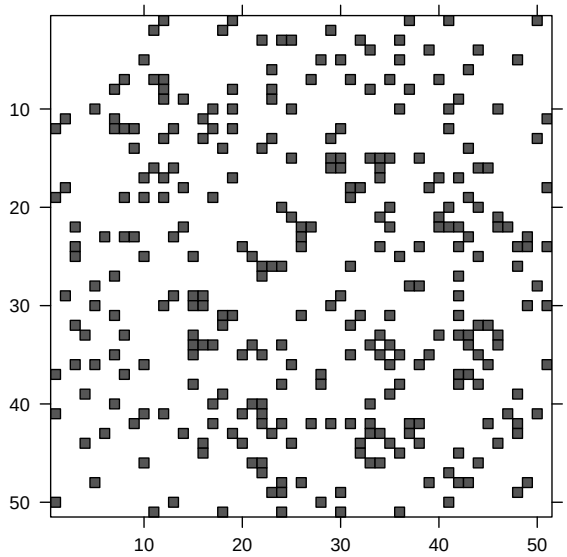



Small World

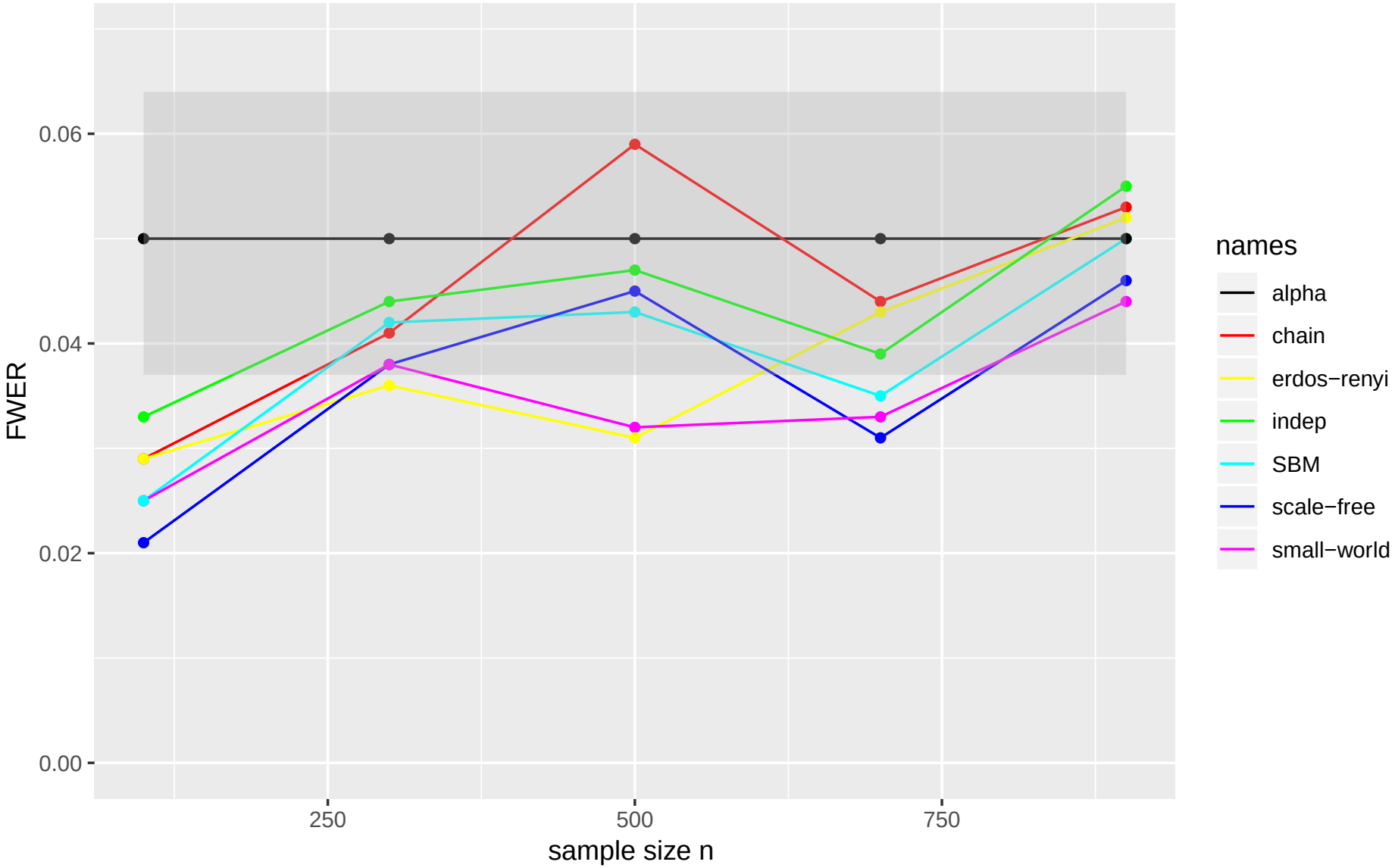
SBM



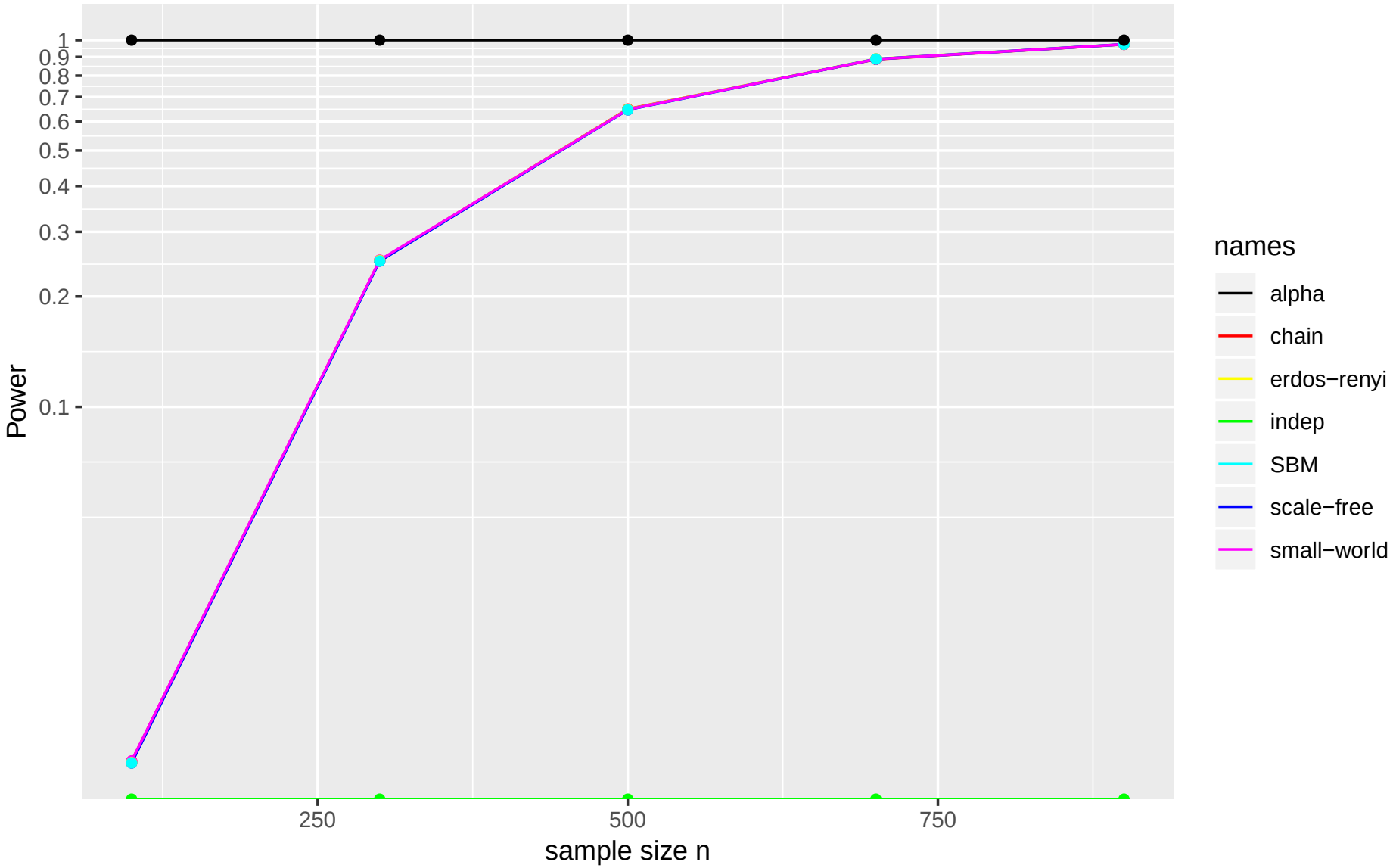




Erdös-Rényi

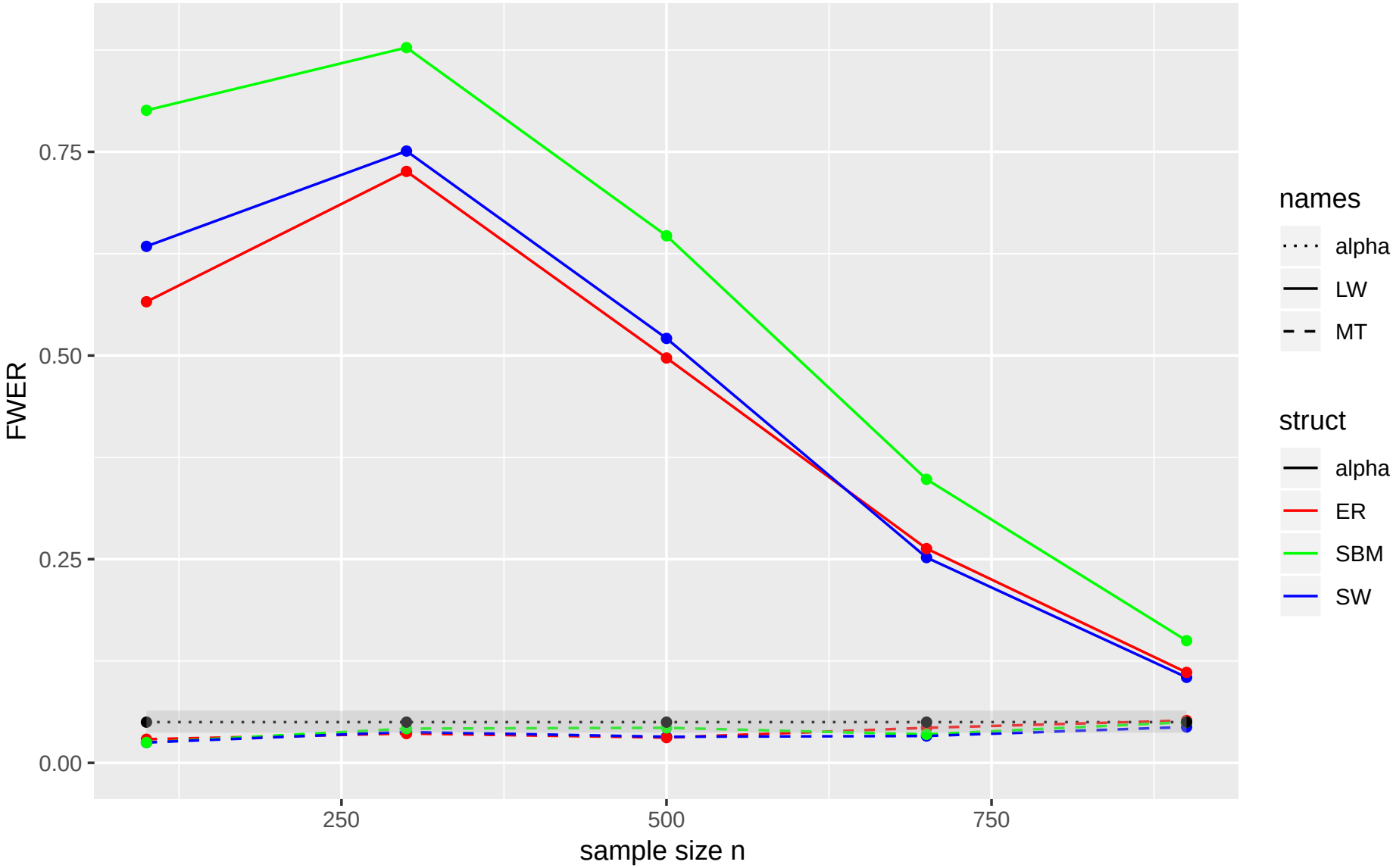
# using Bonferroni: FWER control is preserved as expected



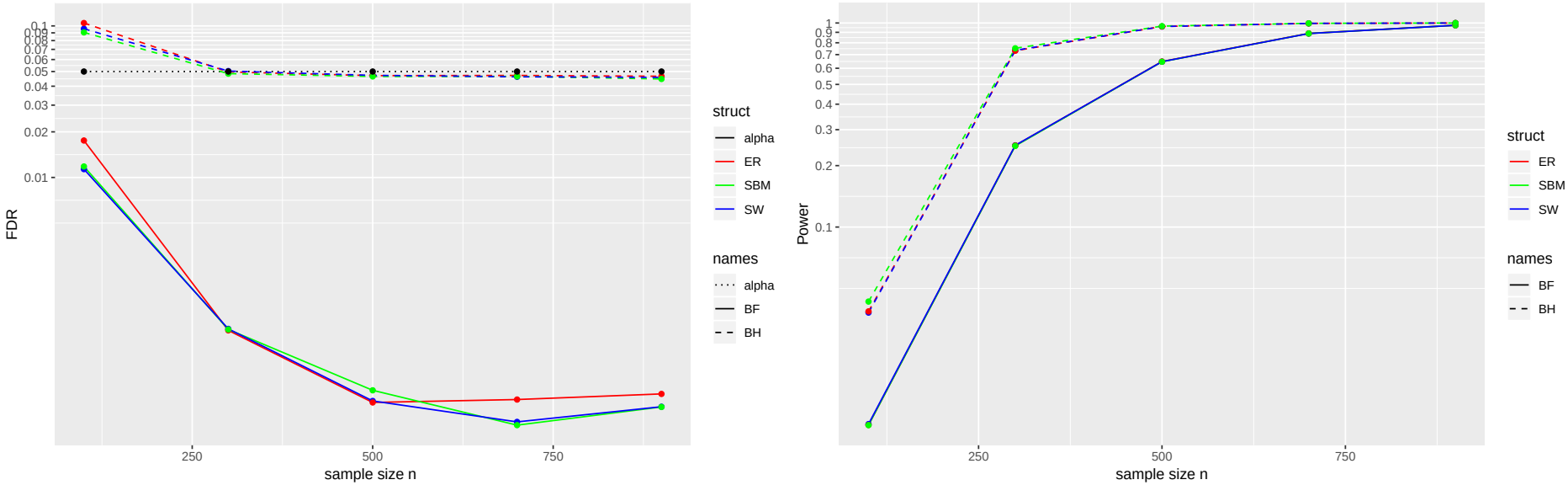
# using Bonferroni: power



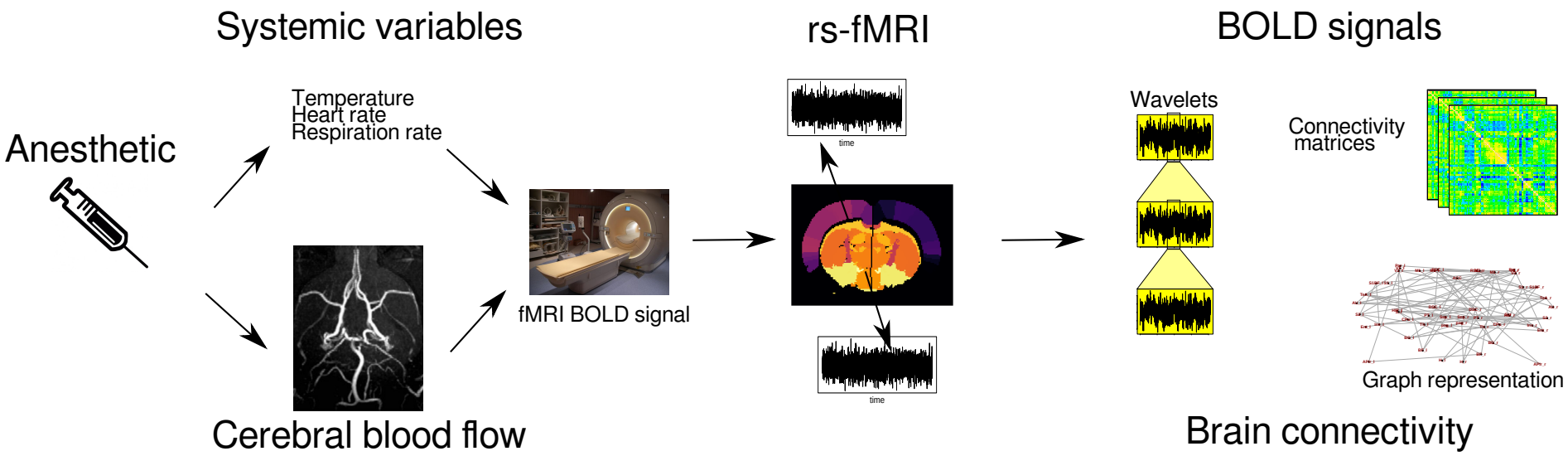
# Comparisons with shrinkage (Ledoit-Wolf)



# Comparisons with FDR



# Real data

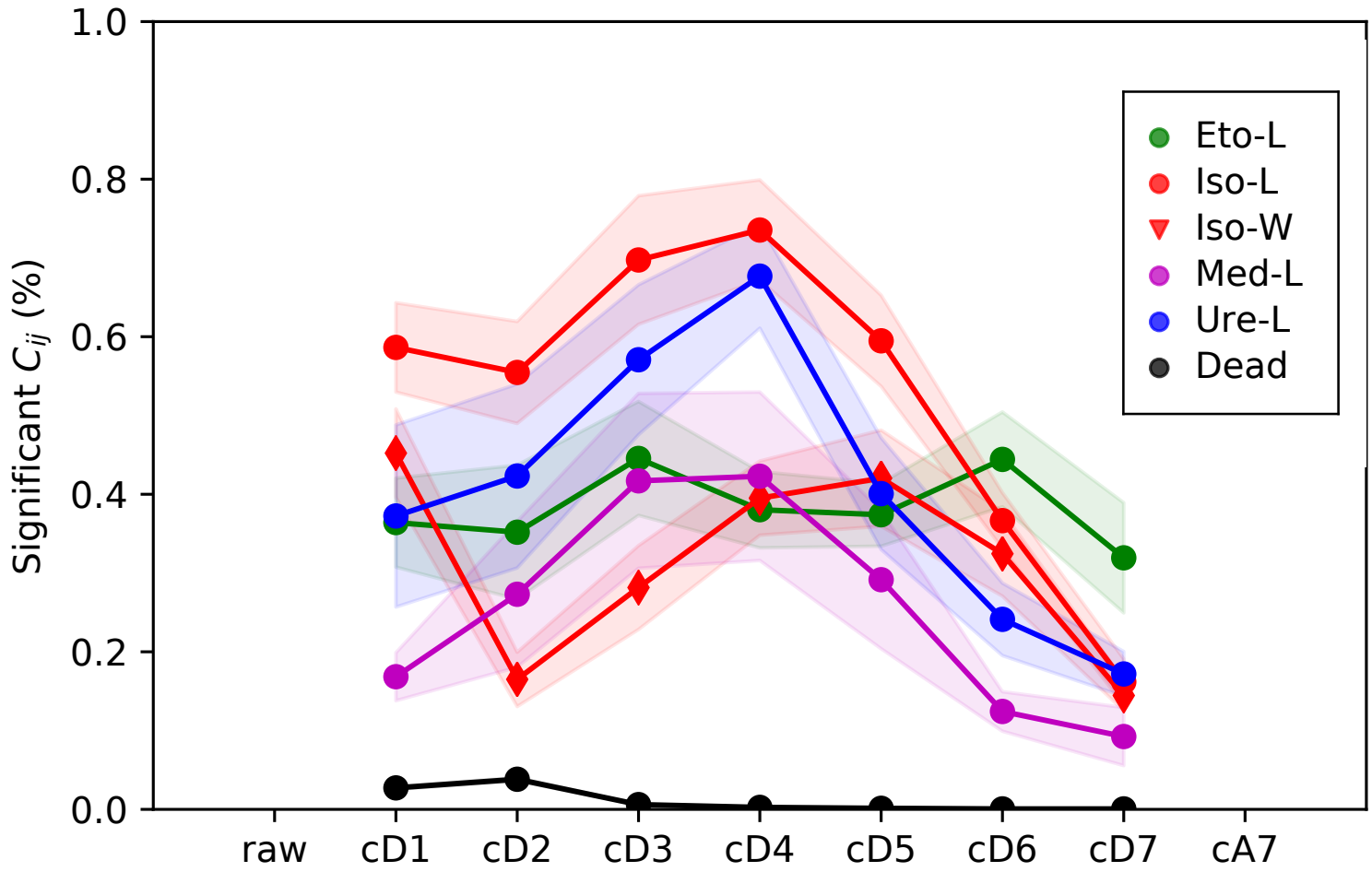



- 4 anesthetics: Isoflurane, Medetomidine, Edetomidate, Urethane
- 5 groups of rats
- 30-minutes resting-state fMRI, TR=0.5s

[**Becq** *et al.* 2018, in preparation]

# Significant edges detection

Opitimisation of signal to noise ratio on edge detection

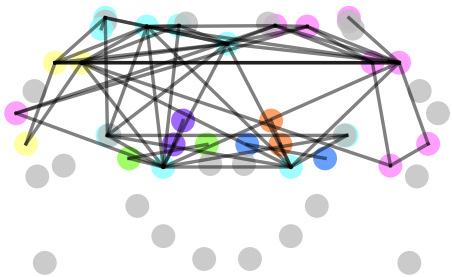


# Graphs for anesthetics

## Spatial reorganization of 5% highest correlation pairs

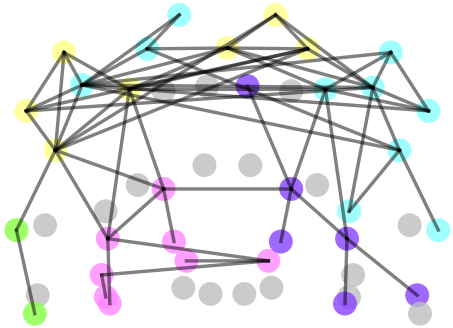
ETO_L

N = 7
C > 0.49
density: 0.05



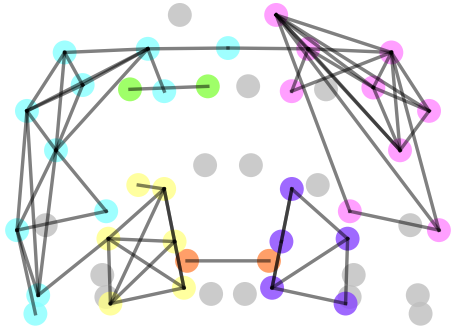
ISO_L

N = 6
C > 0.69
density: 0.05



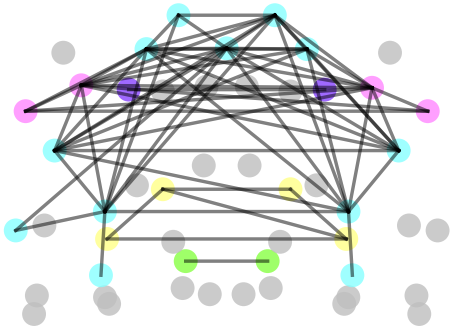
URE_L

N = 7
C > 0.68
density: 0.05



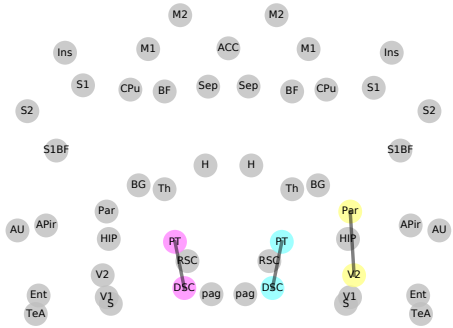
MED_L

N = 7
C > 0.57
density: 0.05



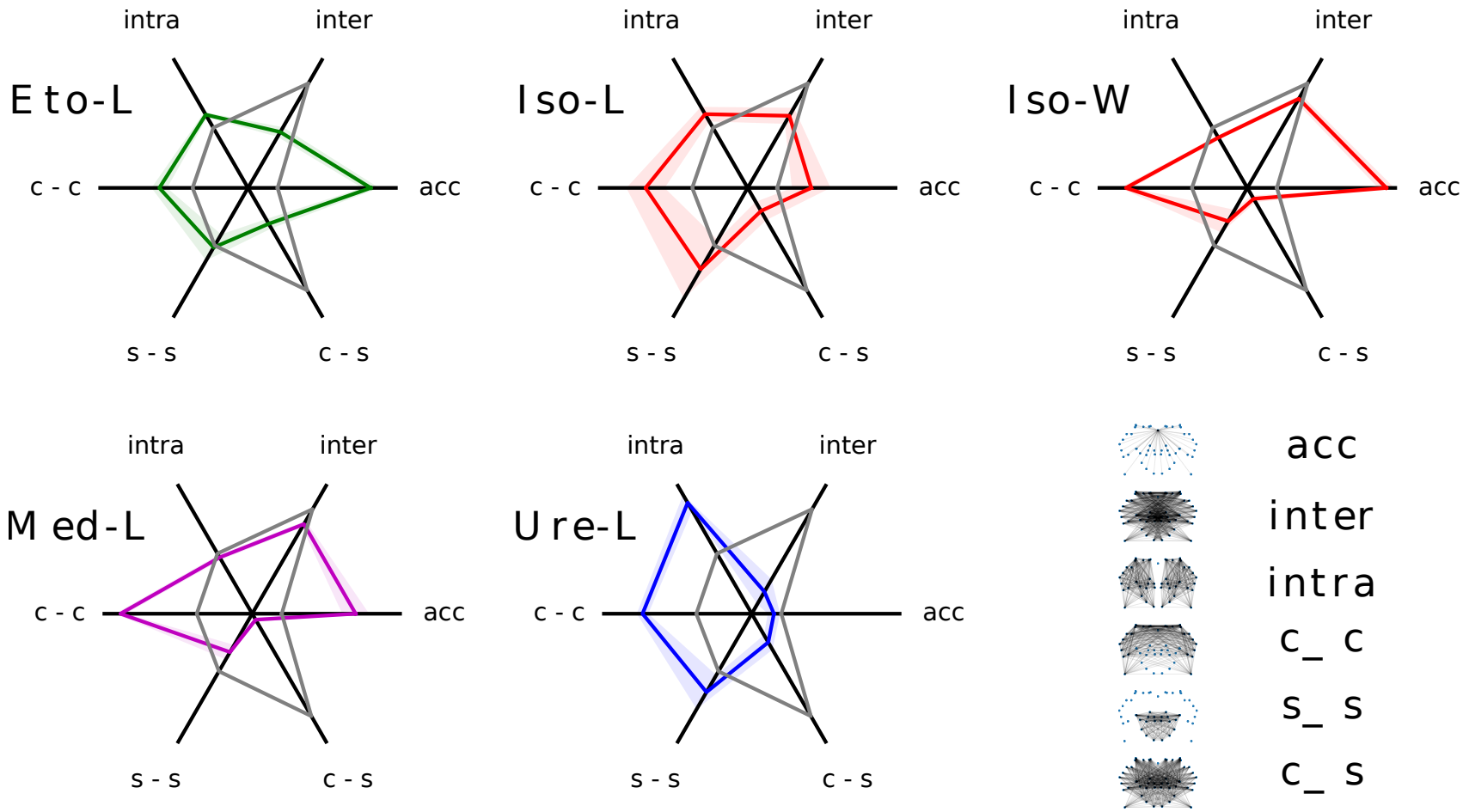
DEAD

N = 4
C > 0.27
density: 0.0



# Anesthetics graphs profiles

Spatial reorganisation of 5% highest correlation pairs



# Conclusion and perspectives

**Conclusion**

- Control of false positive is not easy
- Graph networks properties depend on false positive

**Perspectives**

- Control of false discovery rate may be more appropriate
- Extension to any other mesures of correlation or partial correlation

# Thanks to my collaborators