Constructing credible region for comparison problems in fMRI graph analysis

Alice Chevaux†, Julyan Arbel†, Guillaume Kon Kam King*, Sophie Achard†

†Univ. Grenoble Alpes, CNRS, Inria, LJK, F-38000 Grenoble, France
*Université Paris-Saclay, INRAE, MaIAGE, 78350 Jouy-en-Josas, France

Séminaire Doctorants Inria (Janvier 2026)

1. Introduction
2. Credible Region Construction for Multivariate Distribution
3. Detection of the non-zeros in the matrix
4. Application to real datasets: Do the credible regions overlap ?
5. Conclusion

Motivations for fMRI data

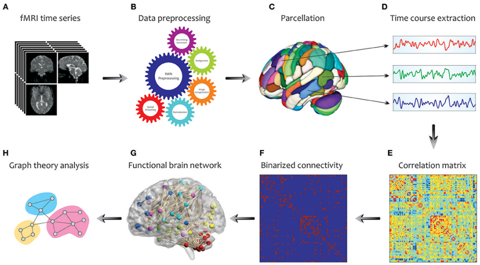


Figure 1: Illustration of the usual inference of graph for fMRI data

- High dimension problem: 100×100 coefficients to estimate
- Graph analysis does not take into account the uncertainty from the correlation estimator

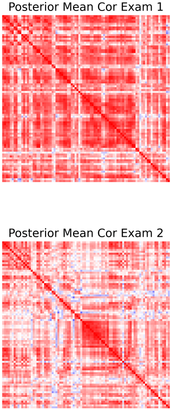


Figure 2: Correlation matrices estimation for the same subject at two different time (D+30 and D+60 after coma)

Empirical correlation matrix

- $\boldsymbol{X} = (X_1, \ldots, X_p)^\top \longrightarrow \mathcal{N}(0, \Sigma)$
- T i.i.d realisations of $\boldsymbol{X}$
- Objective: estimate the correlation matrix

$$\mathrm{R} = \mathrm{D}^{-1}\boldsymbol{\Sigma}\mathrm{D}^{-1} \quad \mathrm{D} = \mathrm{diag}(\Sigma_{11}^{1/2}, \ldots, \Sigma_{pp}^{1/2})$$

- The empirical covariance matrix is given by:

$$\widehat{\boldsymbol{\Sigma}} = \frac{1}{T}\mathrm{y}^\top \mathrm{y}$$

- Defining $\mathrm{D} = \mathrm{diag}(\widehat{\Sigma}_{11}^{1/2}, \ldots, \widehat{\Sigma}_{pp}^{1/2})$, we can define the empirical correlation matrix:

$$\widehat{\mathrm{R}} = \mathrm{D}^{-1}\widehat{\boldsymbol{\Sigma}}\mathrm{D}^{-1}$$

Empirical correlation matrix

- $\boldsymbol{X} = (X_1, \dots, X_p)^\top \longrightarrow \mathcal{N}(0, \Sigma)$
- T i.i.d realisations of $\boldsymbol{X}$
- Objective: estimate the correlation matrix

$$\mathrm{R} = \mathrm{D}^{-1}\boldsymbol{\Sigma}\mathrm{D}^{-1} \quad \mathrm{D} = \mathrm{diag}(\Sigma_{11}^{1/2}, \dots, \Sigma_{pp}^{1/2})$$

- The empirical covariance matrix is given by:

$$\widehat{\boldsymbol{\Sigma}} = \frac{1}{T}\mathrm{y}^\top \mathrm{y}$$

- Defining $\mathrm{D} = \mathrm{diag}(\widehat{\Sigma}_{11}^{1/2}, \dots, \widehat{\Sigma}_{pp}^{1/2})$, we can define the empirical correlation matrix:

$$\widehat{\mathrm{R}} = \mathrm{D}^{-1}\widehat{\boldsymbol{\Sigma}}\mathrm{D}^{-1}$$

- ▶ Unstable estimator in fMRI because $T \approx p$ and we have $\frac{p(p-1)}{2}$ coefficients to estimate
- ▶ Our proposal: quantify the uncertainty of correlation estimation and interpret it for several applications

Uncertainty quantification

Why focus on uncertainty quantification and not point estimate?

Uncertainty quantification

Why focus on uncertainty quantification and not point estimate?

- To quantify uncertainty, which is large

Uncertainty quantification

Why focus on uncertainty quantification and not point estimate?

- To quantify uncertainty, which is large
- To detect the significantly non-zero coefficients of the correlation matrix

Uncertainty quantification

Why focus on uncertainty quantification and not point estimate?

- To quantify uncertainty, which is large
- To detect the significantly non-zero coefficients of the correlation matrix
- To compare two fMRI scans

Bayesian Statistics: Quantifying Uncertainty

Key Paradigm Shift: From Fixed Parameter to Random Variable

Frequentist View

θ is a **fixed** unknown constant

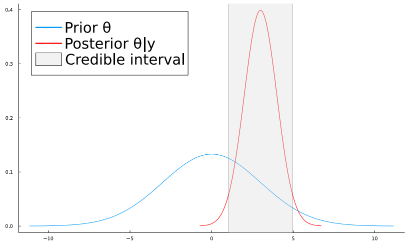
- Point estimate $\hat{\theta}$
- Confidence interval: long-run frequency property

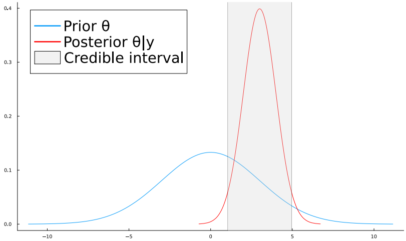
Bayesian View

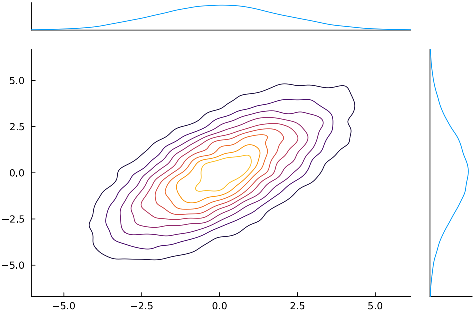
θ is a **random** variable with a prior distribution

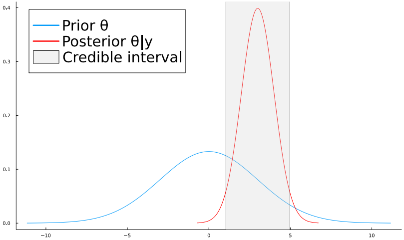
- Full posterior distribution $P(\theta|data)$
- Credible interval: $P(\theta \in \text{interval}|data) = 0.95$

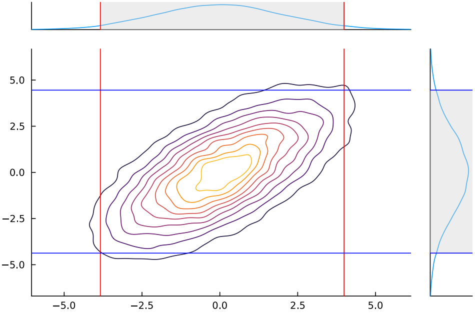
- Natural way to quantify uncertainty through the posterior distribution
- Uncertainty can be propagated for every function $g(\theta)$ via the posterior of $g(\theta|\text{data})$


Prior θ
Posterior θ|y
Credible interval
0.4
0.3
0.2
0.1
0.0
−10
−5
0
5
10


Prior θ
Posterior θ|y
Credible interval
0.4
0.3
0.2
0.1
0.0
−10
−5
0
5
10


5.0
2.5
0.0
−2.5
−5.0
−5.0
−2.5
0.0
2.5
5.0


Prior θ
Posterior θ|y
Credible interval
0.4
0.3
0.2
0.1
0.0
−10
−5
0
5
10


5.0
2.5
0.0
−2.5
−5.0
−5.0
−2.5
0.0
2.5
5.0

Bayesien Model for correlation matrix

- Model:

$$\begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{pmatrix} \sim \mathcal{N}_p\left(0, \Sigma\right)$$

- We have T realisations of X that we will note $\mathrm{y} \in \mathbb{R}^{T \times p}$
- Prior Inverse Wishart:

$$\Sigma \sim \mathcal{IW}(\Phi, \nu)$$

- Conjugate Posterior:

$$\Sigma|\mathrm{y} \sim \mathcal{IW}(\Phi + \mathrm{y}^T\mathrm{y}, \nu + T)$$

Bayesien Model for correlation matrix

- Model:

$$\begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{pmatrix} \sim \mathcal{N}_p(0, \Sigma)$$

- We have T realisations of X that we will note $\mathrm{y} \in \mathbb{R}^{T \times p}$
- Prior Inverse Wishart:

$$\Sigma \sim \mathcal{IW}(\Phi, \nu)$$

- Conjugate Posterior:

$$\Sigma|\mathrm{y} \sim \mathcal{IW}(\Phi + \mathrm{y}^T\mathrm{y}, \nu + T)$$

- At this point we can:
  1. Sample covariance matrices Σ from the posterior distribution in the space $PSD(\mathbb{R}^{p \times p})$
  2. Compute the corresponding correlation matrices $\mathrm{R} \in PSD(\mathbb{R}^{p \times p})$
  3. Project the upper triangle matrix of R in a vectorial space $\mathbb{R}^{\frac{p(p-1)}{2}}$

Choice for the credible region

- Parameter of interest: the triangular superior matrix
  $\theta = (\theta_{12}, \theta_{13}, \cdots, \theta_{p(p-1)}) \in \mathbb{R}^{\frac{p(p-1)}{2}}$
- A region R is a credible region at a level $1-\alpha$ for θ if $\mathbb{P}(\theta \in R|y) \geq 1-\alpha$

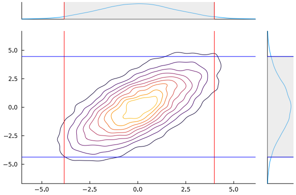
Choice for the credible region

- Parameter of interest: the triangular superior matrix
  $\theta = (\theta_{12}, \theta_{13}, \cdots, \theta_{p(p-1)}) \in \mathbb{R}^{\frac{p(p-1)}{2}}$
- A region R is a credible region at a level $1-\alpha$ for θ if $\mathbb{P}(\theta \in R|y) \geq 1-\alpha$
- We will work with the product of bilateral or unilateral quantile-based credible intervals all at the **same level of quantile** $t^\star$:

$$R_\theta(t^\star) = \prod_{i<j} [q_{ij}(\frac{t^\star}{2}), q_{ij}(1-\frac{t^\star}{2})]$$

$$\text{or} \quad R_\theta(t^\star) = \prod_{i<j} [q_{ij}(t^\star), +\infty[$$

- Objective: Find $t^\star$ such as $\mathbb{P}(\theta \in R_\theta(t^\star)|y) \geq 1-\alpha$



Existence of such region

- **Naive** method could be to use $t^\star = \alpha$, however the product of intervals at a level $1-\alpha$ does not have a probability of $1-\alpha$

Existence of such region

- **Naive** method could be to use $t^\star = \alpha$, however the product of intervals at a level $1-\alpha$ does not have a probability of $1-\alpha$

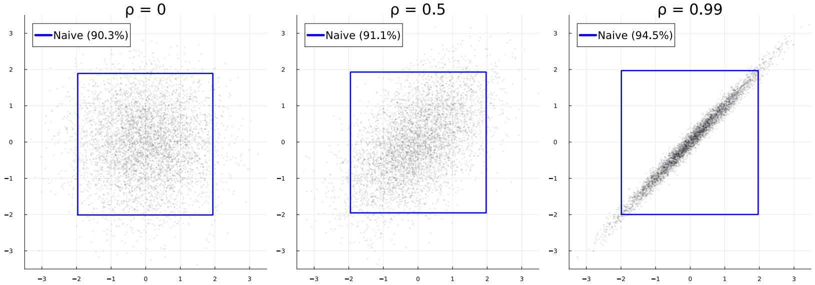
- A sufficient quantile level is $t^\star = \frac{\alpha}{\frac{p(p-1)}{2}}$. We define the corresponding region as a **Bonferroni-type** credible region:

$$R_{\text{Bonf}}^{(1-\alpha)} = R_\theta\left(\frac{\alpha}{\frac{p(p-1)}{2}}\right)$$

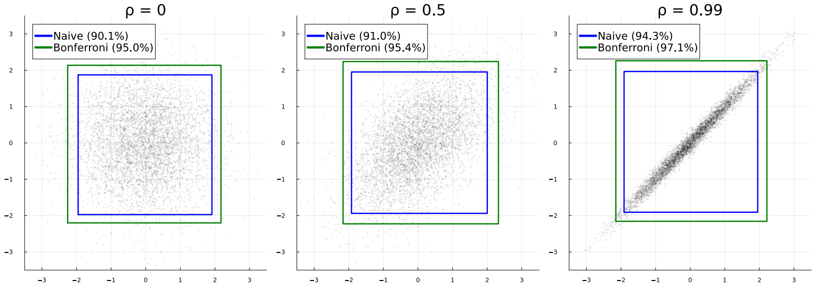
- This region verifies:

$$\mathbb{P}\left((\theta_{12}, \theta_{13}, \cdots, \theta_{(p-1)p}) \in R_{\text{Bonf}}^{(1-\alpha)} | y\right) \geq 1-\alpha$$

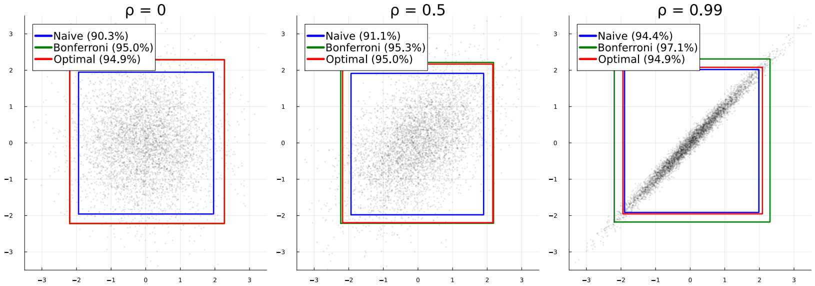
Example in a 2-dimensional setup



Example in a 2-dimensional setup



Example in a 2-dimensional setup



How to find the optimal rectangle region?

Refining the region

- Find numerically a level $t^\star$ between $\frac{\alpha}{\frac{p(p-1)}{2}}$ and α such that

$$\mathbb{P}\left(\theta \in R(t^*)|y\right) = 1 - \alpha$$

- In a Bayesian set-up we can use the posterior joint distribution to evaluate the posterior probability of a region of our choice
- Similar rectangle can be find in the state of the art but we reduced the memory cost from $\mathcal{O}(p^4)$ to $\mathcal{O}(p^2)$ to apply it to our datasets

1. Introduction
2. Credible Region Construction for Multivariate Distribution
3. Detection of the non-zeros in the matrix
4. Application to real datasets: Do the credible regions overlap ?
5. Conclusion

Multiple-testing Error Control

We want to find for each θ_{ij} if they are significantly greater than zero.

- $\frac{p(p-1)}{2}$ hypothesis tests $H_{0,ij} : \theta_{ij} = 0$ and $H_{1,ij} : \theta_{ij} > 0$

Table 1: Terminology on one test

Test Decision	**Ground Truth**	
	$\theta_{ij} > 0$	$\theta_{ij} = 0$
Reject $H_{0,ij}$	True Positive	False Positive
Accept $H_{0,ij}$	False Negative	True Negative

- Usually an hypothesis test is construct to control the risk of False positive at a level α
- If we run 100 tests on independant data, we will likely have $\alpha \times 100$ false positives

Multiple-testing Error Control

We want to find for each θ_{ij} if they are significantly greater than zero.

- $\frac{p(p-1)}{2}$ hypothesis tests $H_{0,ij} : \theta_{ij} = 0$ and $H_{1,ij} : \theta_{ij} > 0$

Table 1: Terminology on one test

Test Decision	**Ground Truth**	
	$\theta_{ij} > 0$	$\theta_{ij} = 0$
Reject $H_{0,ij}$	True Positive	False Positive
Accept $H_{0,ij}$	False Negative	True Negative

- Usually an hypothesis test is construct to control the risk of False positive at a level α
- If we run 100 tests on independant data, we will likely have $\alpha \times 100$ false positives

$$\text{FWER(Family-Wise Error Rate)} = \mathbb{P}\left(\text{there is at least one false positive} \mid \bigcap_{i<j} H_{0,ij}\right)$$

Estimator of the support of the matrix

Using the optimal unilateral region we have:

$$\mathbb{P}\left(\theta \in R(t^*)|\, y\right) = \mathbb{P}\left(\theta \in \prod_{i<j}[q_{ij}(t^\star), +\infty[|y\right) = \mathbb{P}\left(\bigcap_{i<j}\{\theta_{ij} \geq q_{ij}^{(t^*)}\}|y\right) = 1 - \alpha$$

Estimator of the support of the matrix

Using the optimal unilateral region we have:

$$\mathbb{P}\left(\theta \in R(t^*) | y\right) = \mathbb{P}\left(\theta \in \prod_{i<j}[q_{ij}(t^\star), +\infty[|y\right) = \mathbb{P}\left(\bigcap_{i<j}\{\theta_{ij} \geq q_{ij}^{(t^*)}\}|y\right) = 1-\alpha$$

For each dimension we use the decision rule:

$$\text{Consider } \theta_{ij} > 0 \text{ if } q_{ij}^{(t^*)} > 0$$

Estimator of the support of the matrix

Using the optimal unilateral region we have:

$$\mathbb{P}\left(\theta \in R(t^*) | y\right) = \mathbb{P}\left(\theta \in \prod_{i<j} [q_{ij}(t^\star), +\infty[| y\right) = \mathbb{P}\left(\bigcap_{i<j} \{\theta_{ij} \geq q_{ij}^{(t^*)}\} | y\right) = 1 - \alpha$$

For each dimension we use the decision rule:

$$\text{Consider } \theta_{ij} > 0 \text{ if } q_{ij}^{(t^*)} > 0$$

$$\begin{aligned}
\text{FWER}|y &= \mathbb{P}\left(\{\text{there is at least one false positive}\} | y\right) \\
&= \mathbb{P}\left(\bigcup_{i<j} \{q_{ij}^{(t^*)} > 0 \text{ and } \theta_{ij} \leq 0\} | y\right) \\
&\leq \mathbb{P}\left(\bigcup_{i<j} \{\theta_{ij} < q_{ij}^{(t^*)}\} | y\right) \\
&= 1 - \mathbb{P}\left(\bigcap_{i<j} \{\theta_{ij} \geq q_{ij}^{(t^*)}\} | y\right) \\
&= 1 - (1 - \alpha) = \alpha
\end{aligned}$$

Results on simulation

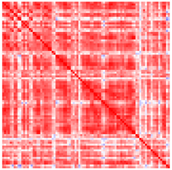
T	True Positive Rate / Method	2.4% TP Accuracy	2.4% TP FWER	24% TP Accuracy	24% TP FWER	48% TP Accuracy	48% TP FWER
50	Multiple test (Bonferroni)	0.9827	0.077	0.8071	0.052	0.4194	0.0211
	Multiple test (Holm-Bonf)	0.9818	0.051	0.7978	0.041	0.3896	0.017
	Bayesian (Bonf threshold)	0.9818	0.051	0.7975	0.039	0.392	0.0105
	Bayesian (Optimal threshold)	**0.9829**	0.094	**0.8108**	0.068	**0.434**	0.0316
100	Multiple test (Bonferroni)	0.987	0.064	0.8558	0.052	0.5547	0.0225
	Multiple test (Holm-Bonf)	0.9868	0.049	0.8482	0.042	0.5416	0.017
	Bayesian (Bonf threshold)	0.9868	0.048	0.8473	0.038	0.5276	0.0112
	Bayesian (Optimal threshold)	**0.9872**	0.08	**0.8586**	0.063	**0.5674**	0.0225
500	Multiple test (Bonferroni)	0.9922	0.046	0.9482	0.046	0.8569	0.0122
	Multiple test (Holm-Bonf)	**0.9938**	0.045	0.9482	0.053	0.8588	0.031
	Bayesian (Bonf threshold)	**0.9938**	0.043	0.9473	0.044	0.8505	0.0122
	Bayesian (Optimal threshold)	0.9923	0.055	**0.9494**	0.061	**0.8613**	0.0244

- Gain in accuracy with control over the FWER (conditionally to different things)

Overlapping credible region

How can we know if two correlation matrices associated to two different fMRI scans are different?

- Impossible with only 2 point estimates
- Possible with two posterior distributions



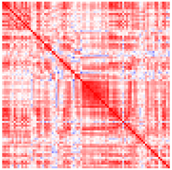


Figure 3: Correlation matrices estimation for the same subject at two different time (D+30 and D+60 after coma)

Overlapping credible region

How can we know if two correlation matrices associated to two different fMRI scans are different?

- Impossible with only 2 point estimates
- Possible with two posterior distributions

- Methodology: check if two credible region at a level $1 - \alpha$ overlap
- Rectangle region overlap $\iff$ Every interval overlap

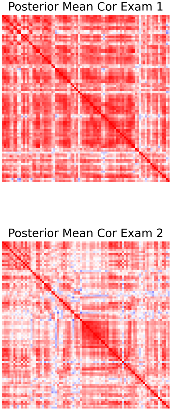


Figure 3: Correlation matrices estimation for the same subject at two different time (D+30 and D+60 after coma)

Dataset presentation

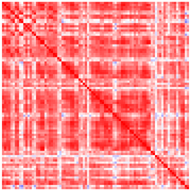
To have a reference we use dataset were we have the same patient who is scanned at two different times:

- 12 patients that were in a coma at the first scan, and conscious for the second exam

- Clinical problem: How to characterize the evolution of the patients between the two scans?

Illustration of one patient

Posterior Mean Cor Exam 1



Posterior Mean Cor Exam 2

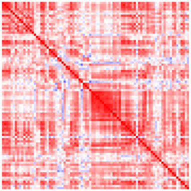


Illustration of one patient

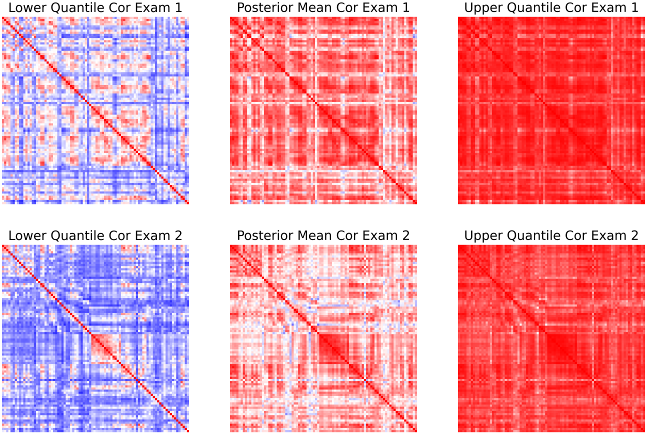
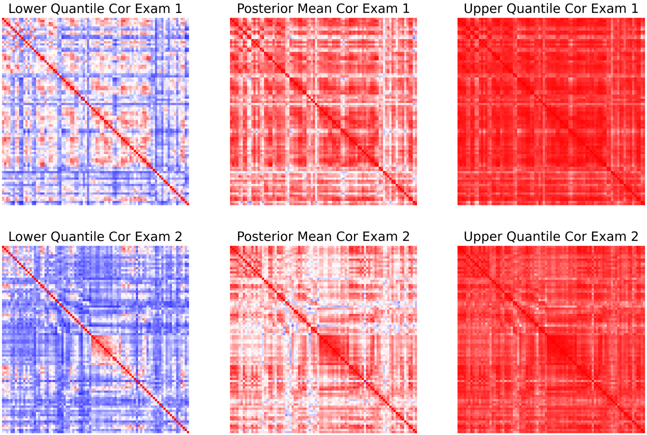


Illustration of one patient



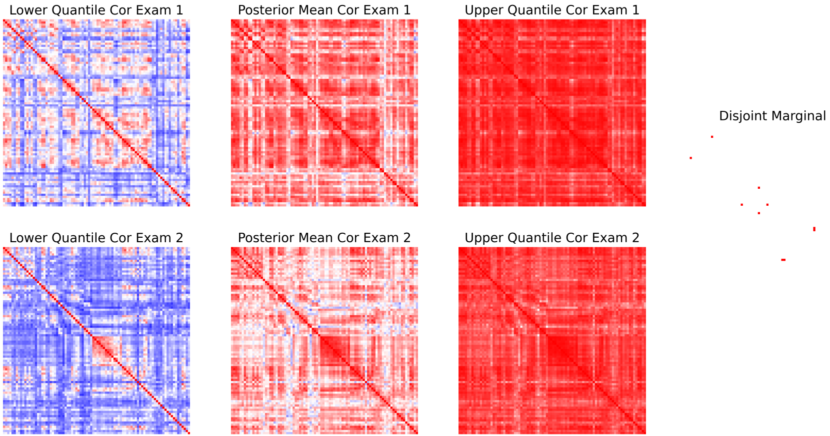
Credible intervals for θ_{12}

- Exam 1: $[0.4416, 0.9449]$

- Exam 2: $[0.5249, 0.9561]$

▶ Intervals overlap

Illustration of one patient



Conclusion

Contributions:

- Bring new methodology to some applications that were not explored before
- For the fMRI community this work advocates for humility over the results and highlights the need for uncertainty quantification

Conclusion

Contributions:

- Bring new methodology to some applications that were not explored before
- For the fMRI community this work advocates for humility over the results and highlights the need for uncertainty quantification

Perspectives:

- Extend the uncertainty quantification on the graph analysis
- Explore the representation of these regions in the space of semi-definite positive matrices

Thank you!