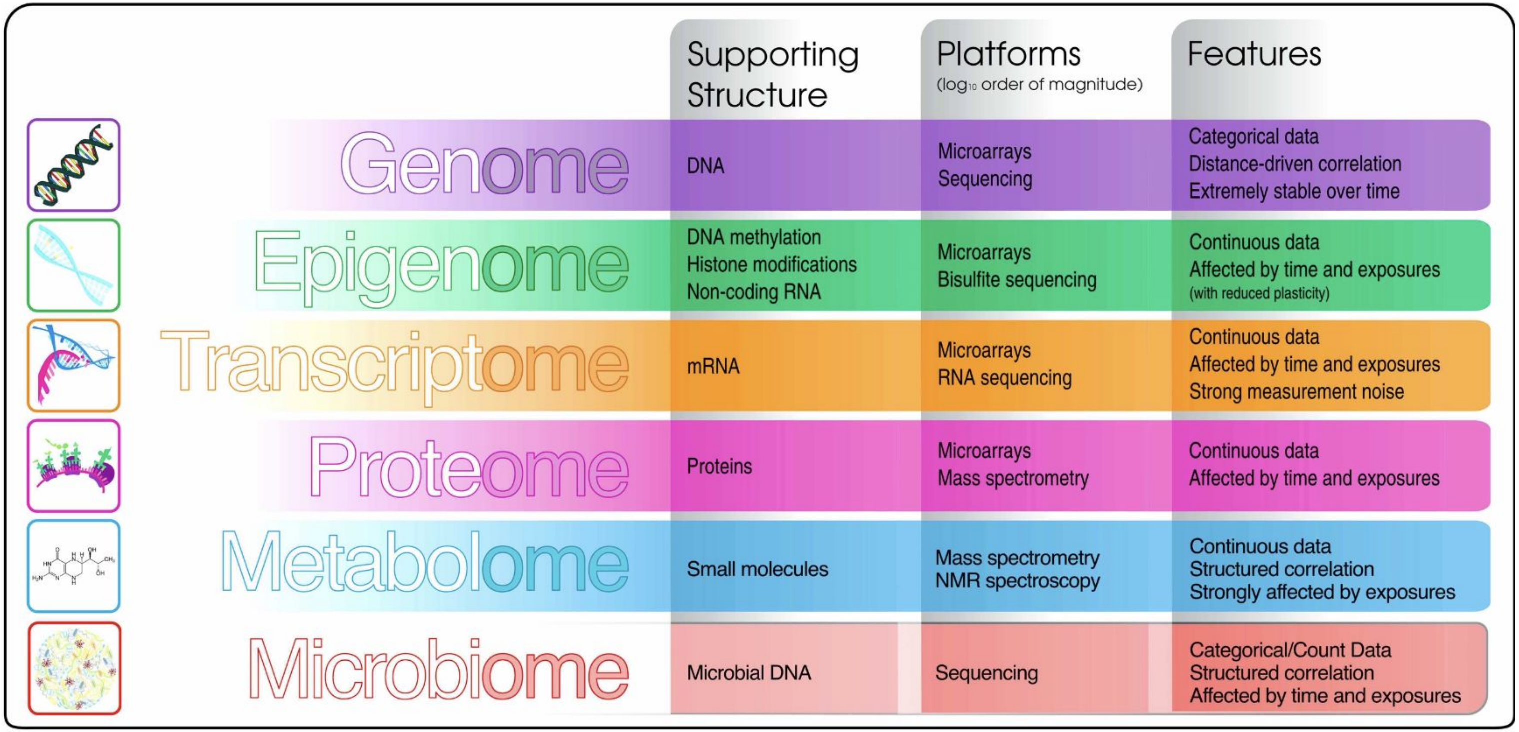
MIA
PARIS-SACLAY

Interpretable variable selection and differential structure estimation in graphical models

Blanche Francheterre, Marc Chadeau-Hyam, Julien Chiquet
28/01/2026

Colorectal - Renal - Pancreatic - Cancers

OMICs

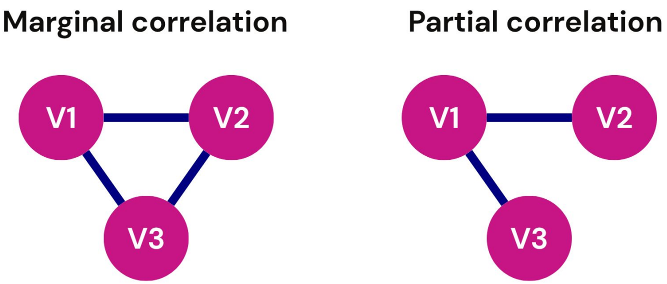


	Supporting Structure	Platforms (log10 order of magnitude)	Features
Genome	DNA	Microarrays Sequencing	Categorical data Distance-driven correlation Extremely stable over time
Epigenome	DNA methylation Histone modifications Non-coding RNA	Microarrays Bisulfite sequencing	Continuous data Affected by time and exposures (with reduced plasticity)
Transcriptome	mRNA	Microarrays RNA sequencing	Continuous data Affected by time and exposures Strong measurement noise
Proteome	Proteins	Microarrays Mass spectrometry	Continuous data Affected by time and exposures
Metabolome	Small molecules	Mass spectrometry NMR spectroscopy	Continuous data Structured correlation Strongly affected by exposures
Microbiome	Microbial DNA	Sequencing	Categorical/Count Data Structured correlation Affected by time and exposures

Motivation

Assumptions

- OMIC variables that are **involved in the same regulatory pathways** tend to be **correlated**
- **Gaussian graphical models** infer **direct associations** between variables by estimating **partial correlations** [Lauritzen, 1996]

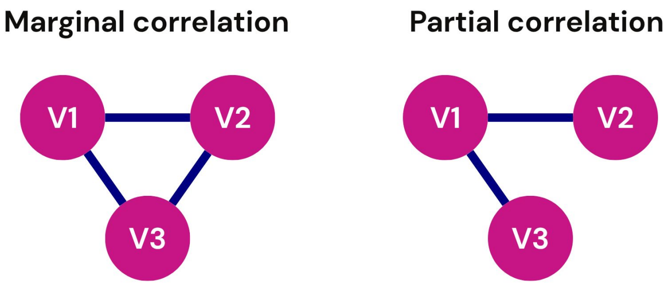


Conditionnally on V1, V2 and V3 are independent

Motivation

Assumptions

- OMIC variables that are **involved in the same regulatory pathways** tend to be **correlated**
- **Gaussian graphical models** infer **direct associations** between variables by estimating **partial correlations** [Lauritzen, 1996]



Conditionnally on V1, V2 and V3 are independent

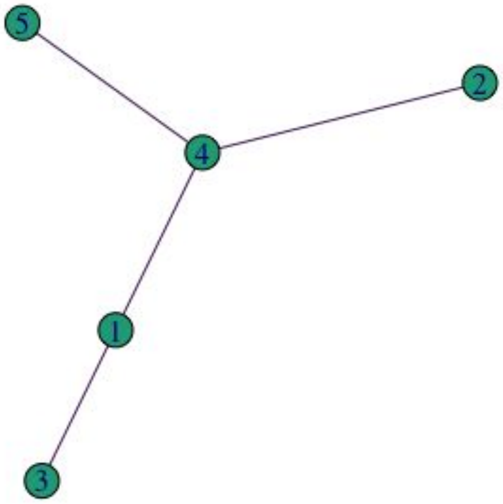
→ These **partial correlations** are encoded in **the precision matrix, Ω**

$$X_i \sim \mathcal{N}(\mu, \Omega^{-1})$$

Motivation

$$\Omega = \begin{pmatrix} 1.71 & 0 & 0.23 & 0.47 & 0 \\ 0 & 1.54 & 0 & -0.54 & 0 \\ 0.23 & 0 & 1.23 & 0 & 0 \\ 0.47 & -0.54 & 0 & 2.52 & 0.5 \\ 0 & 0 & 0 & 0.5 & 1.5 \end{pmatrix}$$

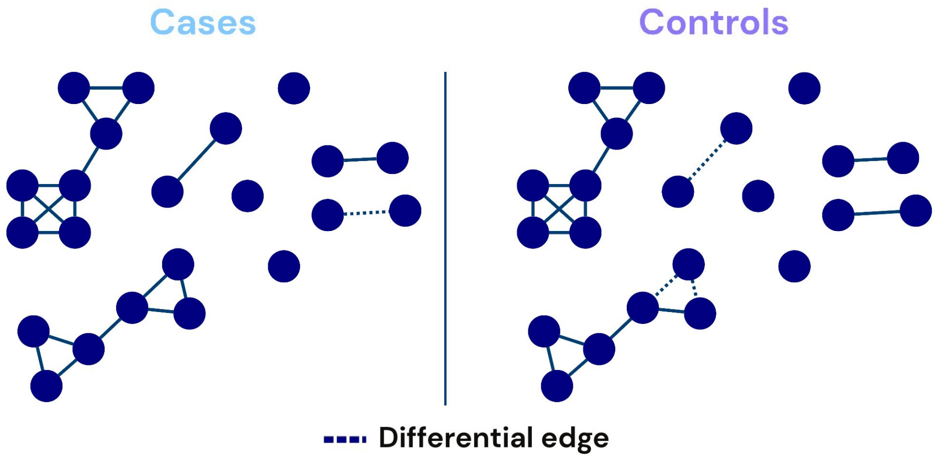
$$A = \begin{pmatrix} 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$



Hypothesis

Let $X \in \mathbb{R}^{n \times p}$ the matrix of observations and $Y \in \{1, 2, .., K\}$ the response variable.
Let $K = 2$, we assume that $X_1 \sim \mathcal{N}(\mu_1, \Sigma_1), X_2 \sim \mathcal{N}(\mu_2, \Sigma_2)$, where $X_1 = (\mathbf{x_i} : Y_i = 1)$ and $X_2 = (\mathbf{x_i} : Y_i = 2)$ such that:

$$\Sigma_1^{-1} = \Omega_1$$

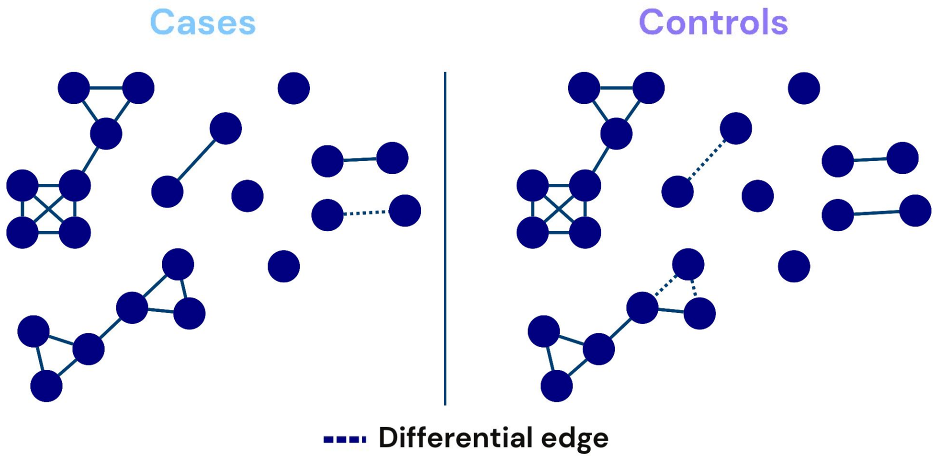


$$\Sigma_2^{-1} = \Omega_2$$

Hypothesis

Let $X \in \mathbb{R}^{n \times p}$ the matrix of observations and $Y \in \{1, 2, .., K\}$ the response variable.
Let $K = 2$, we assume that $X_1 \sim \mathcal{N}(\mu_1, \Sigma_1), X_2 \sim \mathcal{N}(\mu_2, \Sigma_2)$, where $X_1 = (\mathbf{x_i} : Y_i = 1)$ and $X_2 = (\mathbf{x_i} : Y_i = 2)$ such that:

$$\Sigma_1^{-1} = \Omega_1$$



$$\Sigma_2^{-1} = \Omega_2$$

Graphs have a **shared structure** between classes with **differential edges**

Joint network inference

Research questions

➔ Considering joint and independent network estimation approaches, **under what conditions** do they perform best at **recovering shared and condition-specific network structure**?

Research questions

➔ Considering joint and independent network estimation approaches, **under what conditions** do they perform best at **recovering shared and condition-specific network structure**?

Current limitations

- **Various methods** to infer gaussian graphical models and many joint penalties
- Existing studies typically restricted to a **few scenarios**
- Do not clearly distinguish between **model performance** (how well the network structure is recovered) and **model selection** (how penalties are chosen)

Research questions

→ Which **hyperparameter calibration** strategy yields the most reliable edge detection?

Research questions

➔ Which **hyperparameter calibration** strategy yields the most reliable edge detection?

Current limitations

- Do not clearly distinguish between **model performance** (how well the network structure is recovered) and **model selection** (how penalties are chosen)
- Rely on **likelihood-based criteria** (AIC/BIC), which favor dense graphs and reduce interpretability

Joint network inference methods

Method	Loss	Penalties
Graphical lasso (GL) [Yang et al, 2007, Danaher et al, 2014]	$\min_{\{\Omega\}_{k=1}^K} \sum_{k=1}^K n_k \left[\mathrm{tr}\left(\hat{\Sigma}^{(k)}\Omega^{(k)}\right) - \log\det(\Omega^{(k)})\right]$	• L1 penalty on edges • L1 penalty on edges + joint fused penalty • L1 penalty on edges + joint group penalty • L1 penalty on edges + node based penalty
Neighborhood selection (NS) [Meinshausen & Bühlmann, 2006]	$\min_{\{\beta\}_{k=1}^K} \sum_{k=1}^K \sum_{j=1}^p \frac{1}{2n_k} \left\Vert X_j^{(k)} - X_{-j}^{(k)} B_j^{(k)} \right\Vert_2^2$ $\beta_{ij}^{(k)} = 0 \Leftrightarrow \omega_{ij}^{(k)} = 0$	• L1 penalty on edges • L1 penalty on edges + joint fused penalty • L1 penalty on edges + joint group penalty • Data shared lasso
Partial correlation based neighborhood selection (PC) [Peng et al, 2009]	Let $\tilde{X}_i^{(k)} = \sqrt{\frac{\sigma_{ii}}{\sigma_{jj}}} X_i^{(k)}$ $\min_{\{\rho\}_{k=1}^K} \sum_{k=1}^K \sum_{j=1}^p \frac{1}{2n_k} \left\Vert X_j^{(k)} - \sum_{i \neq j} \rho_{ij}^{(k)} \tilde{X}_i^{(k)} \right\Vert_2^2$	• L1 penalty on edges • L1 penalty on edges + joint fused penalty • L1 penalty on edges + joint group penalty
Dtrace [Yuan et al, 2017]	Let $\Omega_{\mathrm{diff}} = \Omega_2 - \Omega_1$ $\min_{\Omega_{\mathrm{diff}}} \left[\frac{1}{2}\,\mathrm{Tr}\left(\Omega_{\mathrm{diff}}^{\mathsf{T}}\hat{\Sigma}_1\,\Omega_{\mathrm{diff}}\,\hat{\Sigma}_2\right) - \mathrm{Tr}\left(\Omega_{\mathrm{diff}}(\hat{\Sigma}_1 - \hat{\Sigma}_2)\right)\right]$	• L1 penalty on differential edges

Penalties

Individual edge penalty

$$\lambda_1 \sum_{k=1}^{K} \sum_{i \neq j} \left| \Omega_{ij}^{(k)} \right|$$

$$+$$

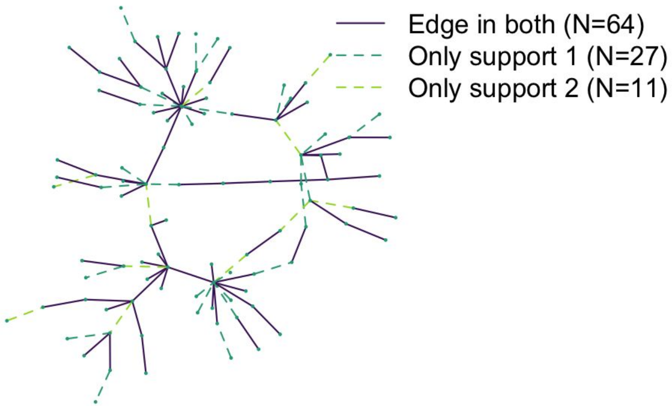
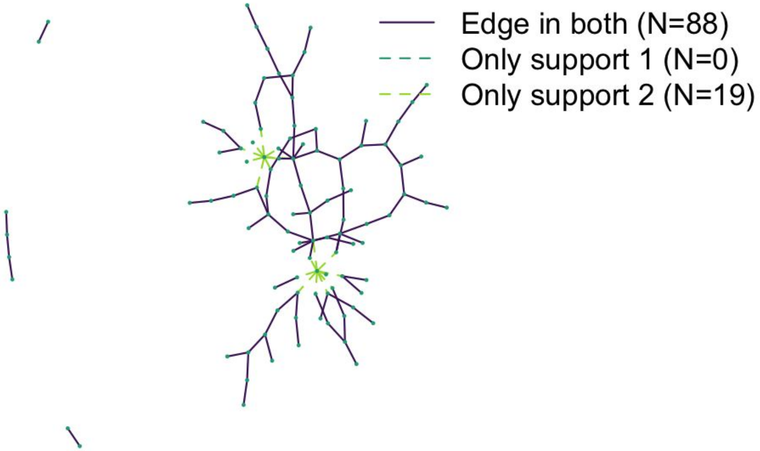
Joint fused penalty

$$\lambda_2 \sum_{k<k'} \sum_{i,j} \left| \Omega_{ij}^{(k)} - \Omega_{ij}^{(k')} \right|$$

Joint group penalty

$$\lambda_2 \sum_{i \neq j} \left(\sum_{k=1}^{K} {\Omega_{ij}^{(k)}}^2 \right)^{1/2}$$

Network generation

Number of nodes, p	**30, 100, 300, 500**
Topology	**Scale-free**  **Random graph with hubs** 
Differential edge	➔ **Random** (a percentage of edges turned off) ➔ **Hub-based**

Data generation

- ➔ Generate **2 positive definite matrices** for the corresponding graphs
- ➔ **Σ_1, Σ_2**: computed as the **inverse of precision matrices**
- ➔ We simulate data from a **multivariate Normal distribution** $N(0,\Sigma_k)$, where $k \in \{1, 2\}$
- ➔ Sample sizes considered: **N = 50, 100** per condition
- ➔ **50 independent replicates** of datasets X_1 and X_2

Performance evaluation

Evaluation of differential support edges
Compute over λ_2 :

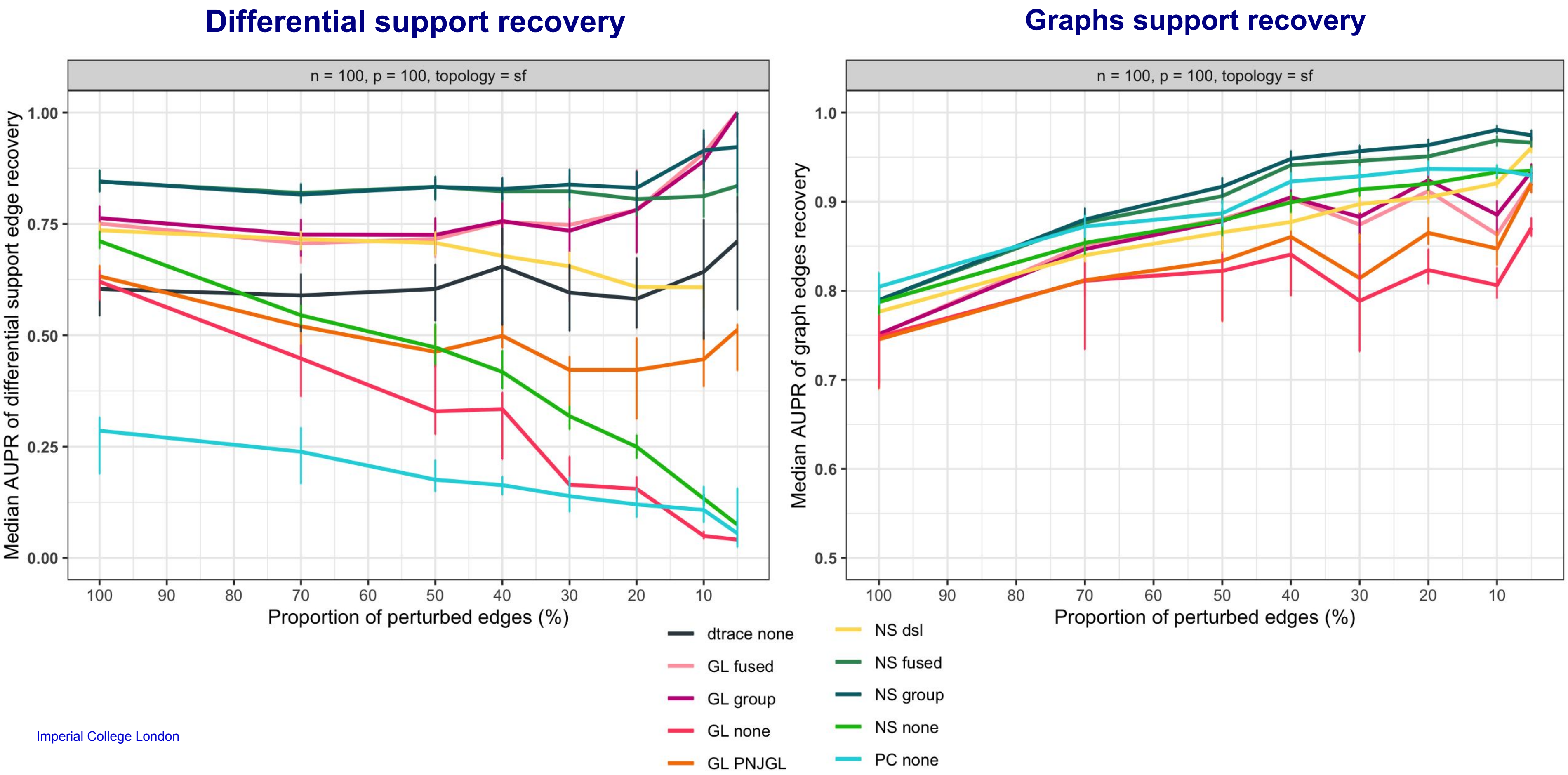
- Area under the precision-recall curve (AUPR)
- Area under recall vs. false-positive rate (AUC)
- Area under power vs. FDR (AUPF)

Evaluation of overall graph inference

- Apply the same metrics to assess graph recovery over λ_1

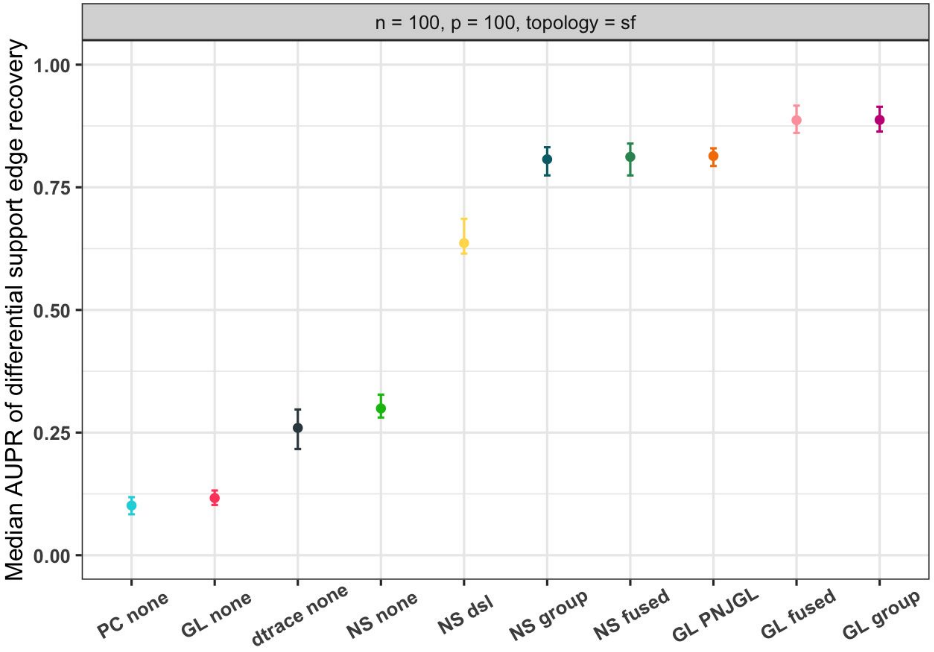
For each method, record the **maximum F1-score** in terms of edge recovery across λ_1 , λ_2 parameters and average it over the 50 simulated datasets

Results: random edge perturbations

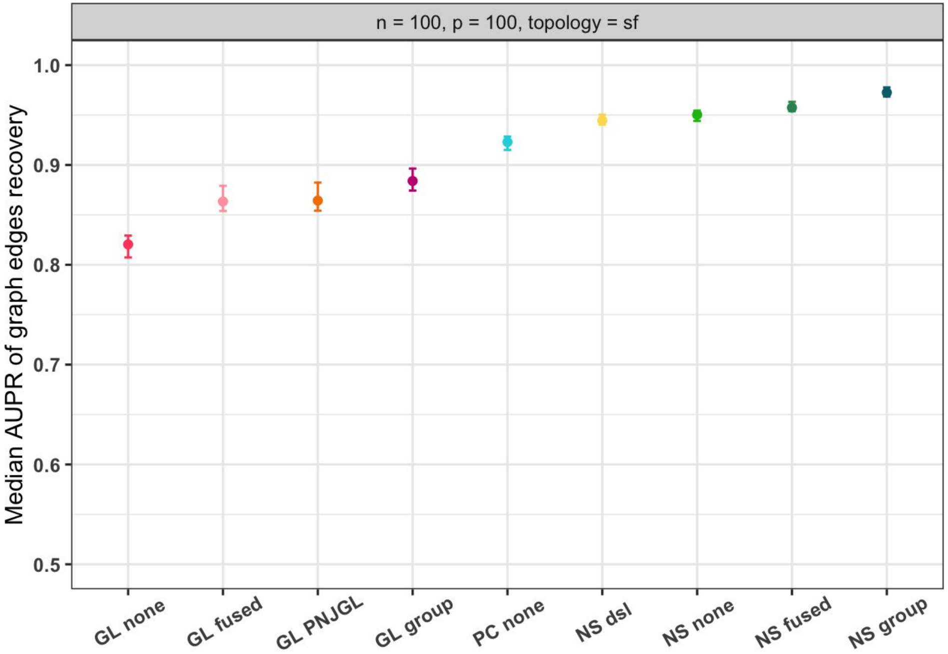


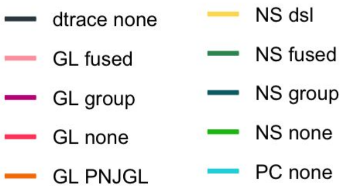
Results: hub based perturbations

Differential support recovery



Graphs support recovery

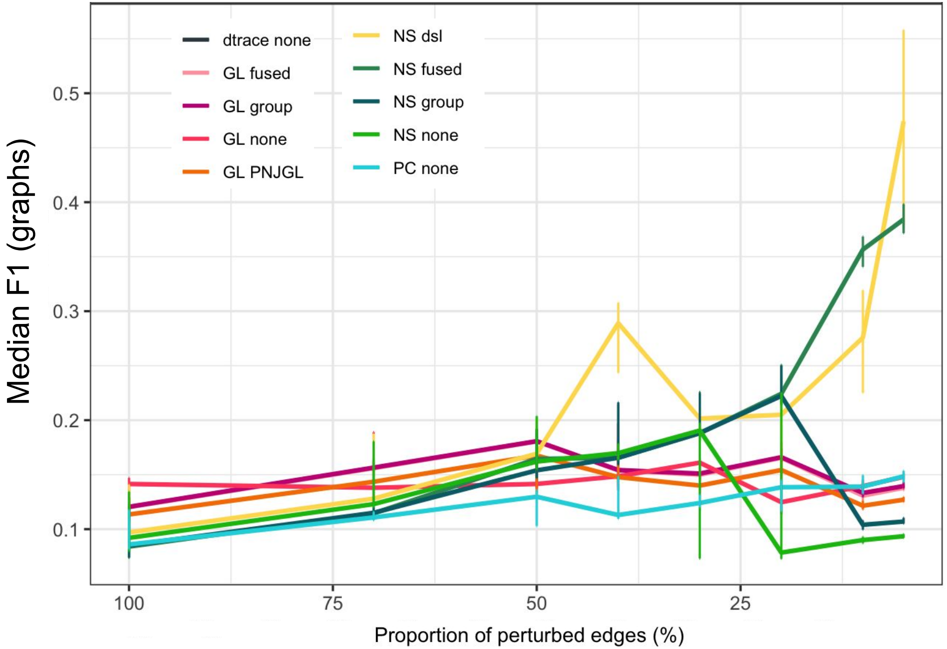




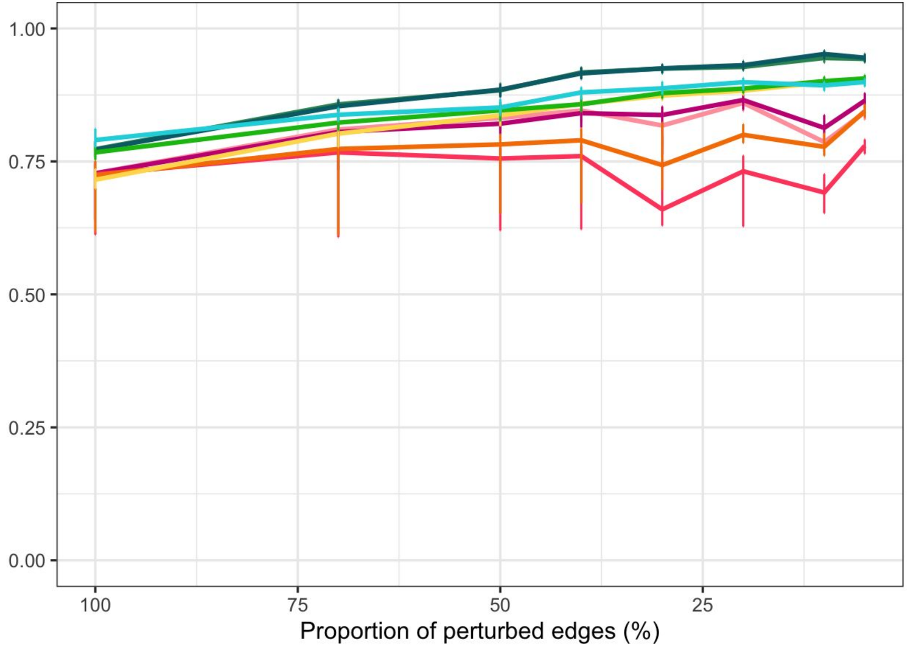
Model calibration

Often relies on **likelihood-based criteria (AIC/BIC)**, which favor **dense graphs** and reduce interpretability

Calibration with AIC



Oracle

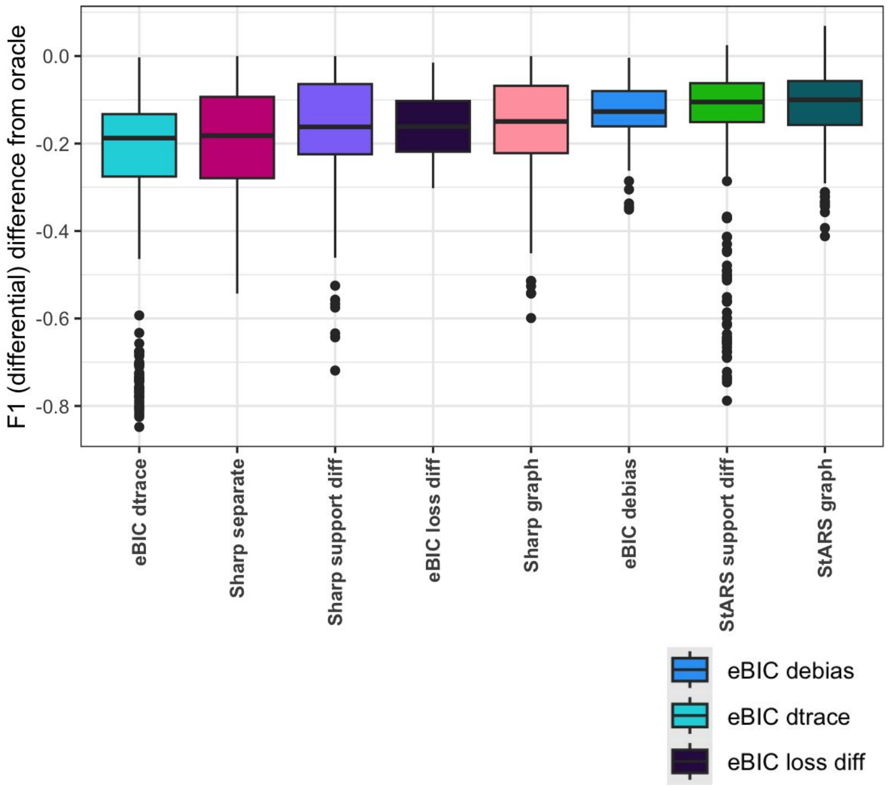


Model calibration

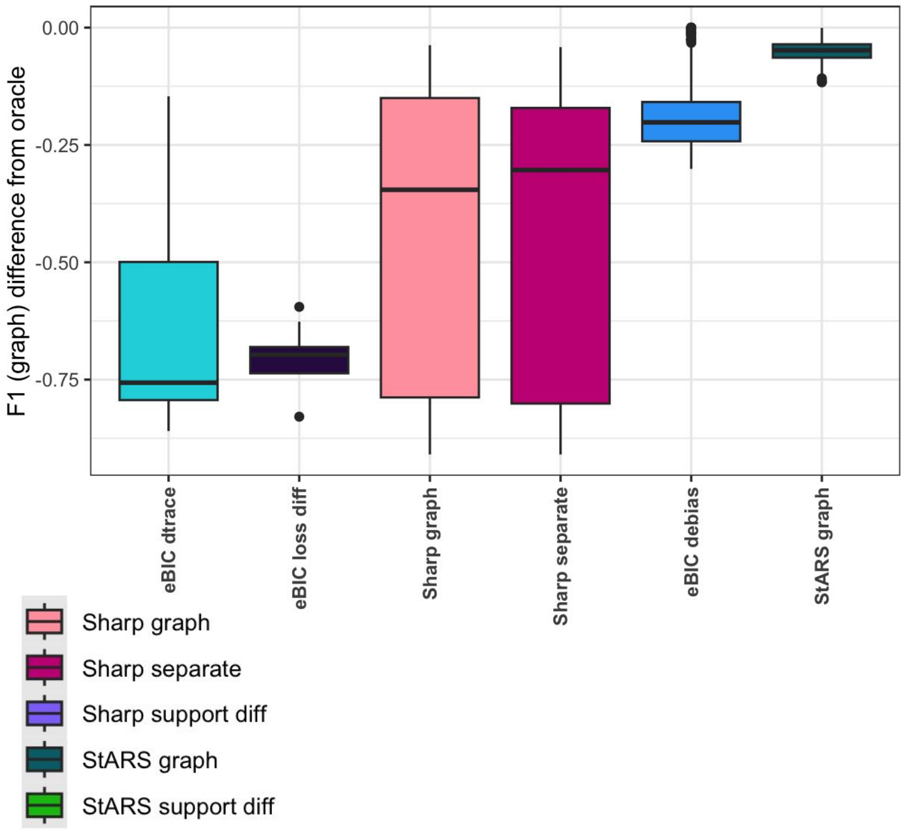
1. **Obtain debiased precision matrices**
2. **Calibrate**
   - $$\mathbf{eBIC}(\lambda_1, \lambda_2) = \sum_{k=1}^{K} n_k \left(-\log\det\left(\hat{\Omega}^{(k)}\right) + \mathrm{tr}\left(\hat{\Sigma}^{(k)}\hat{\Omega}^{(k)}\right)\right) + \log(n_k)\frac{||\hat{\Omega}^{(k)}||_0}{2} + 4\gamma\log(p)\frac{||\hat{\Omega}^{(k)}||_0}{2}$$
   - $$\mathbf{Dtrace\ pseudo\ eBIC}(\lambda_1) = \left[\frac{1}{2}\,\mathrm{Tr}\left(\hat{\Omega}_{\text{diff}}^{\mathsf{T}}\hat{\Sigma}_1\,\hat{\Omega}_{\text{diff}}\,\hat{\Sigma}_2\right) - \mathrm{Tr}\left(\hat{\Omega}_{\text{diff}}(\hat{\Sigma}_1 - \hat{\Sigma}_2)\right)\right]$$
     $$+\log(n_1 + n_2)||\hat{\Omega}_{\text{diff}}||_0 + 4\gamma\log(p)||\hat{\Omega}_{\text{diff}}||_0$$
   - **Stability selection** (StARS [Liu et al, 2010], Sharp [Bodinier et al, 2023])

Model calibration

Differential support recovery



Graphs support recovery



***neighborhood selection** with **sparse group penalty**

Weights

→ **Improving edge selection** and incorporating **prior knowledge**

	Individual edges	Joint fused	Joint group
Penalty	$\lambda_1 \sum_{k=1}^{K} \sum_{i,j} \boldsymbol{w_{ij}^{(k)}} \left\lvert \beta_{ij}^{(k)} \right\rvert$	$\lambda_2 \sum_{k<k'} \sum_{i,j} \boldsymbol{w_{ij}^{(kk')}} \left\lvert \beta_{ij}^{(k)} - \beta_{ij}^{(k')} \right\rvert$	$\lambda_2 \sum_{i,j} \boldsymbol{w_{ij}} \left(\sum_{k=1}^{K} {\beta_{ij}^{(k)}}^2 \right)^{1/2}$
Adaptive weight	$w_{ij}^{(k)} = \frac{1}{\left\lvert \beta_{ij}^{(k)} \right\rvert}$	$w_{ij}^{(kk')} = \frac{1}{\left\lvert \beta_{ij}^{(k)} - \beta_{ij}^{(k')} \right\rvert}$	$w_{ij} = \frac{1}{\left(\sum_{k=1}^{K} {\beta_{ij}^{(k)}}^2 \right)^{1/2}}$
Degree-based weight	$w_{ij}^{(k)} = \frac{1}{\sqrt{(d_i^{(k)}+1)\cdot(d_j^{(k)}+1)}}$	$w_{ij}^{(kk')} = \frac{1}{\left\lvert \sqrt{(d_i^{(k)}+1)\cdot(d_j^{(k)}+1)} - \sqrt{(d_i^{(k')}+1)\cdot(d_j^{(k')}+1)} \right\rvert + \epsilon}$	$w_{ij} = \frac{w_{ij}^{(k)} + w_{ij}^{(k')}}{2}$

Weights

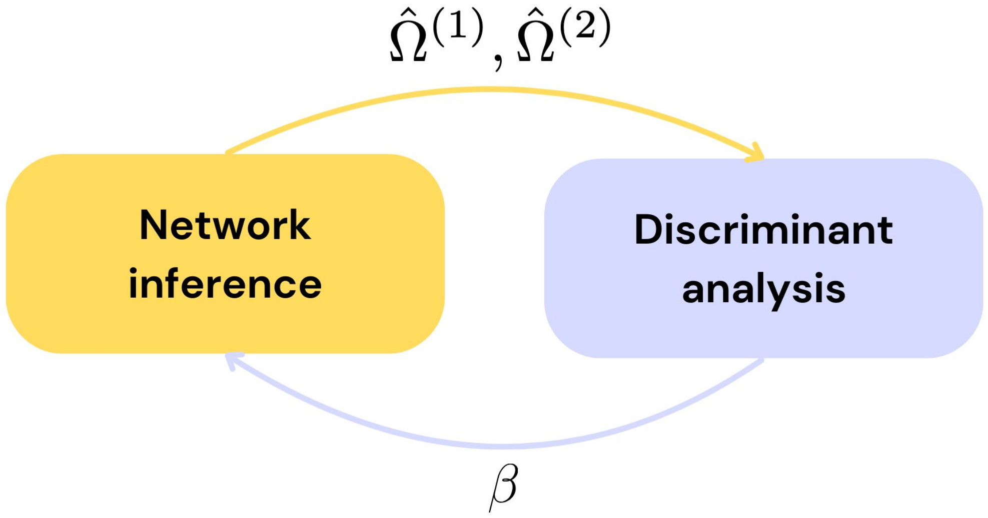
➔ **Improving edge selection** and incorporating **prior knowledge**

- Weights obtained from **Stochastic Block Model** clustering on the adjacency matrix [Ambroise et al, 2009]
- Weights obtained from **Multipartite Block Model** to account for multi-omics [unpublished Chiquet, 2018]

Next steps

➔ Combine joint network estimation and outcome discrimination to:

- improve **interpretability** and **discrimination**
- Identify both **structural** and **predictive differences** across classes



Thank you

Blanche Francheterre, Marc Chadeau-Hyam, Julien Chiquet
28/01/2026