

---

# Network modelling for high-dimensional data

*Instructors:*

Carel F.W. Peeters & Wessel N. van Wieringen  
[cf.peeters@vumc.nl](mailto:cf.peeters@vumc.nl) & [w.vanwieringen@vumc.nl](mailto:w.vanwieringen@vumc.nl)

Dept. Epidemiology and Biostatistics, VUmc  
& Dept. Mathematics, Vrije Universiteit Amsterdam  
Amsterdam, The Netherlands



VU medisch centrum



# Course

---

<i>Outline</i>	<i>210 minutes</i>
Introduction	30 min
Single network	60 min
→ Presentation?	→ 30
→ Practical?	→ 30
-- break --	15 min
Multiple groups	45 min
→ Presentation?	→ 20
→ Practical?	→ 25
-- break --	15 min
Time-course	45 min
→ Presentation	→ 20 min
→ Practical	→ 25 min

# What?

---

Molecular biology aims to understand the molecular processes that occur in the cell.

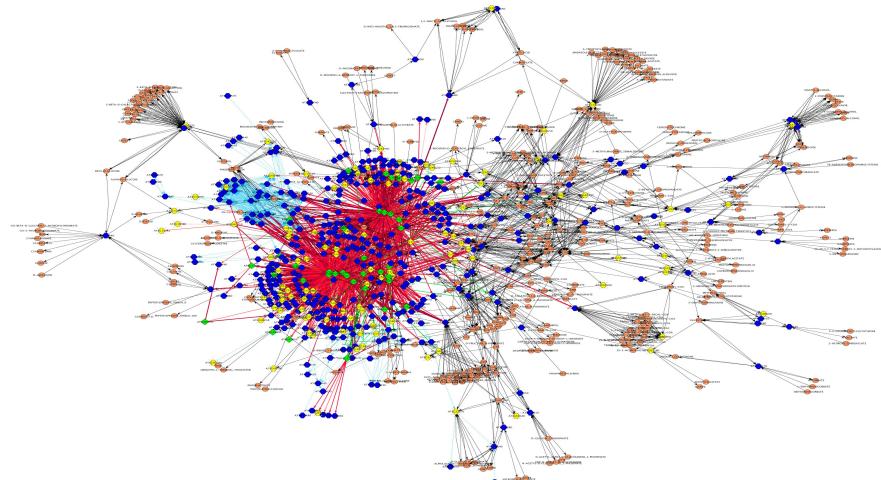
That is, e.g.:

- which molecules present in the cell interact?
- how is this coordinated?

For many cellular processes, it is unknown which genes play what role.

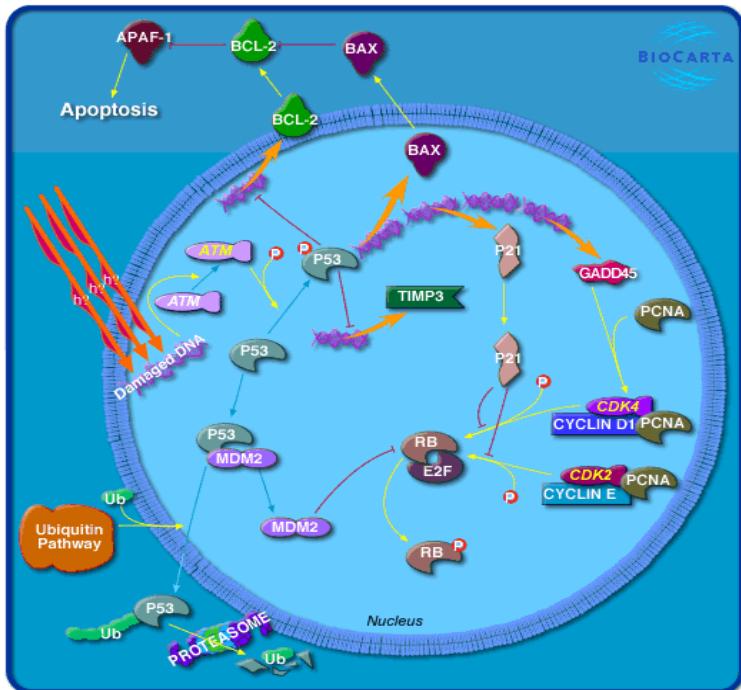
## Goal

Reconstruct the  
cellular regulatory  
network.



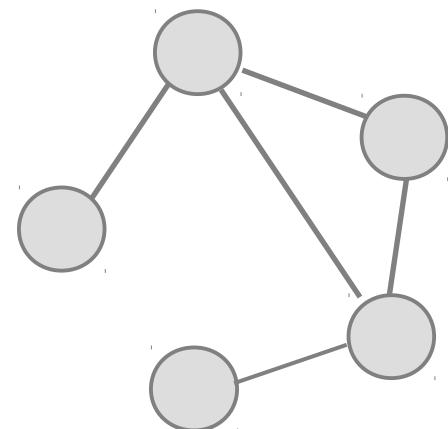
# Network

Pathway is represented by a *network*:  $\mathcal{G} = (\mathcal{V}, \mathcal{E})$

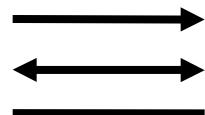


nodes & edges

formalization



node  $\approx$  a molecule.



edge: interaction  
between molecules.

# Conditional independence graph

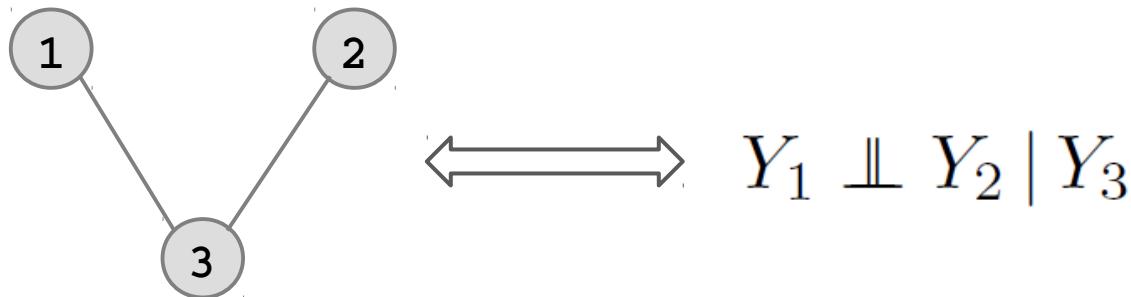
---

Let  $\mathbf{Y}$  be a  $p$ -dimensional random variable and  $\mathcal{V} = \{1, \dots, p\}$  the corresponding set of nodes.

The *conditional independence graph* of  $\mathbf{Y}$  is an undirected graph  $\mathcal{G} = (\mathcal{V}, \mathcal{E})$  s.t.:

$$(j_1, j_2) \notin \mathcal{E} \iff Y_{j_1} \perp\!\!\!\perp Y_{j_2} \mid Y_{\mathcal{V} \setminus \{j_1, j_2\}}$$

*Example*



# With?

---

## Data

Available to reconstruct which molecular interact:

- molecular profiles of  $n$  samples,
- each profile comprises  $p$  features.

	molec. 1	molec. 2	molec. 3	molec. 4	molec. 5
sample 1	-0.21968	-0.42796	0.26441	-5.74971	-0.96908
sample 2	-0.08376	-7.21648	-3.86460	0.77440	-3.18557
sample 3	-1.08336	-1.14688	-1.22544	-2.36134	0.19293
sample 4	0.04333	-0.46377	0.12756	-0.39535	-0.20215
sample 5	1.16542	0.86248	1.16049	1.23941	0.51927
sample 6	-0.29687	0.28602	-0.69624	-1.19779	0.19546
sample 7	1.76249	1.07556	1.46201	1.16076	1.29921
sample 8	0.46387	0.21271	0.49455	0.58267	-0.44349
sample 9	-1.27492	3.95515	-0.26441	-2.95037	-0.77896
...	...	...	...	...	...

≈ activity

## Repository

Prior knowledge on network

## Gaussian graphical model

$$\mathbf{Y} \sim \mathcal{N}(\mu, \Sigma)$$

# Gaussian graphical model

Let  $\mathbf{Y} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$  be partitioned as:

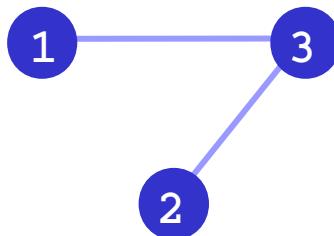
$$\begin{pmatrix} \mathbf{Y}_a \\ \mathbf{Y}_b \\ \mathbf{Y}_c \end{pmatrix} \sim \mathcal{N} \left( \begin{pmatrix} \boldsymbol{\mu}_a \\ \boldsymbol{\mu}_b \\ \boldsymbol{\mu}_c \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma}_{aa} & \boldsymbol{\Sigma}_{ab} & \boldsymbol{\Sigma}_{ac} \\ \boldsymbol{\Sigma}_{ba} & \boldsymbol{\Sigma}_{bb} & \boldsymbol{\Sigma}_{bc} \\ \boldsymbol{\Sigma}_{ca} & \boldsymbol{\Sigma}_{cb} & \boldsymbol{\Sigma}_{cc} \end{pmatrix} \right)$$

Then:

$$\mathbf{Y}_a \perp\!\!\!\perp \mathbf{Y}_b | \mathbf{Y}_c \iff \boldsymbol{\Omega}_{ab} = (\boldsymbol{\Sigma}^{-1})_{ab} = \mathbf{0}.$$

Simple criterion for (conditional) pairwise independence:

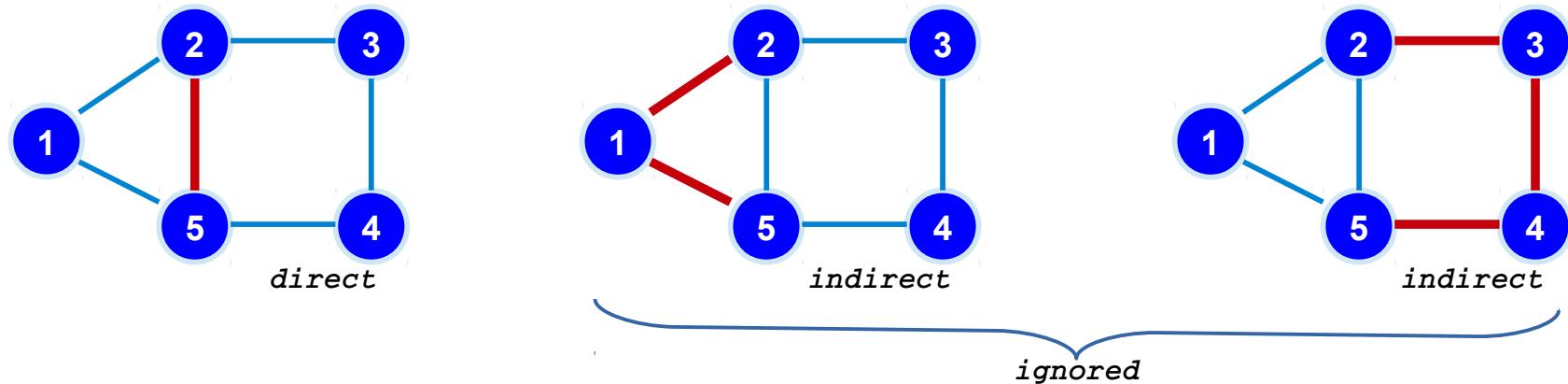
$$(\boldsymbol{\Omega})_{1,2} = 0 \iff (\boldsymbol{\Sigma}^{-1})_{1,2} = 0 \iff Y_1 \perp\!\!\!\perp Y_2 | Y_3, \dots, Y_p \iff$$



# Partial correlation

Elements of  $P$  measure the *direct* relation between two nodes while excluding effects of others.

$\omega_{25}$  : direct association between nodes 2 and 5:



Standardization yields the *partial correlations*, e.g.:

$$\rho(Y_a, Y_b | \mathbf{Y}_c) = \frac{\text{Cov}(Y_a, Y_b | \mathbf{Y}_c)}{\sqrt{\text{Var}(Y_a | \mathbf{Y}_c)} \sqrt{\text{Var}(Y_b | \mathbf{Y}_c)}}$$

= linear dependence between  $y_a$  and  $y_b$  conditional on  $\mathbf{y}_c$ .

# Network reconstruction

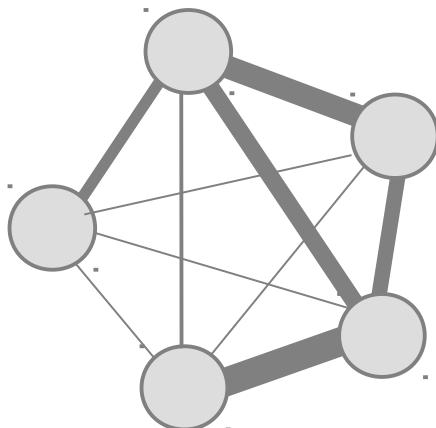
*Low-dimensionally*

The log-likelihood:

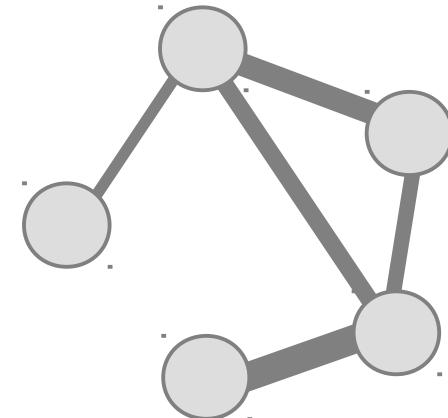
$$\mathcal{L}(\mathbf{S}, \Omega) \propto \log(|\Omega|) - \text{tr}(\mathbf{S}\Omega)$$

with  $\mathbf{S}$  the sample covariance matrix,  
whose inverse estimates the precision matrix  $\Omega$ .

Infer CIG:



$$\frac{\text{edge if reject}}{H_0 : (\Omega)_{1,2} = 0} \rightarrow$$



$$Y_1 \perp Y_2 | Y_3, \dots, Y_p$$

# “Ad hoc” ridge

---

## *High-dimensionally*

The sample covariance matrix is singular.

Consequence, its inverse not defined.

## *Solution*

Patch the singularity and consider a shrunken estimate:

$$\widehat{\Sigma}(\lambda) = (1 - \lambda)\mathbf{S} + \lambda\mathbf{T}$$

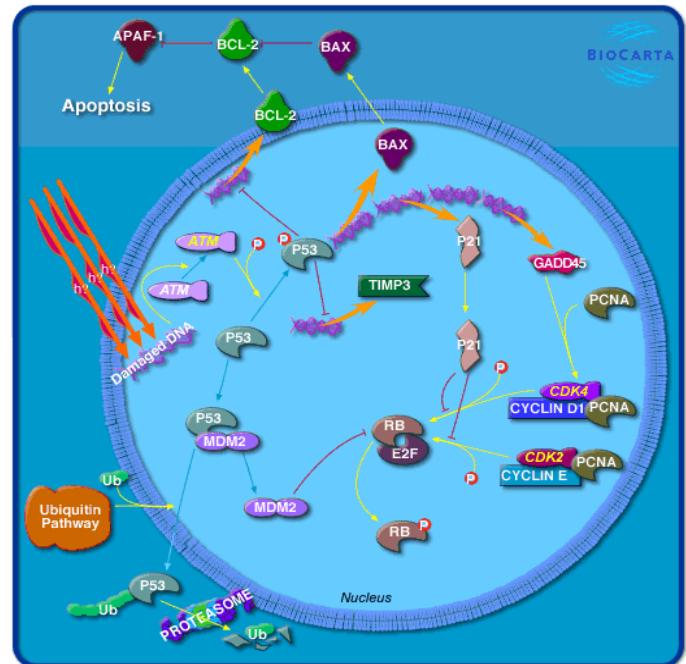
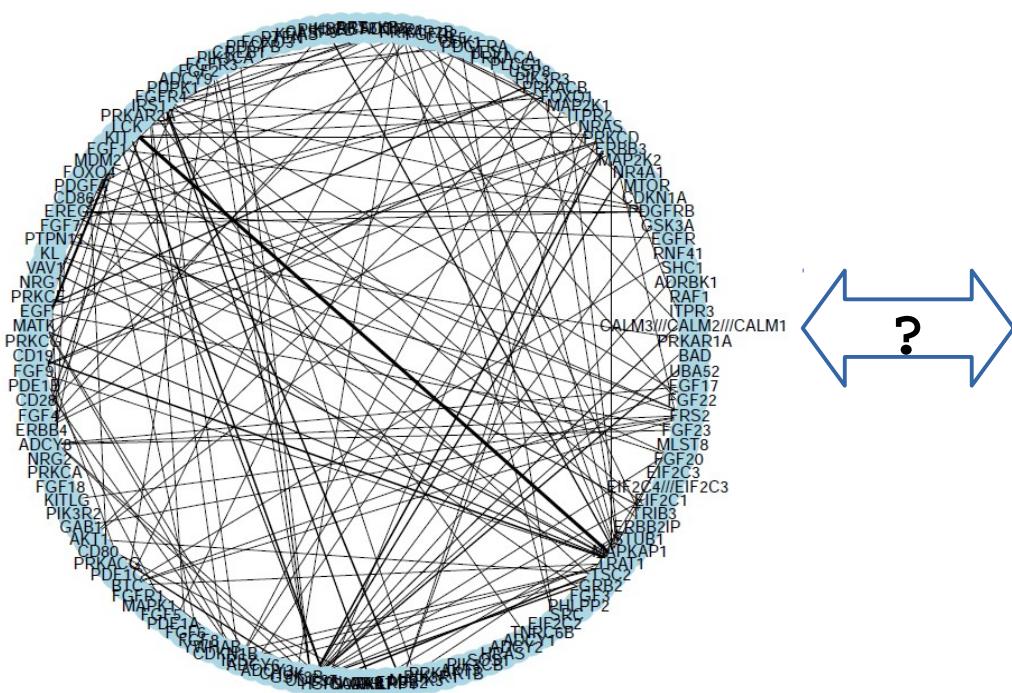
where

- $\mathbf{S}$  : the sample covariance matrix,
- $\lambda \in [0, 1]$  : the *shrinkage parameter*, and
- $\mathbf{T}$  : the *shrinkage target*, a  $p \times p$  dimensional,  
non-random, positive definite, symmetric matrix.

# "Ad hoc" ridge

Why a target?

Poor resemblance inferred network and repository:

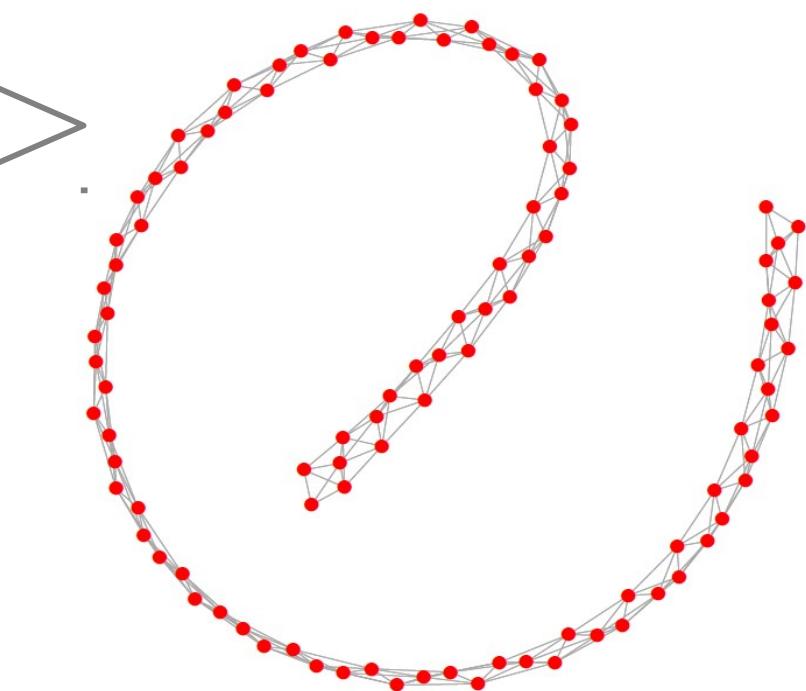
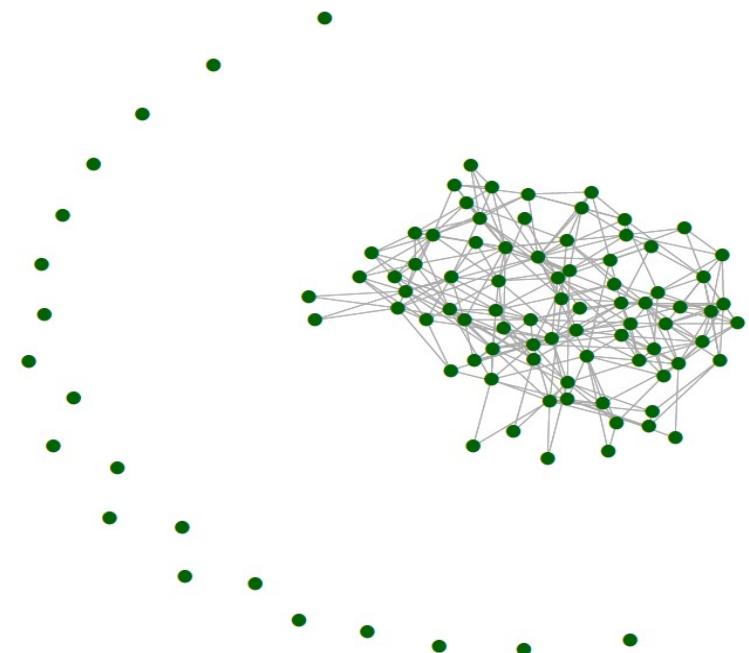
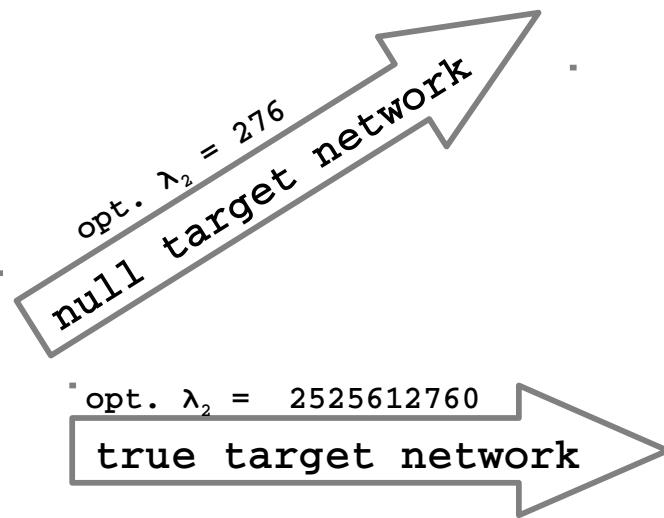
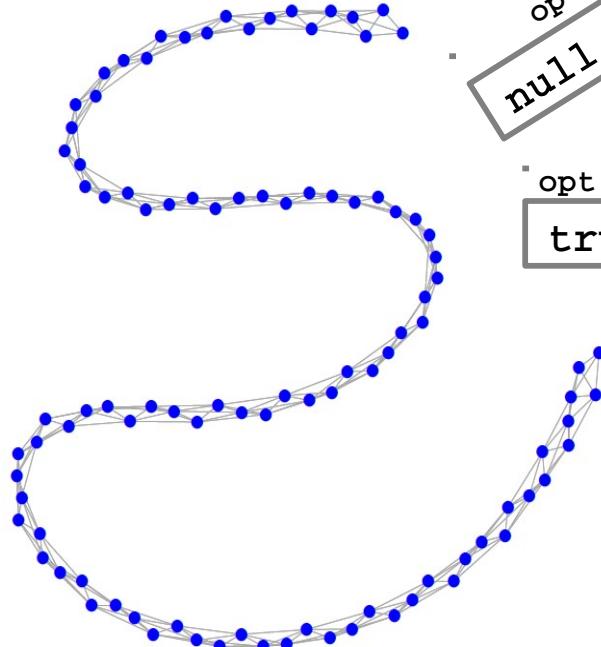


Provide a hint from pilot data.

# "Ad hoc" ridge

*Effect of the target*

True network



# Ridge

---

Instead of an ad-hoc fix,  
look for a ridge estimate which minimizes:

$$\frac{\mathcal{L}(\mathbf{S}, \boldsymbol{\Sigma}^{-1}) - \frac{1}{2} \lambda_2 \|\boldsymbol{\Sigma}^{-1}\|_2^2}{\text{loss function} \quad \text{ridge penalty}}$$

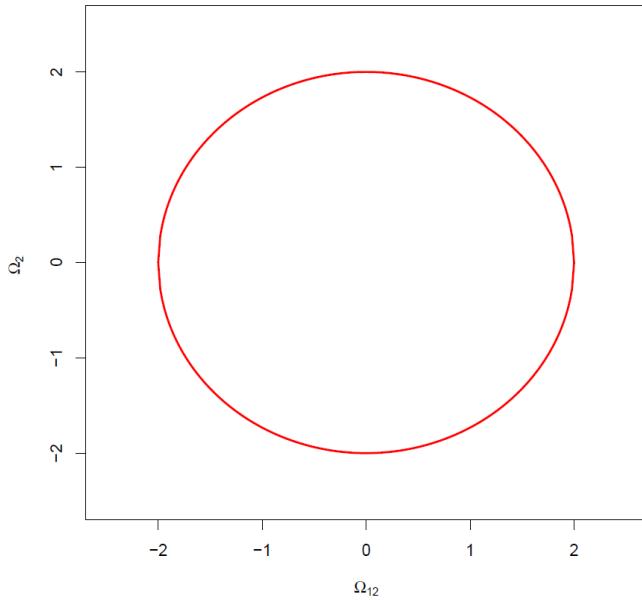

When writing  $\boldsymbol{\Omega} = \boldsymbol{\Sigma}^{-1}$  the ridge penalty is:

$$\|\boldsymbol{\Omega}\|_2^2 = \sum_{j_1, j_2=1}^p [(\boldsymbol{\Omega})_{j_1, j_2}]^2$$

# Ridge

Penalized estimation as constrained estimation .

*Parameter constraint*

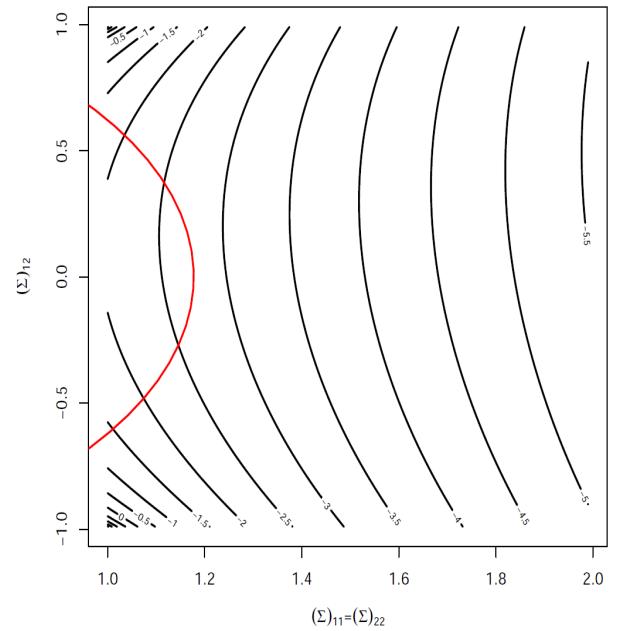


For a  $2 \times 2$  precision matrix the constraint is:

$$[\Omega_{11}]^2 + 2[\Omega_{12}]^2 + [\Omega_{22}]^2 \leq c(\lambda_2)$$

Plot: equal diag. elements.

*Constrained estimation*



Contourplot of the log-likelihood with the ridge parameter constraint.

# Ridge

---

With the inclusion of a target the ridge penalized log-likelihood is:

$$\mathcal{L}(\mathbf{S}, \boldsymbol{\Omega}) - \frac{1}{2}\lambda_2 \text{tr}[(\boldsymbol{\Omega} - \mathbf{T})(\boldsymbol{\Omega} - \mathbf{T})^T]$$

where  $\mathbf{T}$  is a non-random target matrix.

Ridge ML estimator of the precision:

$$\left\{ \frac{1}{2}(\mathbf{S} - \lambda_2 \mathbf{T}) + [\lambda_2 \mathbf{I}_{pp} + \frac{1}{4}(\mathbf{S} - \lambda_2 \mathbf{T})^2]^{1/2} \right\}^{-1}$$

Eigenvalue shrinkage (with zero target):

$$d_j[\widehat{\boldsymbol{\Sigma}}(\lambda_2)] = \frac{1}{2}d_j(\mathbf{S}) + \{\lambda_2 + \frac{1}{4}[d_j(\mathbf{S})]^2\}^{1/2}$$

# Ridge

*Use of the target*

Estimator:

$$\left\{ \frac{1}{2}(\mathbf{S} - \lambda_2 \mathbf{T}) + [\lambda_2 \mathbf{I}_{p \times p} + \frac{1}{4}(\mathbf{S} - \lambda_2 \mathbf{T})^2]^{1/2} \right\}^{-1}$$

$$\widehat{\Sigma}_{\text{pilot}} \xrightarrow{\quad} \mathbf{T} = c \widehat{\Sigma}_{\text{pilot}}^{-1}$$



Pilot data + topology



- Estimated from data
- ≈ “empirical Bayes”;
- Tuning parameter.

# Ridge

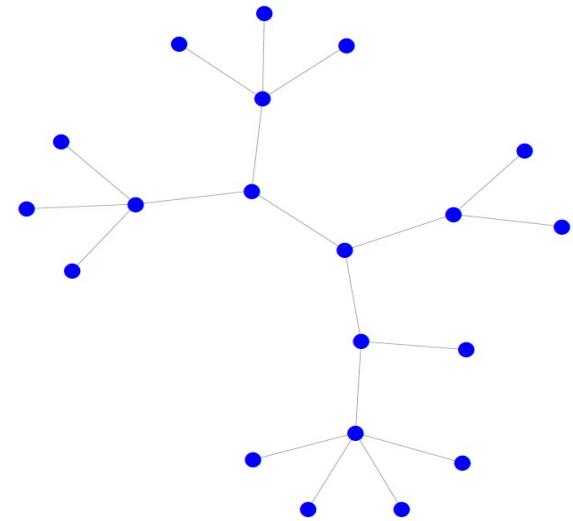
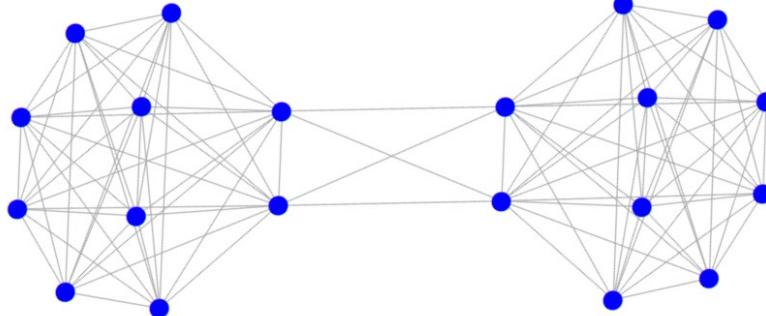
---

*Known absence of edges*

Ridge does not yield zeros. But ...  
... known zeros can be incorporated:

$$\max_{\substack{(\Omega)_{j_1,j_2}=0 \\ \text{for } (j_1,j_2) \notin \mathcal{E}}} \log(|\Omega|) - \text{tr}(\mathbf{S}\Omega) - \lambda_2 f^{\text{pen}}(\Omega, \mathbf{T})$$

Only for *decomposable graphs*,  
although not strictly  
necessary.



# Ridge

---

## *Properties*

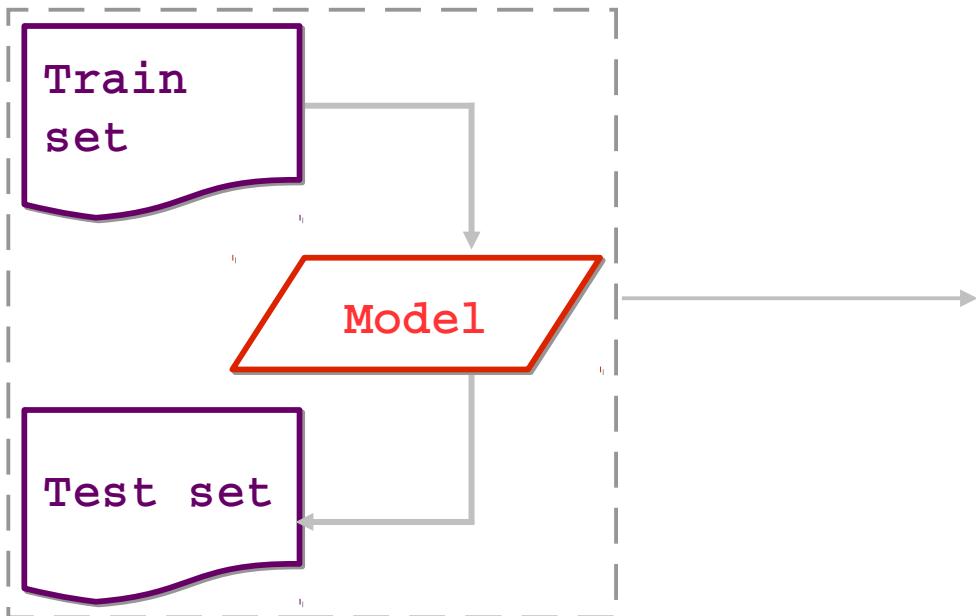
The ridge ML precision estimator is / satisfies:

- $\widehat{\Omega}(\lambda_2) \succ 0$  for all  $\lambda_2 > 0$
- $\lim_{\lambda_2 \downarrow 0} \widehat{\Omega}(\lambda_2) = \mathbf{S}^{-1}$
- $\lim_{\lambda_2 \rightarrow \infty} \widehat{\Omega}(\lambda_2) = \mathbf{T}$
- **asymptotically unbiased,**
- **consistent,**
- **moment approximations,**
- **Bayesian interpretation,**
- **smaller MSE for suitable  $\lambda_2 > 0$ .**

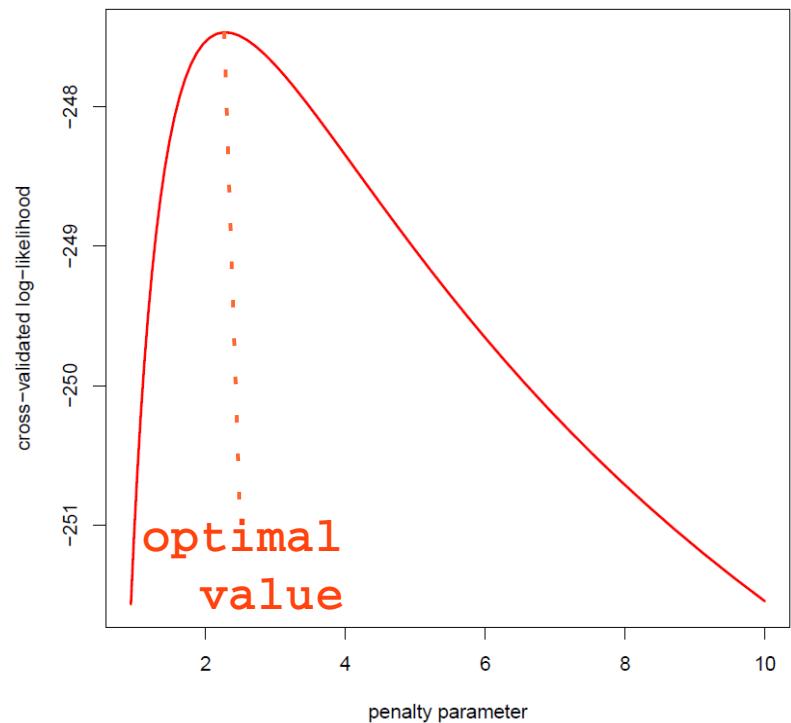
# Ridge

## Penalty selection

### Cross-validation



### Performance evaluation



- **K-fold**
- approximate LOOCV

# Ridge

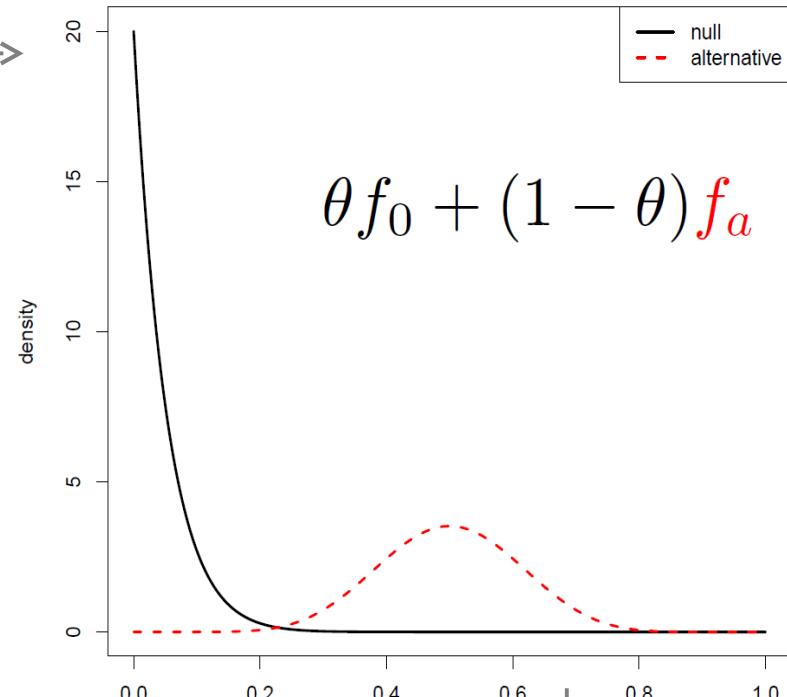
## Support determination

standardized  $\Omega$  estimate

	1	2	3	4	5
1	1.00	0.51	0.42	-0.03	0.34
2	*	1.00	0.40	0.11	-0.09
3	*	*	1.00	-0.57	0.04
4	*	*	*	1.00	0.07
5	*	*	*	*	1.00



model as mixture



inferred  $\Omega$  support

	1	2	3	4	5
1	1	1	1	0	1
2	*	1	1	0	0
3	*	*	1	1	0
4	*	*	*	1	0
5	*	*	*	*	1

← cut-off

$$P[(1, 2) \notin \mathcal{E}_G \mid (\widehat{\Omega})_{1,2}]$$



# Lasso

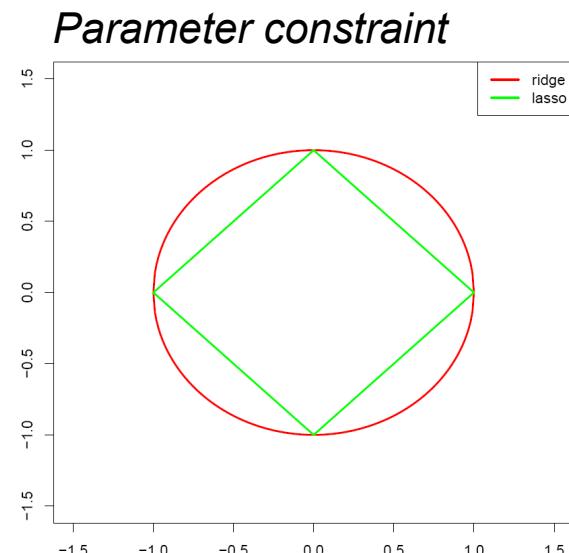
The *graphical lasso*, is a proper lasso estimator of the inverse covariance matrix, in the sense that it minimizes:

$$\mathcal{L}(\mathbf{S}; \Sigma^{-1}) + \lambda \|\Sigma^{-1}\|_1$$

loss                              penalty

When writing  $\Omega = \Sigma^{-1}$   
the lasso penalty is:

$$\|\Omega\|_1 = \sum_{j_1, j_2=1}^p |(\Omega)_{j_1, j_2}|$$



For a 2x2 precision matrix the constraint is:

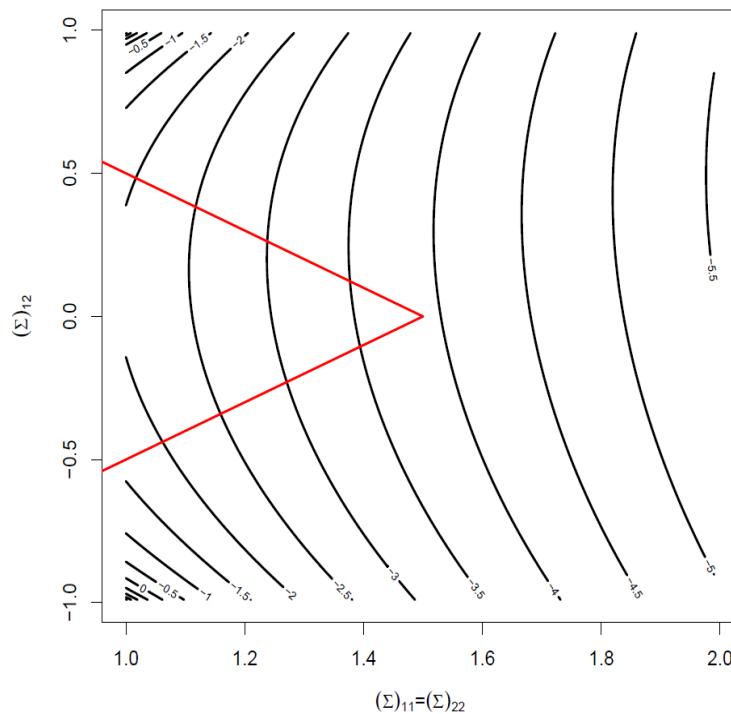
$$|\Omega_{11}| + 2|\Omega_{12}| + |\Omega_{22}| \leq c(\lambda_1)$$

For plot: equal diagonal elements.

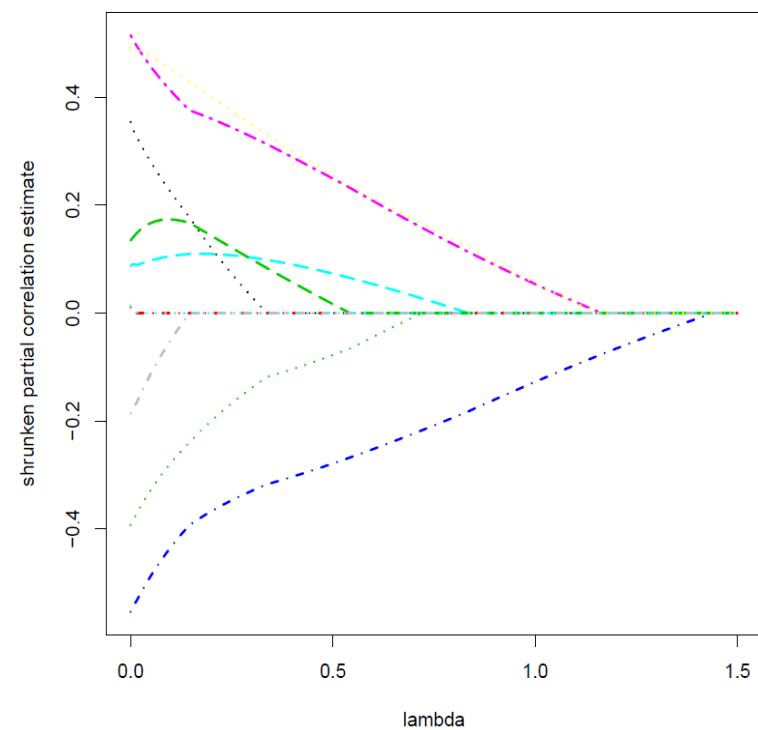
# Lasso

As the lasso, the graphical lasso selects!

Contourplot of the loglikelihood with the lasso parameter constraint.



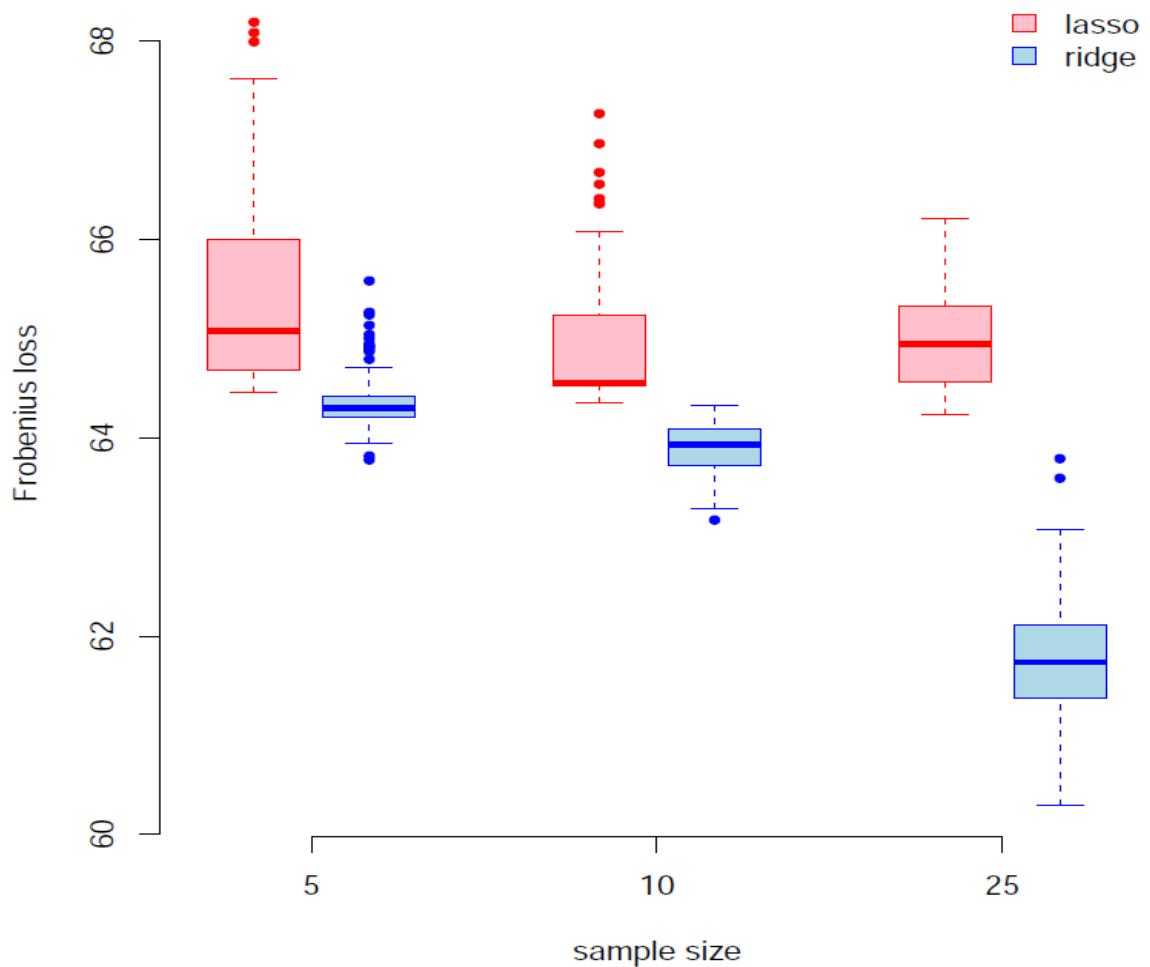
Effect of  $\lambda$  on the partial correlation lasso estimates.



# Ridge vs. lasso

## *Frobenius loss*

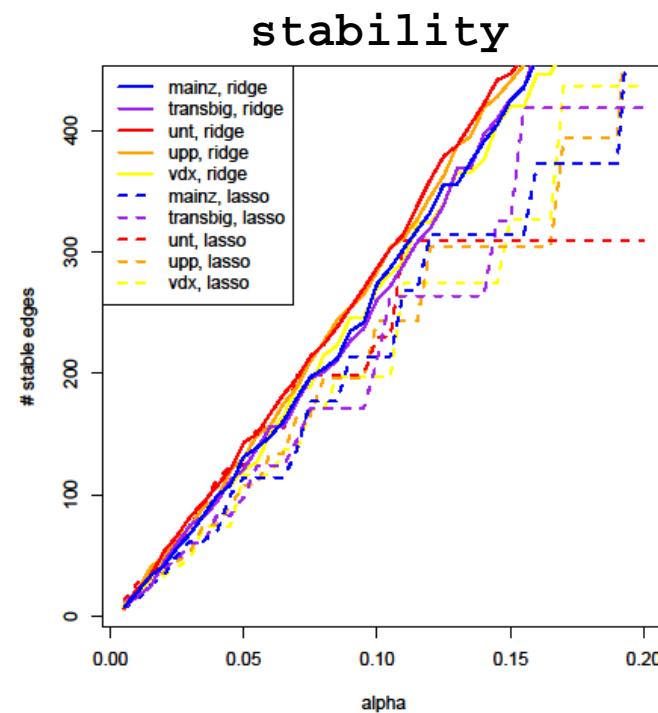
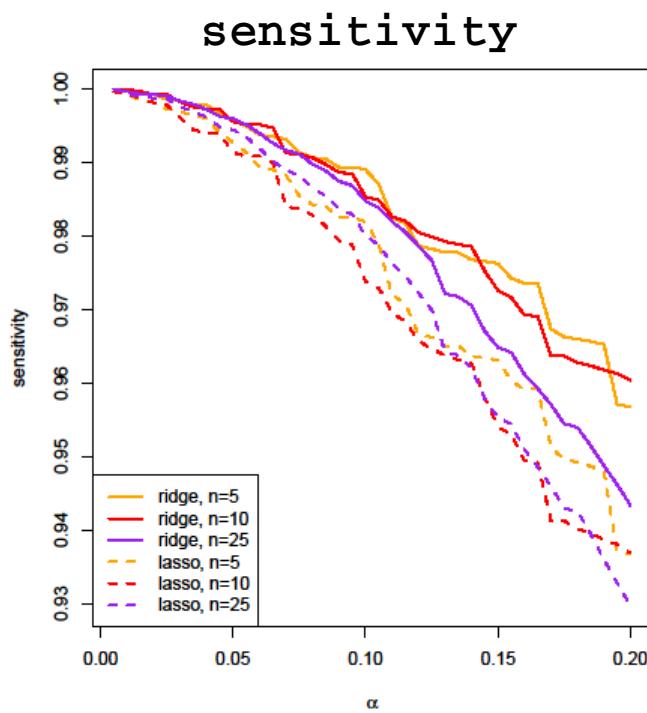
- breast cancer data
- subsample (100x)
- $n=5, 10, 25$
- LOOCV
- population  
= all samples



# Ridge vs. lasso

## *Selection and stability*

- 5 breast cancer data sets
- consensus truth
- 4 out of 5
- true + stable



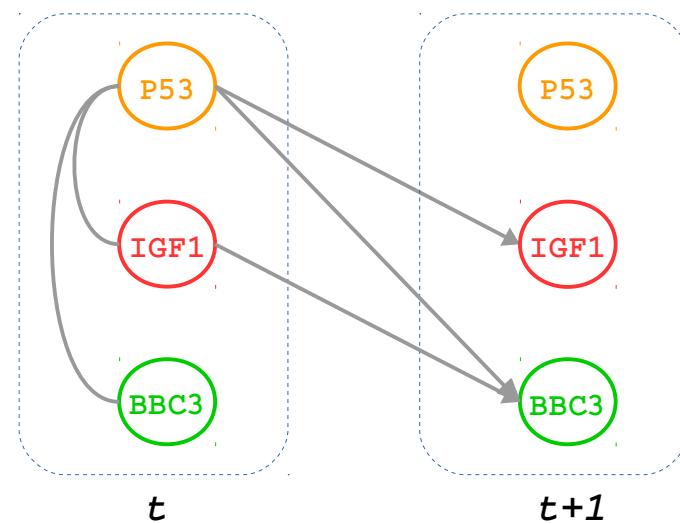
# Extensions

---

*Multiple groups*



*Time-course*

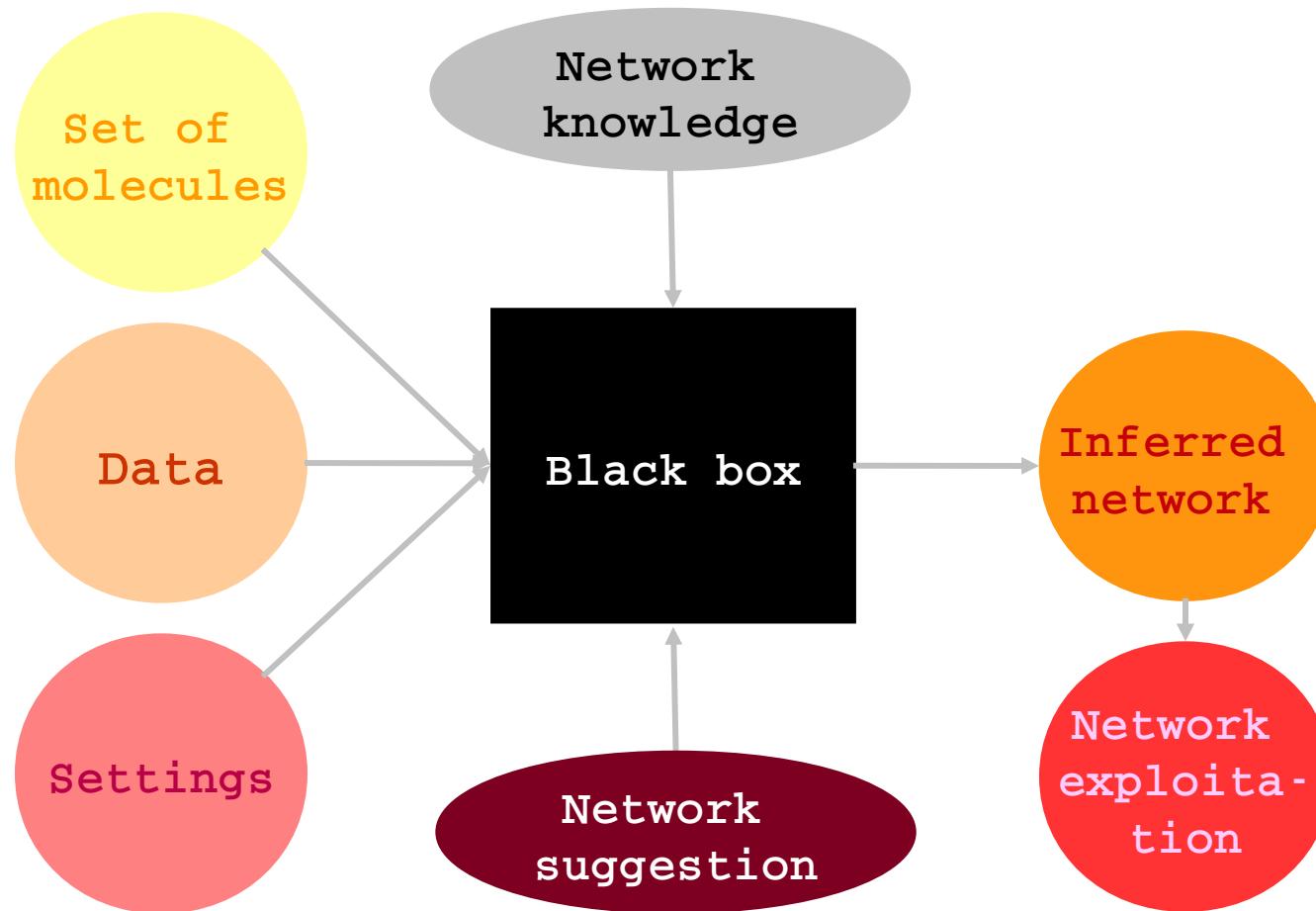


# How?

---

R-packages for network reconstruction:

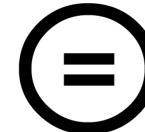
- `rags2ridges`: static data
- `ragt2ridges`: time course data.



# References

---

- Bilgrau, A. E., Peeters, C. F.W., Eriksen, P. S., Bøgsted, M., van Wieringen, W. N. (2015). "Targeted Fused Ridge Estimation of Inverse Covariance Matrices from Multiple High-Dimensional Data Classes". arXiv preprint arXiv:1509.07982.
- Efron, B. (2004). "Large-scale simultaneous hypothesis testing". *Journal of the American Statistical Association*, 99(465).
- Friedman, J., Hastie, T., Tibshirani, R. (2008), "Sparse inverse covariance estimation with the graphical lasso", *Biostatistics*, 9(3), 432-441.
- Ledoit, O., & Wolf, M. (2004). "A well-conditioned estimator for large-dimensional covariance matrices". *Journal of Multivariate Analysis*, 88(2), 365-411.
- Miok, V., Wilting, S.M., van Wieringen, W.N. (2016), "Ridge estimation of the VAR(1) model and its time series chain graph from multivariate time-course omics data", *Biometrical Journal*, 59(1), 172-191.
- Shafer, J., Strimmer, K. (2005), "A Shrinkage Approach to Large-Scale Covariance Matrix Estimation and Implications for Functional Genomics", *Statistical Applications in Genetics and Molecular Biology* , 4, Article 32.
- van Wieringen, W.N., Peeters, C.F.W. (2016), "Ridge estimation of inverse covariance matrices from high-dimensional data", *Computational Statistics and Data Analysis*, 103, 284-303.
- van Wieringen, W.N. (2017) "On the mean squared error of the ridge estimator of the covariance and precision matrix", *Statistics and Probability Letters*, 123, 88-92.
- Whittaker, J. (2009), Graphical models in applied multivariate statistics, Wiley.



This material is provided under the Creative Commons Attribution/Share-Alike/Non-Commercial License.

See <http://www.creativecommons.org> for details.

Slides by: Wessel N. van Wieringen