rags2ridges
Ridge estimation and graphical modeling for high-dimensional precision matrices

Carel F.W. Peeters* Anders E. Bilgrau† Wessel N. van Wieringen*‡

*Dept. of Epidemiology & Biostatistics, VU University medical center, Amsterdam
† Dept. of Mathematical Sciences, Aalborg University
‡Dept. of Mathematics, VU University, Amsterdam

useR! 2015
Aalborg, Denmark
July 1, 2015

cf.peeters@vumc.nl

Twinkies don't go bad

The graphical Lasso will always result in bliss

G +

=

graphical modeling bliss

Gaussian graphical modeling

Graphical modeling

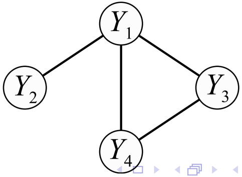
Class of models using graphs to express conditional (in)dependence relations between random variables

Gaussian setting

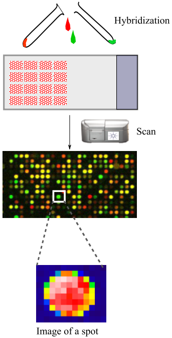
- Vertices: Correspond to random variables with normal distribution
- Edges: Correspond to the dependence structure
- Say $\mathbf{y} \sim \mathcal{N}_p(\mathbf{0}, \mathbf{\Sigma})$, and define $\mathbf{\Sigma}^{-1} \equiv \mathbf{\Omega}$. Then, for $a, b \in$ vertex set V, $a \neq b$

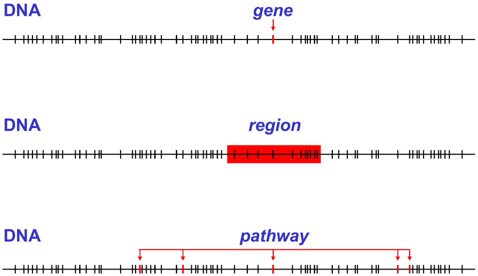
$$-\frac{\omega_{ab}}{\sqrt{\omega_{aa}\omega_{bb}}} = 0 \iff \omega_{ab} = 0 \iff a \perp\!\!\!\perp b | V \setminus \{a, b\} \iff a \not\sim b$$

$$\begin{bmatrix} \omega_{11} & \omega_{12} & \omega_{13} & \omega_{14} \\ \omega_{21} & \omega_{22} & 0 & 0 \\ \omega_{31} & 0 & \omega_{33} & \omega_{34} \\ \omega_{41} & 0 & \omega_{43} & \omega_{44} \end{bmatrix}$$



Challenge: High-dimensional data





Problem

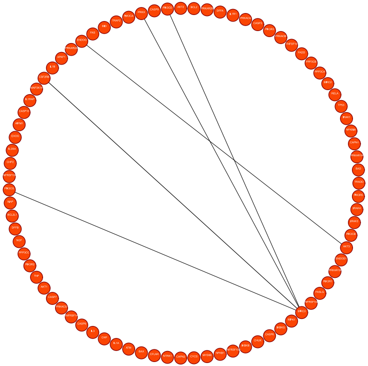
- Let $\hat{\mathbf{\Sigma}}$ denote the sample covariance matrix on $\mathbf{y}_i$
- When $p \approx n$ or $p > n$, $\hat{\mathbf{\Sigma}}$ is ill-behaved or singular
- The precision $\hat{\mathbf{\Sigma}}^{-1} \equiv \hat{\mathbf{\Omega}}$ is then undefined

ℓ_1-Penalization: The graphical Lasso

Maximize

$$\underbrace{\ln|\mathbf{\Omega}| - \mathrm{tr}(\hat{\mathbf{\Sigma}}\mathbf{\Omega})}_{\mathrm{log-likelihood}} - \underbrace{\lambda\|\mathbf{\Omega}\|_1}_{\ell_1-\mathrm{penalty}}$$

λ = Tuning parameter



ℓ_2-Penalization: Proper ridge

Maximize

$$\underbrace{\ln|\boldsymbol{\Omega}| - \mathrm{tr}(\hat{\boldsymbol{\Sigma}}\boldsymbol{\Omega})}_{\text{log}-\text{likelihood}} - \underbrace{\frac{\lambda}{2}\|\boldsymbol{\Omega} - \mathbf{T}\|_2^2}_{\ell_2-\text{penalty}}$$

- $\mathbf{T}$ denotes a p.d. symmetric target matrix
- $\lambda \in (0, \infty)$ denotes a penalty parameter

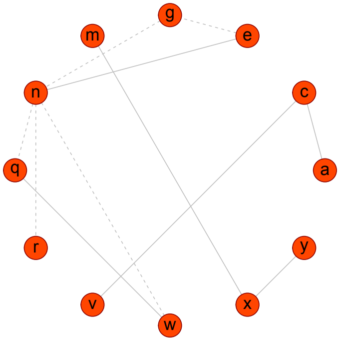
Analytic penalized ML estimator

$$\hat{\boldsymbol{\Omega}}(\lambda) = \left\{\left[\lambda\mathbf{I}_p + \frac{1}{4}(\hat{\boldsymbol{\Sigma}} - \lambda\mathbf{T})^2\right]^{1/2} + \frac{1}{2}(\hat{\boldsymbol{\Sigma}} - \lambda\mathbf{T})\right\}^{-1}$$

Implementation

Code

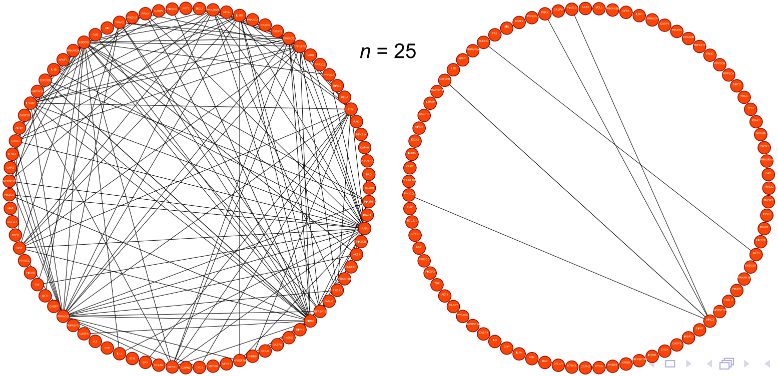
```
> CVres <- optPenalty.LOOCVauto(Y, lambdaMin = .001, lambdaMax = 30)
> rPrec <- ridgeS(covML(Y), CVres$optLambda)
> P0 <- sparsify(rPrec, threshold = "localFDR", FDRcut = 0.90)
> Ugraph(P0$sparseParCor, type = "fancy", prune = TRUE)
```



Comparison

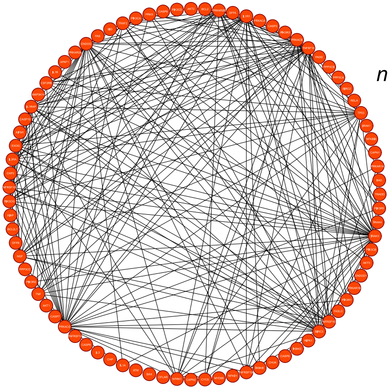
Data

- UPP ER+ breast cancer data (http://www.bioconductor.org/)
- Apoptosis pathway genes (as defined by KEGG)
- $p = 83$

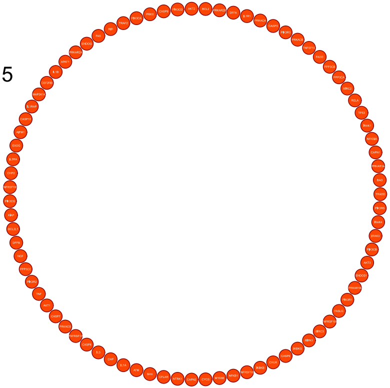


Comparison

$n = 5$



Graphical ridge



graphical lasso

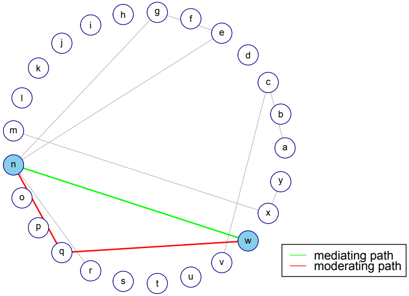
Network analysis

Node and network statistics

```
$degree
a b c d e f g h i j k l m n o p q r s t u v w x y
1 0 2 0 2 0 2 0 0 0 0 0 1 5 0 0 2 1 0 0 0 1 2 2 1

$betweenness
a b c d e f g h i j k l m n o p q r s t u v w x y
0 0 1 0 0 0 0 0 0 0 0 0 0 8 0 0 0 0 0 0 0 0 0 1 0
```

Path statistics



Message

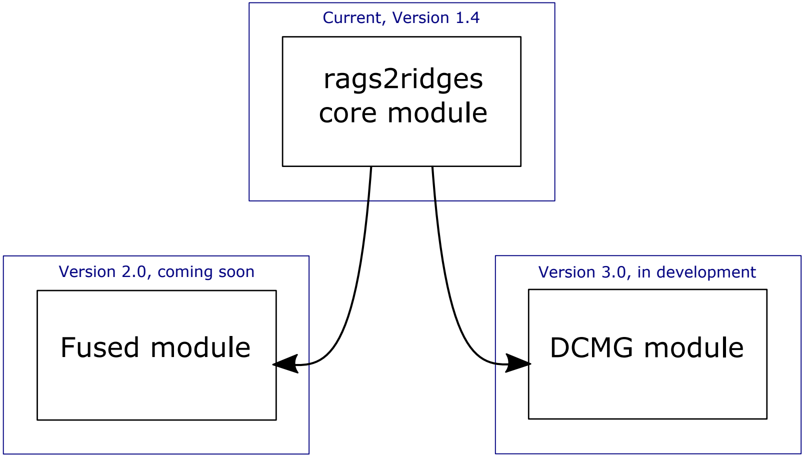
Viable alternative

rags2ridges contender in high-dimensional graphical modeling

Perks

- Fast
- Can handle high p/n ratios
- Integrates network inference and analysis

Developments



References

Inspirations

- `glasso` (Friendman, Hastie, & Tibshirani)
- `GeneNet` (Schaefer, Opgen-Rhein, & Korbinian Strimmer)

Making things easier

Imports

- `igraph` (Csardi *et al.*)
- `expm` (Goulet *et al.*)
- `reshape` (Wickham)
- `ggplot2` (Wickham)
- `Hmisc` (Harrell)
- `fdrtool` (Klaus & Strimmer)
- `snowfall` (Knaus)

References

- Bilgrau, A.E., C.F.W. Peeters, P.S. Eriksen, M. Bøgsted, & W.N. van Wieringen (in progress). Targeted Fused Ridge Estimation of Multiple Inverse Class Covariance Matrices from High-Dimensional Data.
- van Wieringen, W.N., & C.F.W. Peeters (to appear). 'Application of a New Ridge Estimator of the Inverse Covariance Matrix to the Reconstruction of Gene-Gene Interaction Networks', in: Proceedings of the Eleventh International Meeting on Computational Intelligence Methods for Bioinformatics and Biostatistics (CIBB 2014). New York: Springer (Lecture Notes in Computer Science series)
- van Wieringen, W.N., & C.F.W. Peeters (under review). Ridge Estimation of Inverse Covariance Matrices from High-Dimensional Data. arXiv:1403.0904 [stat.ME]


The End
thank you for listening