


Graph Wavelets to Analyze Genomic Data with Biological Networks

Yunlong Jiao and Jean-Philippe Vert



"Emerging Topics in Biological Networks and Systems Biology"
symposium, Swedish Collegium for Advanced Study, Uppsala,
October 11, 2017

abelprize.no




THE ABEL PRIZE

[ABOUT THE ABEL PRIZE](#) [LAUREATES](#) [THE ABEL YEARS](#) [MEDIA](#) [NIELS HE](#)

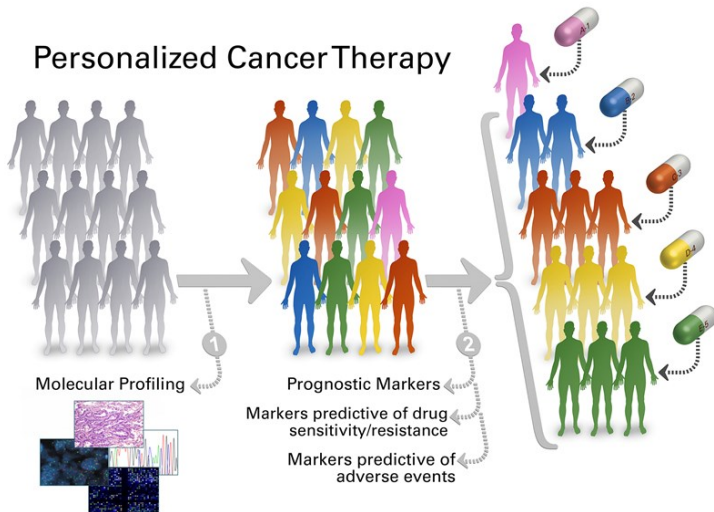
Yves Meyer received the Abel Prize from H.M. King Harald

H.M. King Harald presented the Abel Prize to Yves Meyer of the École normale supérieure Paris-Saclay, France at an award ceremony in Oslo on 23 May. He receives the prize for his pivotal role in the development of the mathematical theory of wavelets, says John Rognes, chair of the Abel committee. Among the prominent guests attending the award ceremony was the French ambassador to Norway, Jean-François Dobelle and the Norwegian Minister of Education and Research, Torbjørn Røe Isaksen.

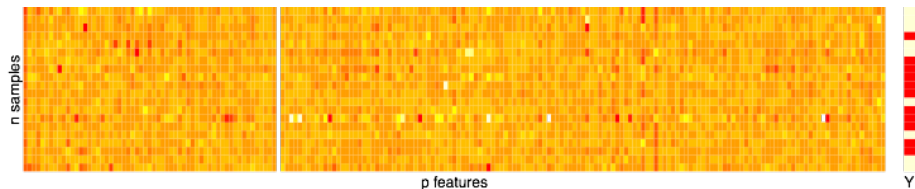


Motivation

Personalized Cancer Therapy

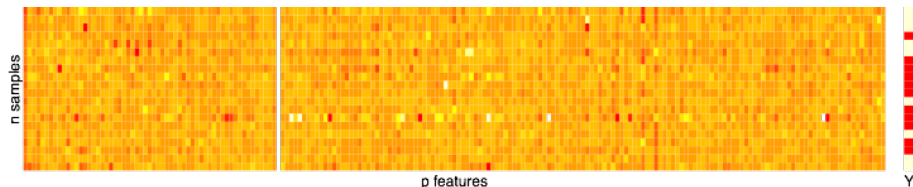


Typical problem



- $X \in \mathbb{R}^{n \times p}$ **gene expression** profile of each patient
- $Y \in \mathcal{Y}^n$ survival information of each patient
- $n = 10^2 \sim 10^4$
- $p = 2 \times 10^4$
- Goal: learn to predict Y from X
- Difficult ($n < p$)

Regularized linear models



Fit a linear model $\beta \in \mathbb{R}^p$ by solving

$$\min_{\beta \in \mathbb{R}^p} R(Y, X\beta) + \lambda J(\beta),$$

where

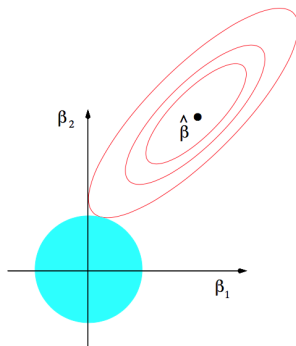
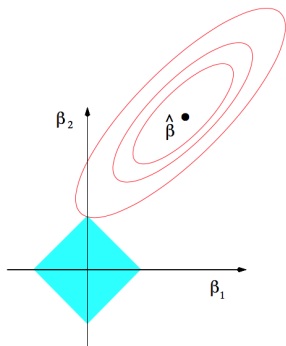
- $R(Y, X\beta)$ is an **empirical risk** to measures the fit to the training data
- $J(\beta)$ is a **penalty** to control the complexity of the model
- $\lambda > 0$ is a **regularization parameter**

Standard regularizations

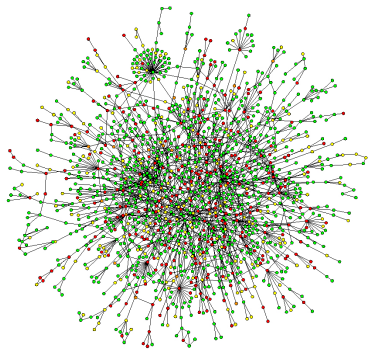
$$\min_{\beta \in \mathbb{R}^p} R(Y, X\beta) + \lambda J(\beta)$$

where

- **Lasso**: $J(\beta) = \|\beta\|_1$ for gene selection.
- **Ridge**: $J(\beta) = \|\beta\|_2^2$ to address $n \gg m$.
- **Elastic net**: $J(\beta) = \alpha\|\beta\|_2^2 + (1 - \alpha)\|\beta\|_1$

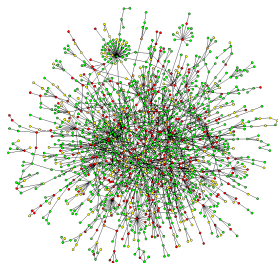


Network-based regularizations



- $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ a graph of genes
- $J_{\mathcal{G}}(\beta) = ?$
 - β should be "smooth" on the graph?
 - Selected genes should be connected?

Examples



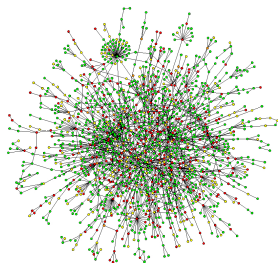
$$J_G(\beta) = \sum_{i \sim j} (\beta_i - \beta_j)^2 \quad (\text{Rapaport et al., 2007})$$

$$J_G(\beta) = a \|\beta\|_1 + (1 - a) \sum_{i \sim j} (\beta_i - \beta_j)^2 \quad (\text{Li and Li, 2008})$$

$$J_G(\beta) = \sup_{\alpha \in \mathbb{R}^p : \forall i \sim j \alpha_i^2 + \alpha_j^2 \leq 1} \alpha^\top \beta \quad (\text{Jacob et al., 2009})$$

$$J_G(\beta) = a \|\beta\|_1 + (1 - a) \sum_{i \sim j} |\beta_i - \beta_j| \quad (\text{Hoefling, 2010})$$

Examples



$$J_G(\beta) = \sum_{i \sim j} (\beta_i - \beta_j)^2 \quad (\text{Rapaport et al., 2007})$$

$$J_G(\beta) = a \|\beta\|_1 + (1 - a) \sum_{i \sim j} (\beta_i - \beta_j)^2 \quad (\text{Li and Li, 2008})$$

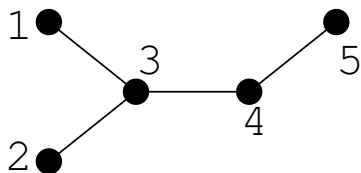
$$J_G(\beta) = \sup_{\alpha \in \mathbb{R}^p : \forall i \sim j \alpha_i^2 + \alpha_j^2 \leq 1} \alpha^\top \beta \quad (\text{Jacob et al., 2009})$$

$$J_G(\beta) = a \|\beta\|_1 + (1 - a) \sum_{i \sim j} |\beta_i - \beta_j| \quad (\text{Hoeffling, 2010})$$

From smoothness penalty to Laplacian

$$\sum_{i \sim j} (\beta_i - \beta_j)^2 = \beta^\top L \beta$$

where $L = D - A$ is the **graph Laplacian**.



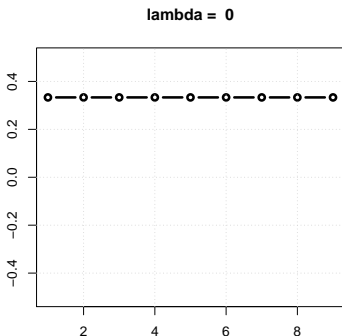
$$L = \begin{pmatrix} 1 & 0 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 \\ -1 & -1 & 3 & -1 & 0 \\ 0 & 0 & -1 & 2 & -1 \\ 0 & 0 & 0 & -1 & 1 \end{pmatrix}$$

From Laplacian to Fourier analysis: $L = U\Lambda U^T$

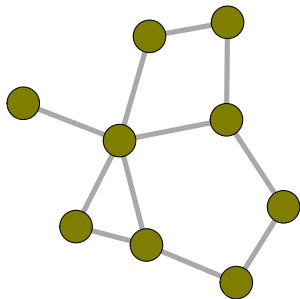
- Eigenvectors U of L form the Fourier basis:

$$\hat{\beta} = U^T \beta$$

- Eigenvalues $\Lambda = (0 = \lambda_1 \leq \dots \leq \lambda_p)$ represent the "frequencies" of the Fourier basis



Lambda = 0



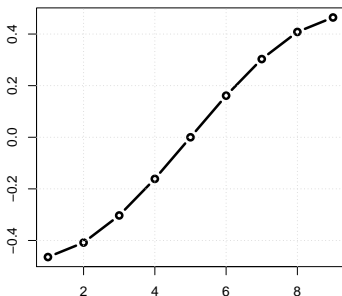
From Laplacian to Fourier analysis: $L = U\Lambda U^T$

- Eigenvectors U of L form the Fourier basis:

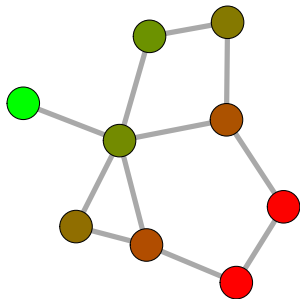
$$\hat{\beta} = U^T \beta$$

- Eigenvalues $\Lambda = (0 = \lambda_1 \leq \dots \leq \lambda_p)$ represent the "frequencies" of the Fourier basis

lambda = 0.12



Lambda = 0.76

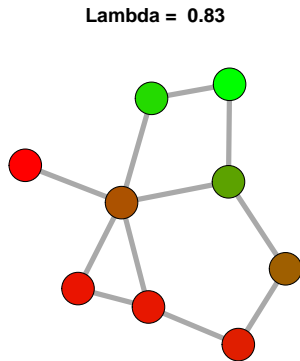
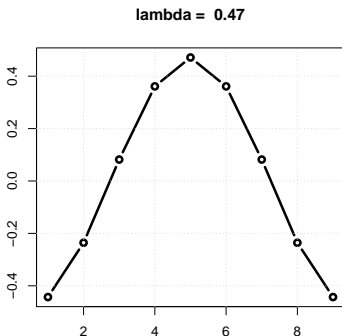


From Laplacian to Fourier analysis: $L = U\Lambda U^T$

- Eigenvectors U of L form the Fourier basis:

$$\hat{\beta} = U^T \beta$$

- Eigenvalues $\Lambda = (0 = \lambda_1 \leq \dots \leq \lambda_p)$ represent the "frequencies" of the Fourier basis

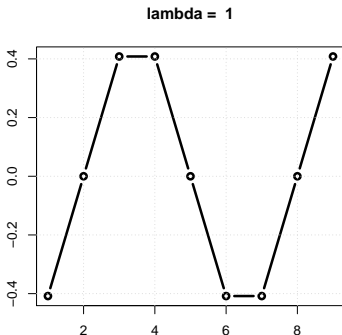


From Laplacian to Fourier analysis: $L = U\Lambda U^T$

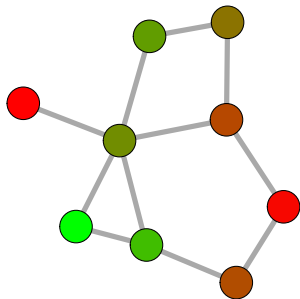
- Eigenvectors U of L form the Fourier basis:

$$\hat{\beta} = U^T \beta$$

- Eigenvalues $\Lambda = (0 = \lambda_1 \leq \dots \leq \lambda_p)$ represent the "frequencies" of the Fourier basis



Lambda = 1.3

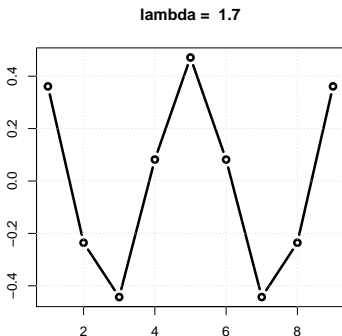


From Laplacian to Fourier analysis: $L = U\Lambda U^T$

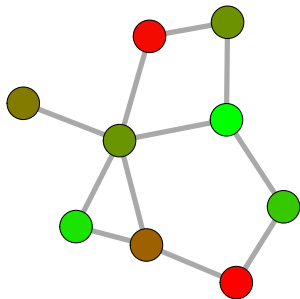
- Eigenvectors U of L form the Fourier basis:

$$\hat{\beta} = U^T \beta$$

- Eigenvalues $\Lambda = (0 = \lambda_1 \leq \dots \leq \lambda_p)$ represent the "frequencies" of the Fourier basis



Lambda = 2.2

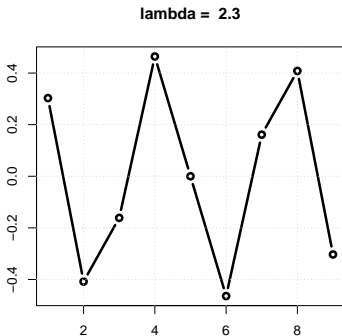


From Laplacian to Fourier analysis: $L = U\Lambda U^T$

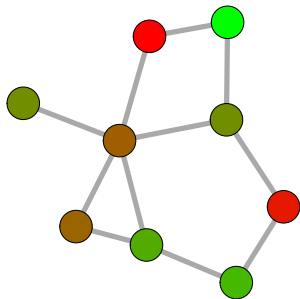
- Eigenvectors U of L form the Fourier basis:

$$\hat{\beta} = U^T \beta$$

- Eigenvalues $\Lambda = (0 = \lambda_1 \leq \dots \leq \lambda_p)$ represent the "frequencies" of the Fourier basis



Lambda = 2.8

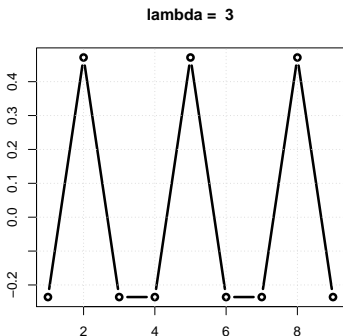


From Laplacian to Fourier analysis: $L = U\Lambda U^T$

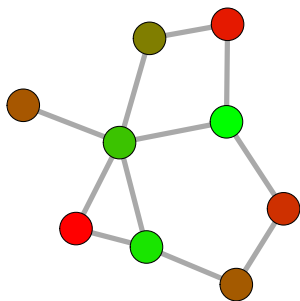
- Eigenvectors U of L form the Fourier basis:

$$\hat{\beta} = U^T \beta$$

- Eigenvalues $\Lambda = (0 = \lambda_1 \leq \dots \leq \lambda_p)$ represent the "frequencies" of the Fourier basis



Lambda = 3.6

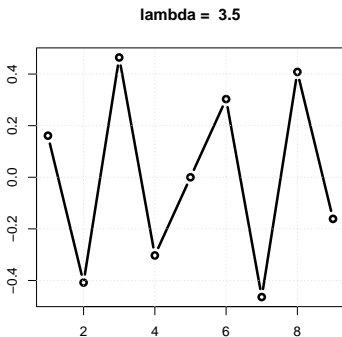


From Laplacian to Fourier analysis: $L = U\Lambda U^T$

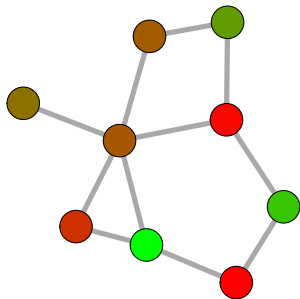
- Eigenvectors U of L form the Fourier basis:

$$\hat{\beta} = U^T \beta$$

- Eigenvalues $\Lambda = (0 = \lambda_1 \leq \dots \leq \lambda_p)$ represent the "frequencies" of the Fourier basis



Lambda = 4.2



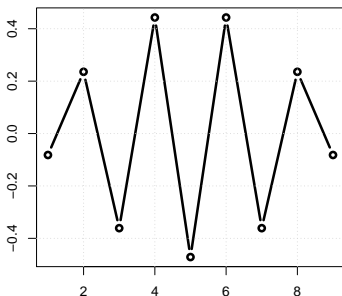
From Laplacian to Fourier analysis: $L = U\Lambda U^T$

- Eigenvectors U of L form the Fourier basis:

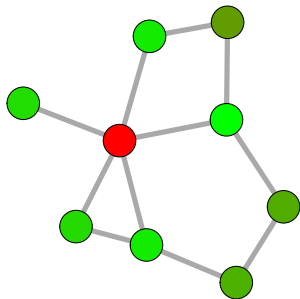
$$\hat{\beta} = U^T \beta$$

- Eigenvalues $\Lambda = (0 = \lambda_1 \leq \dots \leq \lambda_p)$ represent the "frequencies" of the Fourier basis

lambda = 3.9



Lambda = 6.3



Smoothness in the Fourier domain

- Therefore, the smoothness penalty penalizes Fourier coefficients corresponding to high frequencies:

$$J(\beta) = \sum_{i \sim j} (\beta_i - \beta_j)^2 = \beta^\top L \beta = \beta^\top U \Lambda U^\top \beta = \hat{\beta}^\top \Lambda \hat{\beta} = \sum_{i=1}^p \lambda_i \hat{\beta}_i^2$$

*"the linear model mapped on the graph
should have little energy at high frequency"*

- Rapaport et al. (2007) extends this to more general penalties:

$$J_\phi(\beta) = \sum_{i=1}^p \phi(\lambda_i) \hat{\beta}_i^2 \quad \text{s.t.} \quad \beta = U \hat{\beta}$$

for $\phi : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ non-decreasing.

Fourier vs wavelets

Fourier



Localized in frequency

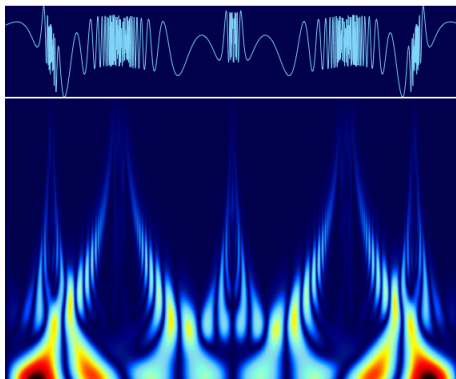
Wavelets



Localized in frequency AND space

Wavelets on graphs

- A family of vectors $\{\Psi_{v,s}\} \subset \mathbb{R}^p$ where
 - $v \in [1, p]$ is a vertex (space)
 - $s \in \mathbb{R}^+$ is a scale (frequency)
- In practice we choose a small number of scales $s_1 < \dots < s_S$
- This results in $p \times S$ vectors (overcomplete basis)



Example on graphs

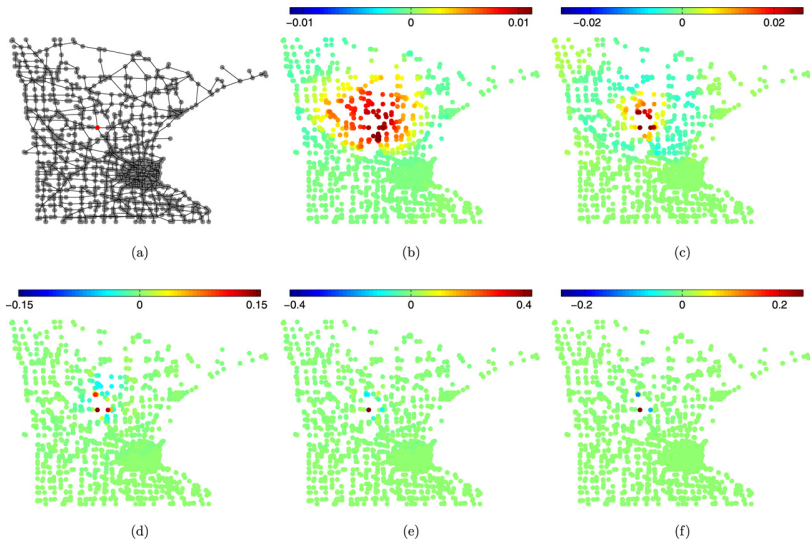
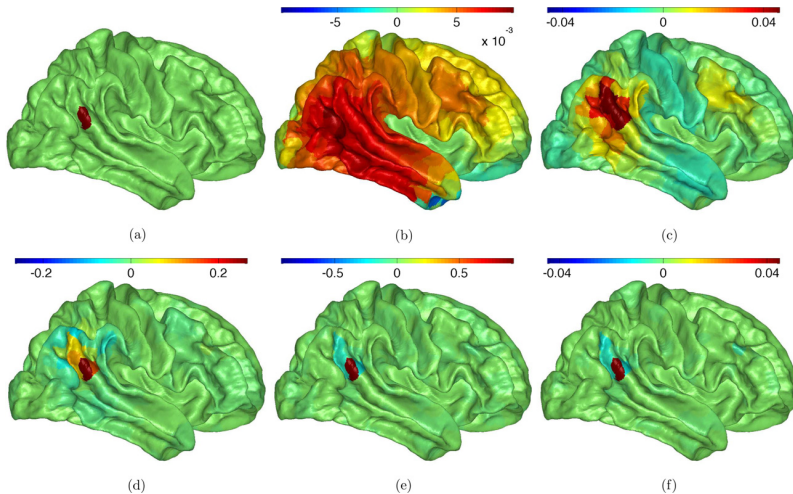


Fig. 4. Spectral graph wavelets on Minnesota road graph, with $K = 100$, $J = 4$ scales. (a) Vertex at which wavelets are centered, (b) scaling function, (c)–(f) wavelets, scales 1–4.

Example on graphs



(Hammond et al., 2011)

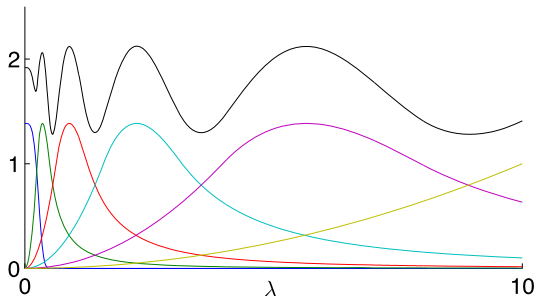
How to make the wavelet basis $\{\Psi_{v,t}\}$?

- Hammond et al. (2011) propose **spectral graph wavelets**
- Formally, at scale $s > 0$,

$$\Psi_s = (\Psi_{1,s} \mid \dots \mid \Psi_{p,s}) = Ug(s\Lambda)U^\top$$

where $g : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ is a function that satisfies

- g is a band-pass filter (localization in frequency)
- g is smooth near 0 (this ensures localization in space)



Graph wavelet-based regularization

- Given a graph, compute a redundant set of $S \times p$ wavelet basis at different scales:

$$B = (\Psi_{s_1} \mid \dots \mid \Psi_{s_S})$$

- Take for penalty the atomic norm:

$$J(\beta) = \min \left\{ \sum_{i=1}^{S \times p} |c_i| : \beta = \sum_{i=1}^{S \times p} c_i B_i \right\}$$

- $J(\beta)$ is small when β is a sum of a few atoms
- The atom $\Psi_{s,v}$ has weights in a neighborhood of v of "size" s

Summary: Fourier vs. wavelet penalty

$$\min_{\beta \in \mathbb{R}^p} R(Y, X\beta) + \lambda J(\beta)$$

- Fourier:

$$J_\phi(\beta) = \|\phi(\Lambda)\hat{\beta}\|_2^2 \quad \text{s.t.} \quad \beta = U\hat{\beta}$$

β will be **smooth** on the graph

- Wavelet

$$J(\beta) = \min \{\|c\|_1 : \beta = Bc\}$$

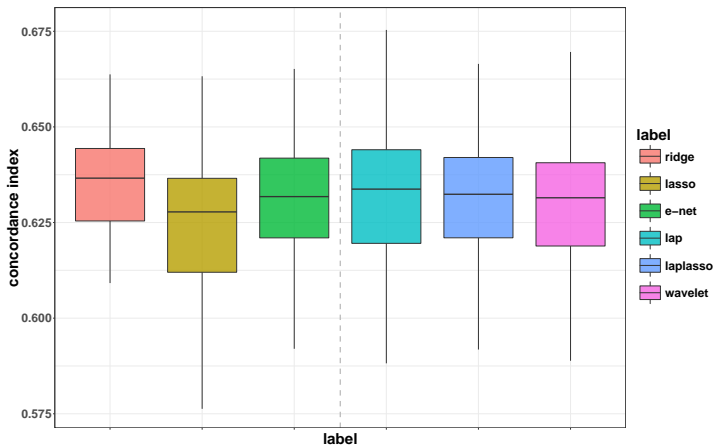
β will decompose as a **sum of localized functions** on the graph (pathways?)

Experiment

- Protein-protein interaction (PPI) network obtained from Human Protein Reference Database (HPRD).
- METABRIC breast cancer dataset
 - $n = 1,981$ breast cancer samples paired with survival information of patients.
 - Expression data of a total of 24,771 genes available, among which $p = 9,117$ genes are found with known interaction in HPRD.
- Benchmark study comparing **6 penalty functions**:

Label	Penalty function $J(\beta)$	Network-based	Gene selection
ridge	$\ \beta\ _2^2$		
lasso	$\ \beta\ _1$		✓
e-net	$a\ \beta\ _1 + (1-a)\ \beta\ _2^2$		✓
lap	$\sum_{i \sim j} (\beta_i - \beta_j)^2$	✓	
laplasso	$a\ \beta\ _1 + (1-a) \sum_{i \sim j} (\beta_i - \beta_j)^2$	✓	✓
wavelet	$\min_{\theta} \ \theta\ _1 \quad \text{s.t. } \beta = \Psi\theta$	✓	✓

Prediction performance



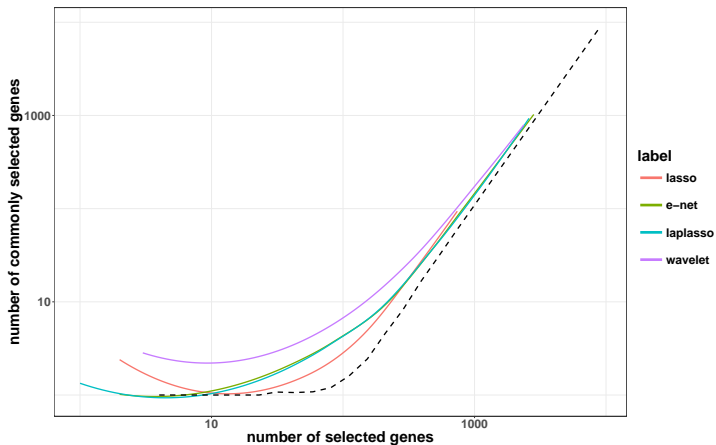
Boxplots on survival risk prediction performance evaluated by concordance index scores over 5-fold cross-validation repeated 10 times of the METABRIC data.

Prediction performance

Label	Mean CI scores (\pm SD)	Network-based	Gene selection
ridge	0.636 (\pm 0.018)		
lap	0.632 (\pm 0.0193)	✓	
laplasso	0.6312 (\pm 0.0185)	✓	✓
e-net	0.6304 (\pm 0.0183)		✓
wavelet	0.6295 (\pm 0.0198)	✓	✓
lasso	0.626 (\pm 0.0177)		✓

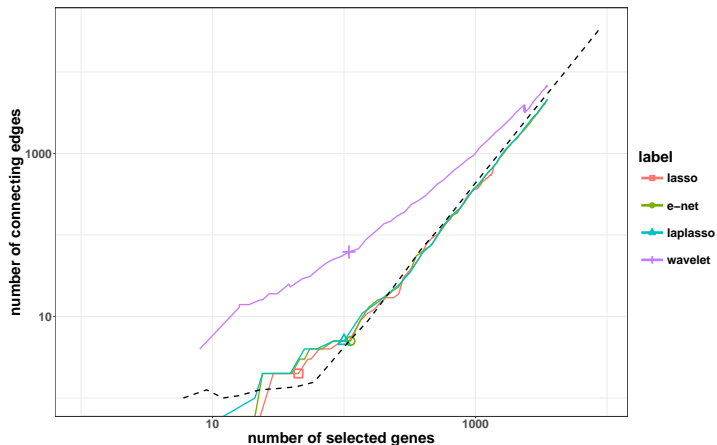
Mean concordance index (CI) scores (\pm standard deviation) of survival risk prediction over 5-fold cross-validation repeated 10 times of the METABRIC data. Methods are ordered by decreasing mean CI scores.

Gene selection performance: Stability



Stability performance of gene selection related to breast cancer survival, estimated over 100 random experiments. The black dotted curve denotes random selection.

Gene selection performance: Connectivity



Connectivity performance of gene selection related to breast cancer survival, where special marks correspond to the number tuned by cross-validation. The black dotted curve denotes random selection.

Gene selection performance: Interpretability

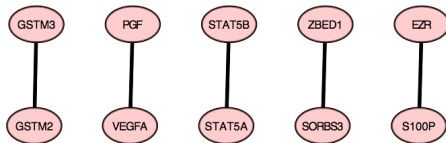


Figure: Gene subnetworks related to breast cancer survival identified by regularization methods identified by the elastic net (10 genes connected out of 112 selected) or the Laplacian lasso (10 genes connected out of 100 selected).

Gene selection performance: Interpretability

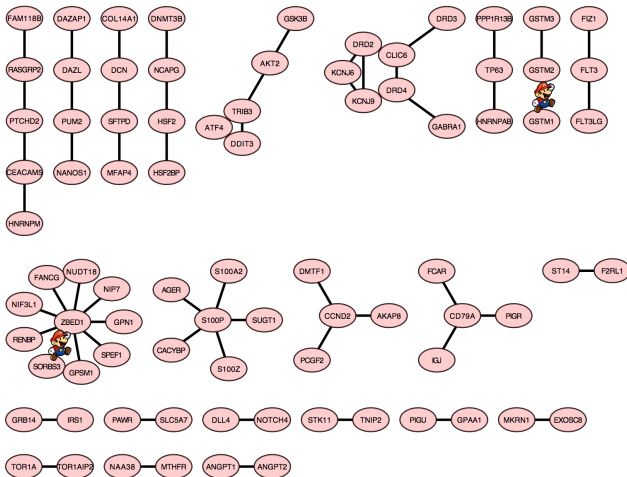


Figure: Gene subnetworks related to breast cancer survival identified by regularization methods identified by network-based wavelet smoothing (82 genes connected out of 109 selected).

Conclusion

- Can biological networks help define a **structure** on high-dimensional omics data?
- Fourier-based penalties (smoothness, diffusion...) already exist
- Wavelets-based penalties decompose a signal over a basis
 - localized in space
 - localized in frequency
- Preliminary results on gene expression classification

Thanks



Inserm

Institut national
de la santé et de la recherche médicale



ENS
ÉCOLE NORMALE
SUPÉRIEURE



The Adolph C. and Mary Sprague
Miller Institute for Basic
Research in Science
University of California, Berkeley



SIMONS
INSTITUTE
for the Theory of Computing

References

- D. K. Hammond, P. Vandergheynst, and R. Gribonval. Wavelets on graphs via spectral graph theory. *Applied and Computational Harmonic Analysis*, 30(2):129–150, 2011. ISSN 1063-5203. doi: 10.1016/j.acha.2010.04.005.
- H. Hoefling. A path algorithm for the Fused Lasso Signal Approximator. *J. Comput. Graph. Stat.*, 19(4):984–1006, 2010. doi: 10.1198/jcgs.2010.09208. URL <http://dx.doi.org/10.1198/jcgs.2010.09208>.
- L. Jacob, G. Obozinski, and J.-P. Vert. Group lasso with overlap and graph lasso. In *ICML '09: Proceedings of the 26th Annual International Conference on Machine Learning*, pages 433–440, New York, NY, USA, 2009. ACM. ISBN 978-1-60558-516-1. doi: 10.1145/1553374.1553431. URL <http://dx.doi.org/10.1145/1553374.1553431>.
- C. Li and H. Li. Network-constrained regularization and variable selection for analysis of genomic data. *Bioinformatics*, 24:1175–1182, May 2008. ISSN 1367-4811. doi: 10.1093/bioinformatics/btn081.
- F. Rapaport, A. Zinovyev, M. Dutreix, E. Barillot, and J.-P. Vert. Classification of microarray data using gene networks. *BMC Bioinformatics*, 8:35, 2007. doi: 10.1186/1471-2105-8-35. URL <http://dx.doi.org/10.1186/1471-2105-8-35>.
- R. Tibshirani. Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. Ser. B*, 58(1): 267–288, 1996. URL <http://www.jstor.org/stable/2346178>.