Contribution à l'analyse de données non vectorielles

Nathalie Villa-Vialaneix

nathalie.villa@toulouse.inra.fr
http://www.nathalievilla.org

Soutenance HDR, 13 novembre 2014
INRA, Toulouse

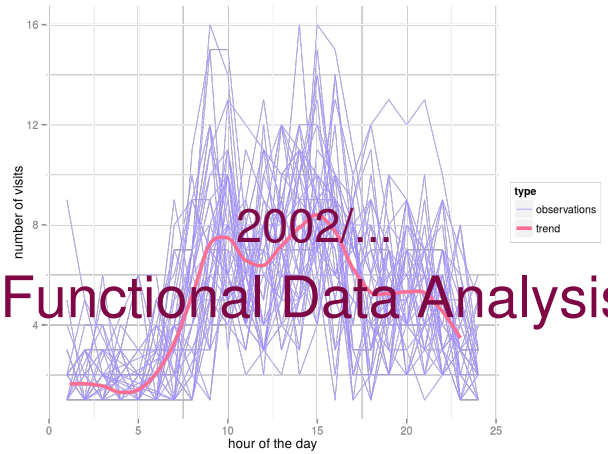
Overview of the topics

data: non vectorial data,
high dimension...
methods: machine learning & data
mining, kernel methods, neuronal
methods...

Overview of the topics

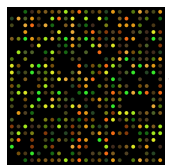
data: non vectorial data, high dimension...

methods: machine learning & data mining, kernel methods, neuronal methods...



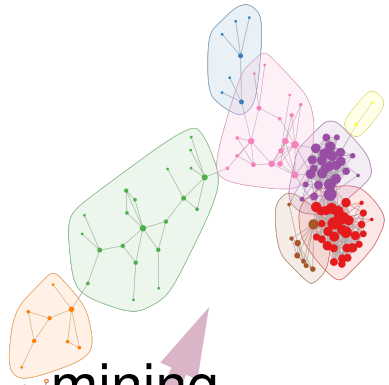
2002/...

Functional Data Analysis



inference

2012/...



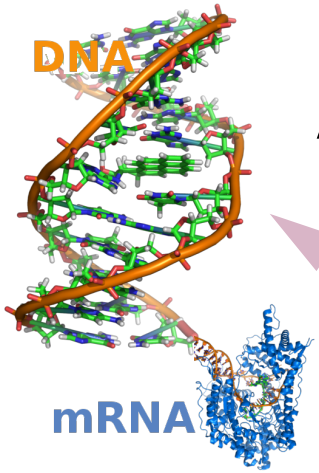
mining

visualization, clustering

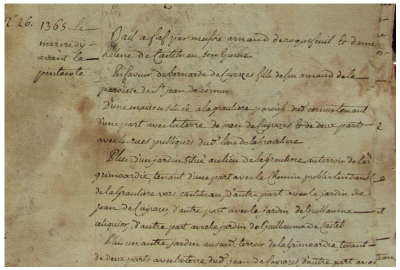
2006/...

Graph/Network

Overview of the topics



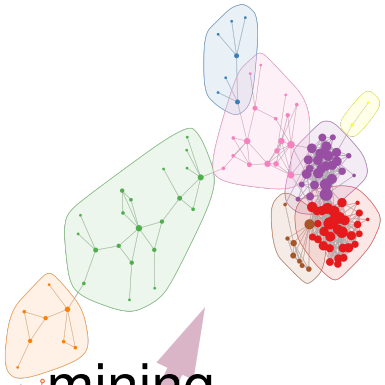
Applications in biology



Applications in human sciences

data: non vectorial data, high dimension...

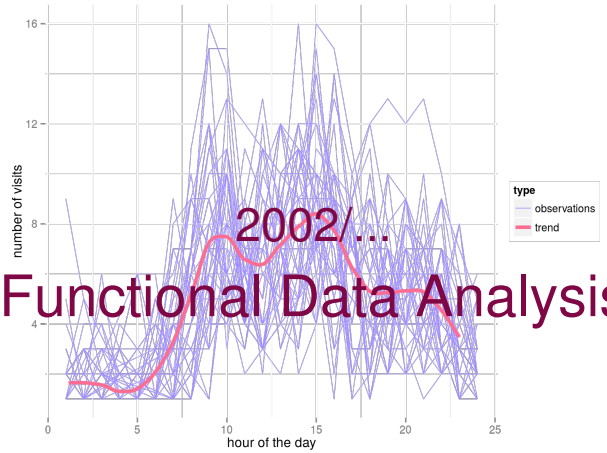
methods: machine learning & data mining, kernel methods, neuronal methods...



mining

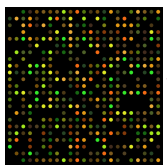
visualization, clustering

2006/...



2002/...

Functional Data Analysis



inference

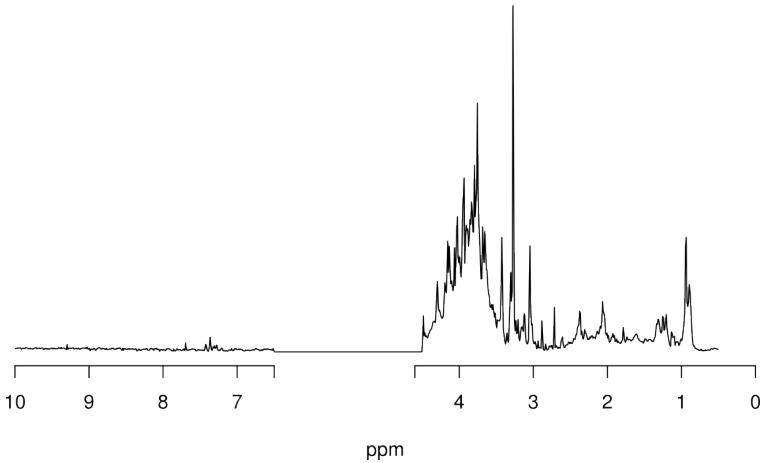
2012/...

Graph/Network

Functional Data Analysis

FDA framework and examples

Observations of a random variable X taking values in a space of functions $\mathcal{X}$ (typically, an infinite dimensional Hilbert space as L^2):
$x_1, \ldots, x_n$.
The function is observed at $t_{i1}, \ldots, t_{id_i}$.



Applications in: time series analysis, speech recognition, biochemistry (NIR spectra for instance), metabolomic data (NMR), weather data...

Typical issues in FDA

1. variance operator used in standard model does not have a continuous inverse: $\Gamma_X = \mathbb{E}(X \otimes X) - \mathbb{E}(X) \otimes \mathbb{E}(X)$ is a Hilbert Schmidt operator
   $\Rightarrow \Gamma_X^{-1}$ is not bounded.
   As a consequence, $\Gamma_X^n = \frac{1}{n} \sum_i x_i \otimes x_i - \overline{x} \otimes \overline{x}$ is ill-conditionned (and its inverse a bad estimate of Γ_X^{-1}).
   $\Rightarrow$ regularization or penalization techniques are needed.

Typical issues in FDA

1. variance operator used in standard model does not have a continuous inverse: $\Gamma_X = \mathbb{E}(X \otimes X) - \mathbb{E}(X) \otimes \mathbb{E}(X)$ is a Hilbert Schmidt operator
   $\Rightarrow \Gamma_X^{-1}$ is not bounded.
   As a consequence, $\Gamma_X^n = \frac{1}{n}\sum_i x_i \otimes x_i - \overline{x} \otimes \overline{x}$ is ill-conditionned (and its inverse a bad estimate of Γ_X^{-1}).
   $\Rightarrow$ regularization or penalization techniques are needed.
2. $(x_i)_i$ are never perfectly observed and only a digitized (sometimes noisy) estimation is available.
   $\Rightarrow$ reconstruction techniques are needed to provide a functional representation of the data, remove noise and measurement artefacts (translation, scaling, ..., of the functions).

Overview of my contributions

Supervised learning framework

(X, Y) st $X \in \mathcal{X}$, $Y \in \mathbb{R}$ (regression) or $Y \in \{-1, 1\}$ (binary classification).
Observations: $(x_i, y_i)_{i=1,\dots,n}$
Purpose: estimate $\hat{y}$ for a new x. $(x_i, y_i)_i$ is used to define a prediction function Φ^n.

Overview of my contributions

Supervised learning framework

(X, Y) st $X \in \mathcal{X}$, $Y \in \mathbb{R}$ (regression) or $Y \in \{-1, 1\}$ (binary classification).
Observations: $(x_i, y_i)_{i=1,\dots,n}$
Purpose: estimate $\hat{y}$ for a new x. $(x_i, y_i)_i$ is used to define a prediction function Φ^n.

- Inverse methods use $\mathcal{L}(X|Y)$ to estimate $\mathcal{L}(Y|X)$
  - [2] (Ferré & Villa, *Scandinavian Journal of Statistics* , 2006) in the FIR (Functional Inverse Regression) model

$$Y = F\left(\langle X, \beta_1 \rangle, \dots, \langle X, \beta_d \rangle, \epsilon\right),$$

with F and $(\beta_j)_j$ to be estimated, smooth estimation of the $(\beta_j)_j$ combined with the estimation of F by multi-layer perceptron.

Overview of my contributions

Supervised learning framework

(X, Y) st $X \in \mathcal{X}$, $Y \in \mathbb{R}$ (regression) or $Y \in \{-1, 1\}$ (binary classification).
Observations: $(x_i, y_i)_{i=1,\dots,n}$
Purpose: estimate $\hat{y}$ for a new x. $(x_i, y_i)_i$ is used to define a prediction function Φ^n.

- Inverse methods use $\mathcal{L}(X|Y)$ to estimate $\mathcal{L}(Y|X)$
  - [2] (Ferré & Villa, *Scandinavian Journal of Statistics* , 2006)
  - [3] (Hernández, *et al.*, *Statistica Sinica*, 2014) calibration problem in chemiometry: under the assumption that $\mathcal{L}(X|Y)$ is Gaussian, estimation of

$$f(x|y) = \exp\left[\sum_{j\geq 1} \frac{r_j(y)}{\lambda_j}\left(x_j - \frac{r_j(y)}{2}\right)\right]$$

which is used in a plug-in estimate of $\mathbb{E}(Y|X = x)$.

Overview of my contributions

Supervised learning framework

(X, Y) st $X \in \mathcal{X}$, $Y \in \mathbb{R}$ (regression) or $Y \in \{-1, 1\}$ (binary classification).
Observations: $(x_i, y_i)_{i=1,\dots,n}$
Purpose: estimate $\hat{y}$ for a new x. $(x_i, y_i)_i$ is used to define a prediction function Φ^n.

- Inverse methods use $\mathcal{L}(X|Y)$ to estimate $\mathcal{L}(Y|X)$
  - [2] (Ferré & Villa, *Scandinavian Journal of Statistics* , 2006)
  - [3] (Hernández, *et al.*, *Statistica Sinica*, 2014)
- Kernel methods (less sensitive to high dimension, can include some functional pre-processing)
  - [6] (Rossi & Villa, *Neurocomputing*, 2006) SVM for functional data analysis
  - [8] (Rossi & Villa-Vialaneix, *Pattern Recognition Letter*, 2011) derivative-based kernel methods for functional classification and regression

Overview of my contributions

Supervised learning framework

(X, Y) st $X \in \mathcal{X}$, $Y \in \mathbb{R}$ (regression) or $Y \in \{-1, 1\}$ (binary classification).
Observations: $(x_i, y_i)_{i=1,\dots,n}$
Purpose: estimate $\hat{y}$ for a new x. $(x_i, y_i)_i$ is used to define a prediction function Φ^n.

- Inverse methods use $\mathcal{L}(X|Y)$ to estimate $\mathcal{L}(Y|X)$
  - [2] (Ferré & Villa, *Scandinavian Journal of Statistics* , 2006)
  - [3] (Hernández, *et al.*, *Statistica Sinica*, 2014)
- Kernel methods
  - [6] (Rossi & Villa, *Neurocomputing*, 2006)
  - [8] (Rossi & Villa-Vialaneix, *Pattern Recognition Letter*, 2011)
- [5, 12], (Rohart, *et al.*, *Journal of Animal Science*, 2012), (Villa-Vialaneix, *et al.*, *Communication in Statistics*, 2014) biomarker identification from metabolomic (NMR) data using functional approaches

Kernels for functional data

Given $K : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ st

- symmetry: $K(x, x') = K(x', x)$
- positivity: $\forall\, N \in \mathbb{N}, \forall\, (\alpha_i) \subset \mathbb{R}^N, \forall\, (x_i) \subset \mathcal{X}^N, \sum_{i,j} \alpha_i \alpha_j K(x_i, x_j) \geq 0.$

$\exists!\ (\mathcal{H}, \langle ., . \rangle_{\mathcal{H}})$ (RKHS) and $\Psi : \mathcal{X} \to \mathcal{H}$ st $K(x, x') = \langle \Psi(x), \Psi(x') \rangle_{\mathcal{H}}$

General form for FD

pre-processing: $\mathcal{P} : \mathcal{X} \to \mathcal{D}$

$$\forall\, x, x' \in \mathcal{X}, Q(x, x') = K(\mathcal{P}(x), \mathcal{P}(x')).$$

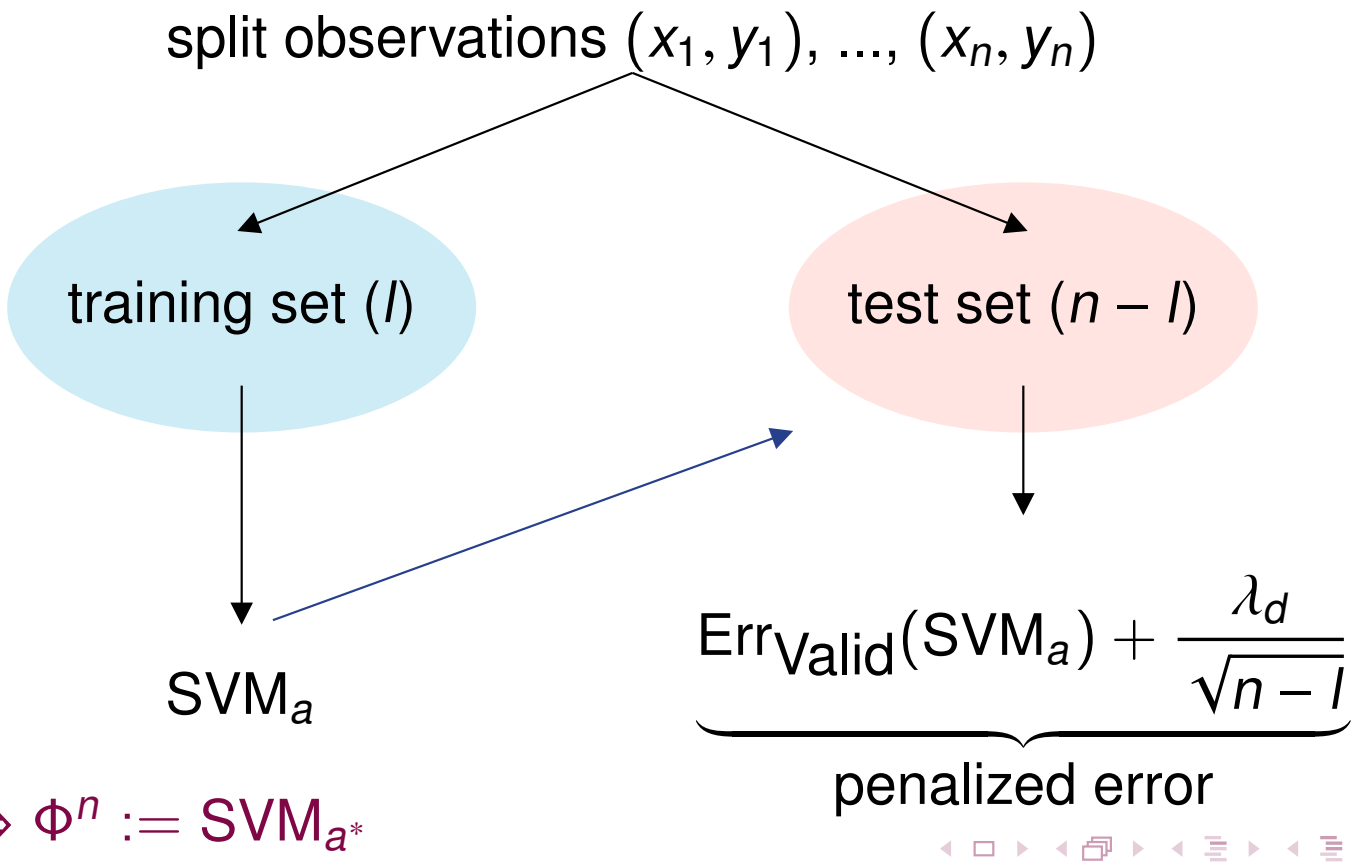
1. projections: for $V_d = \text{Vect}\{\psi_1, \ldots, \psi_d\}$, $\mathcal{P}(x) = \sum_{j=1}^{d} \langle x, \psi_j \rangle \psi_j$ (and $K = K_d$, standard kernel on $\mathbb{R}^d$).
2. functional transformation: $\mathcal{P}(x) = D^q x, \ldots$
3. FIR...

A consistent approach

[6] (Rossi & Villa, *Neurocomputing*, 2006)

SVM for functional data: $(X, Y) \in \mathcal{X} \times \{-1, 1\}$ with kernel $Q = K_d \circ \mathcal{P}_{V_d}$, K_d standard kernel on $\mathbb{R}^d$

selection of $a = (d, K_d, C) \in \mathbb{N} \times \mathcal{I}_d \times [0, C_d]$:



$\Rightarrow \Phi^n := \mathrm{SVM}_{a^*}$

Assumptions

[6] (Rossi & Villa, *Neurocomputing*, 2006)

Assumptions on the law of X

(A1) X takes its values in a bounded subspace of $\mathcal{X}$.

Assumptions on the parameters: $\forall\, d \geq 1$,

(A2) $\mathcal{J}_d$ is a finite set;
(A3) $\exists K_d \in \mathcal{J}_d$ (kernels on $\mathbb{R}^d$) st: K_d is universal and
$\exists \nu_d > 0 : \ \mathcal{N}(K_d, \epsilon) = \mathcal{O}(\epsilon^{-\nu_d})$;
(A4) $C_d > 1$;
(A5) $\sum_{d\geq 1} |\mathcal{J}_d| e^{-2\lambda_d^2} < +\infty$.

Assumptions on the training/validation sets

(A6) $\lim_{n\to+\infty} l = +\infty$;
(A7) $\lim_{n\to+\infty} n - l = +\infty$;
(A8) $\lim_{n\to+\infty} \frac{l \log(n-l)}{n-l} = 0$.

Convergence to the Bayes error

[6] (Rossi & Villa, *Neurocomputing*, 2006)

Theorem: universal consistency

Under the assumptions **(A1)-(A8)**,

$$L\Phi^n \xrightarrow{n\to+\infty} L^*,$$

where

- $L\Phi^n = \mathbb{P}(\Phi^n(X) \neq Y)$
- $L^* = \mathbb{P}(\Phi^*(X) \neq Y)$ with $\Phi^*(x) = \begin{cases} 1 & \text{if } \mathbb{P}(Y=1|X=x) > 1/2, \\ -1 & \text{otherwise.} \end{cases}$

Convergence to the Bayes error

[6] (Rossi & Villa, *Neurocomputing*, 2006)

Theorem: universal consistency

Under the assumptions **(A1)-(A8)**,

$$L\Phi^n \xrightarrow{n\to+\infty} L^*,$$

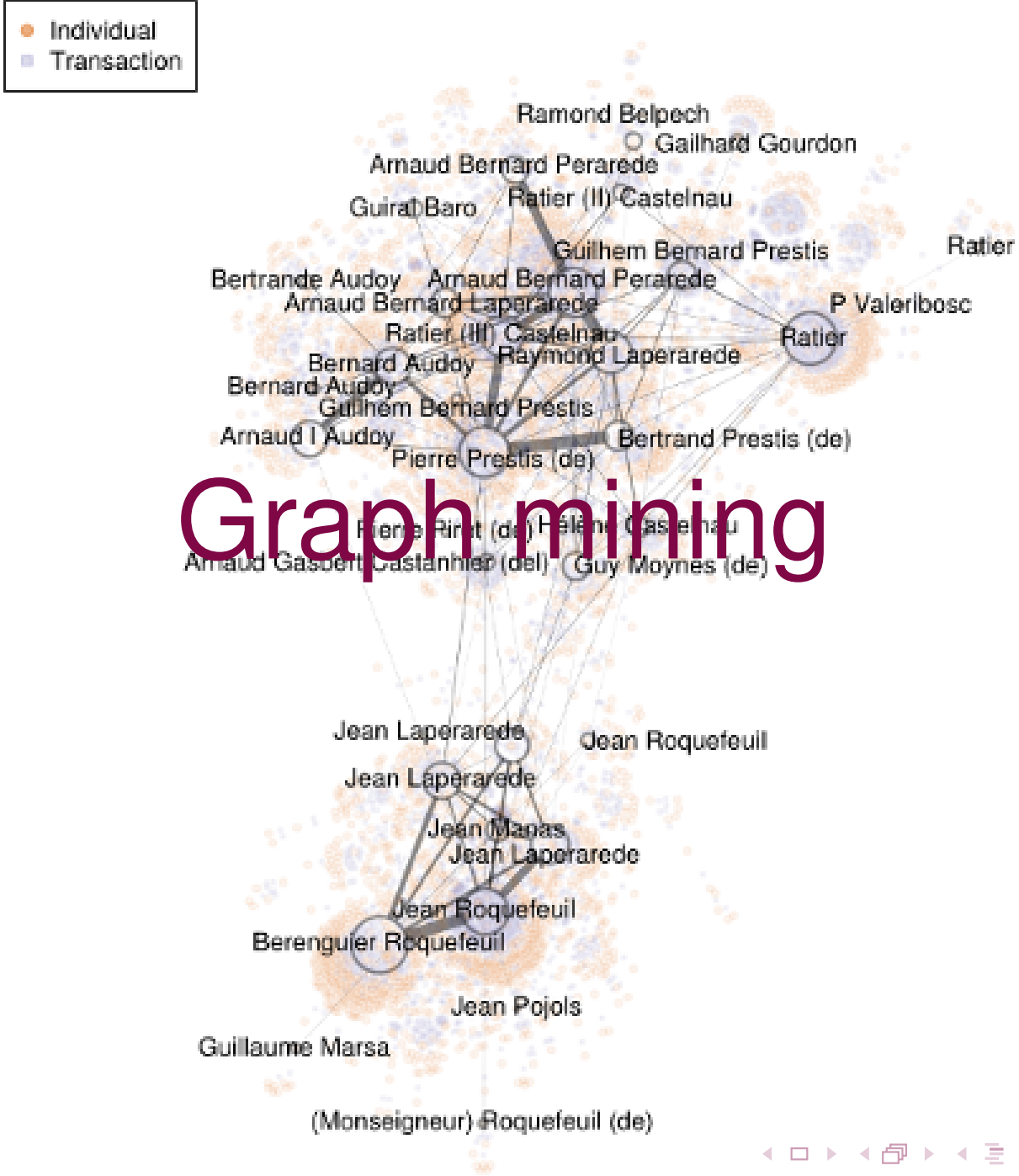
where

- $L\Phi^n = \mathbb{P}(\Phi^n(X) \neq Y)$
- $L^* = \mathbb{P}(\Phi^*(X) \neq Y)$ with $\Phi^*(x) = \begin{cases} 1 & \text{if } \mathbb{P}(Y=1|X=x) > 1/2, \\ -1 & \text{otherwise.} \end{cases}$

[8] (Rossi & Villa-Vialaneix, *Pattern Recognition Letter*, 2011) alternative kernel for:

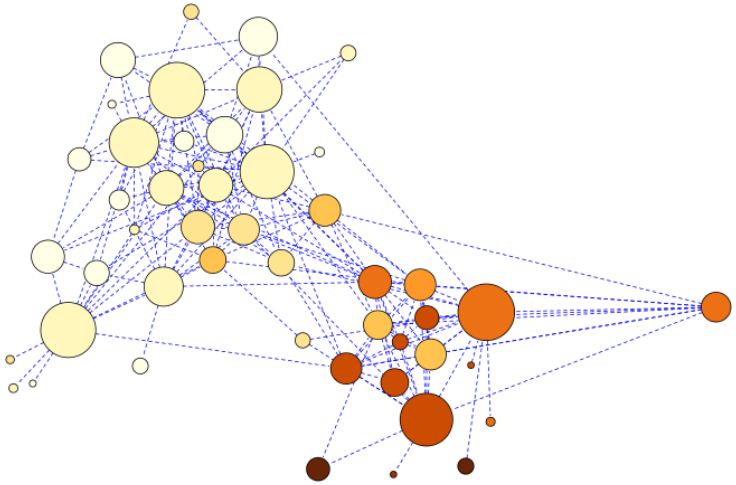
- digitized observations
- derivation preprocessing

with a general consistency result in regression and binary classification.



Graph mining

Framework of this section



A graph (network) $\mathcal{G} = (V, E, W)$ with

- n vertices $V = \{x_1, \ldots, x_n\}$;
- a set of edges, E, weighted by $W_{ij} = W_{ji} \geq 0$ ($W_{ii} = 0$).

Visualization: a tool for graph mining

Standard approach for graph visualization: force directed placement algorithms (FDP). Drawbacks:

- slow (impracticable for large graphs);
- based on aesthetic criteria rather than on interpretability:
  - *trend*: short and uniform length edges;
  - *negative consequence*: nodes with the largest degrees are grouped in the middle of the layout.

Visualization: a tool for graph mining

Standard approach for graph visualization: force directed placement algorithms (FDP). Drawbacks:

- slow (impracticable for large graphs);
- based on aesthetic criteria rather than on interpretability:
  - *trend*: short and uniform length edges;
  - *negative consequence*: nodes with the largest degrees are grouped in the middle of the layout.

A more natural way to explore a graph:

1. highlight the macroscopic structure: find “communities” and relations between them;
2. eventually focus on finer details in some communities.

Overview of my contributions

- topographic maps: combine clustering and visualization using a prior structure (*ie* a map):
  - batch kernel SOM [1] (Boulet, *et al.*, *Neurocomputing*, 2008) and on-line multiple relational SOM [4] (Olteanu & Villa-Vialaneix, *Neurocomputing*, 2015) (not restricted to graphs; can handle multiple graphs or labeled graphs)

Overview of my contributions

- topographic maps: combine clustering and visualization using a prior structure (*ie* a map):
  - batch kernel SOM [1] (Boulet, *et al.*, *Neurocomputing*, 2008) and on-line multiple relational SOM [4] (Olteanu & Villa-Vialaneix, *Neurocomputing*, 2015) (not restricted to graphs; can handle multiple graphs or labeled graphs)
  - maps based on an organized modularity criterion [7] (Rossi & Villa-Vialaneix, *Neurocomputing*, 2010)

Overview of my contributions

- topographic maps: combine clustering and visualization using a prior structure (*ie* a map):
  - batch kernel SOM [1] (Boulet, *et al.*, *Neurocomputing*, 2008) and on-line multiple relational SOM [4] (Olteanu & Villa-Vialaneix, *Neurocomputing*, 2015) (not restricted to graphs; can handle multiple graphs or labeled graphs)
  - maps based on an organized modularity criterion [7] (Rossi & Villa-Vialaneix, *Neurocomputing*, 2010)
- [9] (Rossi & Villa-Vialaneix, *Journal de la SFdS*, 2011) hierarchical approach (based on modularity maximization) for graph clustering and visualization;

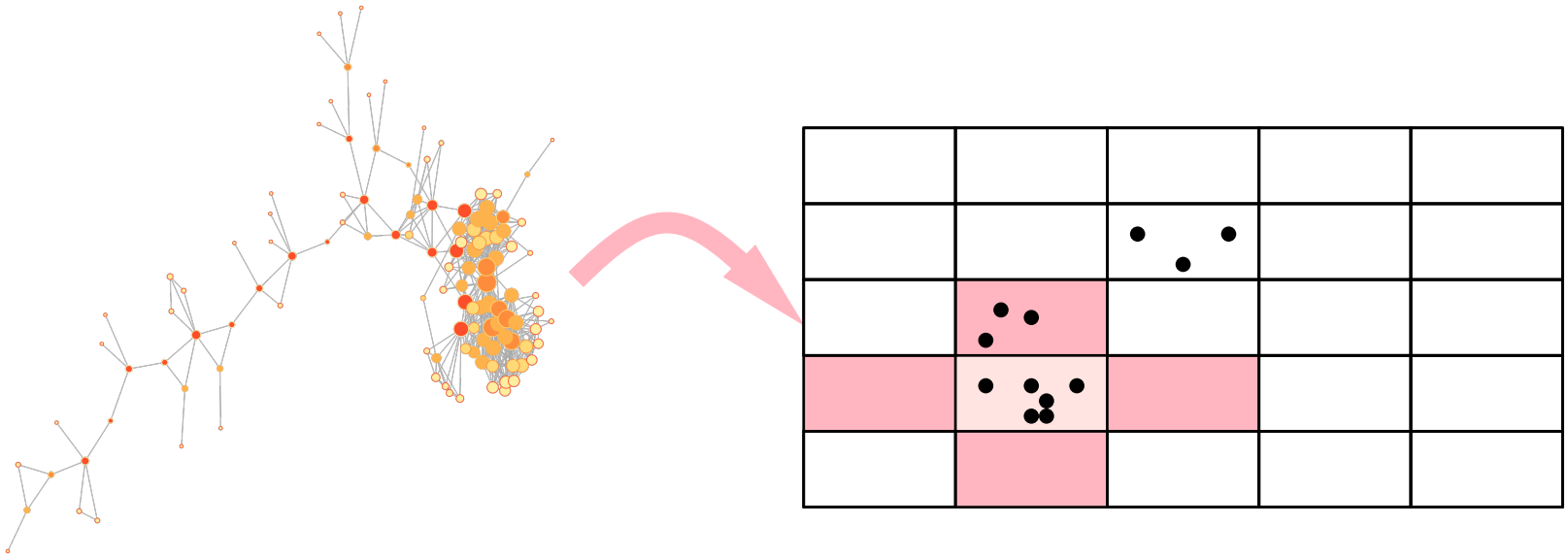
Overview of my contributions

- topographic maps: combine clustering and visualization using a prior structure (*ie* a map):
  - batch kernel SOM [1] (Boulet, *et al.*, *Neurocomputing*, 2008) and on-line multiple relational SOM [4] (Olteanu & Villa-Vialaneix, *Neurocomputing*, 2015) (not restricted to graphs; can handle multiple graphs or labeled graphs)
  - maps based on an organized modularity criterion [7] (Rossi & Villa-Vialaneix, *Neurocomputing*, 2010)
- [9] (Rossi & Villa-Vialaneix, *Journal de la SFdS*, 2011) hierarchical approach (based on modularity maximization) for graph clustering and visualization;
- applications
  - for mining a network extracted from a corpus of medieval documents [10] (Rossi, *et al.*, *Digital Medievalist*, 2013)

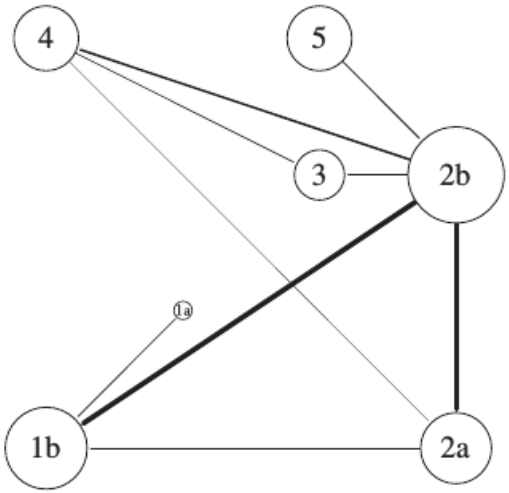
Overview of my contributions

- topographic maps: combine clustering and visualization using a prior structure (*ie* a map):
  - batch kernel SOM [1] (Boulet, *et al.*, *Neurocomputing*, 2008) and on-line multiple relational SOM [4] (Olteanu & Villa-Vialaneix, *Neurocomputing*, 2015) (not restricted to graphs; can handle multiple graphs or labeled graphs)
  - maps based on an organized modularity criterion [7] (Rossi & Villa-Vialaneix, *Neurocomputing*, 2010)
- [9] (Rossi & Villa-Vialaneix, *Journal de la SFdS*, 2011) hierarchical approach (based on modularity maximization) for graph clustering and visualization;
- applications
  - for mining a network extracted from a corpus of medieval documents [10] (Rossi, *et al.*, *Digital Medievalist*, 2013)
  - for co-expression network analysis [13] (Villa-Vialaneix, *et al.*, *PLoS ONE*, 2013) (clustering in co-expression network with ontology validation)

Using topographic maps to visualize a clustered graph



- vertices are clustered in U units
- units are arranged on a prior 2D-grid equipped with a topology
- prior positions of the units are used to produce a simplified representation



How to extend topographic maps to graph?

Generic approach using kernel/dissimilarity

[4] (Olteanu & Villa-Vialaneix, *Neurocomputing*, 2015), [1] (Boulet, *et al.*, *Neurocomputing*, 2008) graphs can be described by pairwise relations between nodes:

- with a kernel (ex: $K^\beta = e^{-\beta L}$, with L the graph Laplacian, heat kernel)
- with a dissimilarity (ex: shortest path length)

$\Rightarrow$ extension of SOM to kernel/dissimilarity data using prototypes of the form $p_u = \sum_{i=1}^{n} \gamma_{ui} x_i$.

How to extend topographic maps to graph?

Generic approach using kernel/dissimilarity

[4] (Olteanu & Villa-Vialaneix, *Neurocomputing*, 2015), [1] (Boulet, *et al.*, *Neurocomputing*, 2008) graphs can be described by pairwise relations between nodes:

- with a kernel (ex: $K^\beta = e^{-\beta L}$, with L the graph Laplacian, heat kernel)
- with a dissimilarity (ex: shortest path length)

$\Rightarrow$ extension of SOM to kernel/dissimilarity data using prototypes of the form $p_u = \sum_{i=1}^{n} \gamma_{ui} x_i$.

Graph specific approach

[7] (Rossi & Villa-Vialaneix, *Neurocomputing*, 2010) Extension of the modularity quality criterion to a criterion taking into account the prior topology

$$O = \frac{1}{2m} \sum_{ijk} M_{ik} S_{kl} M_{jk} B_{ij}$$

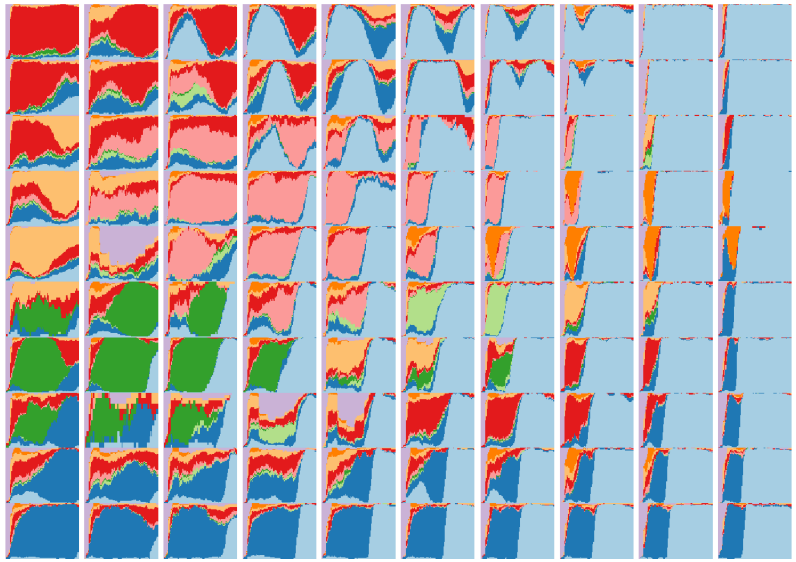
$B = \left(W_{ij} - \frac{d_i d_j}{2m}\right)$ modularity matrix ($d_i = \sum_j W_{ij}$ and $m = \frac{\sum_i d_i}{2}$); $M_{ik} = \mathbf{1}_{x_i \in k}$; S encodes the topology of the prior map (similarity)

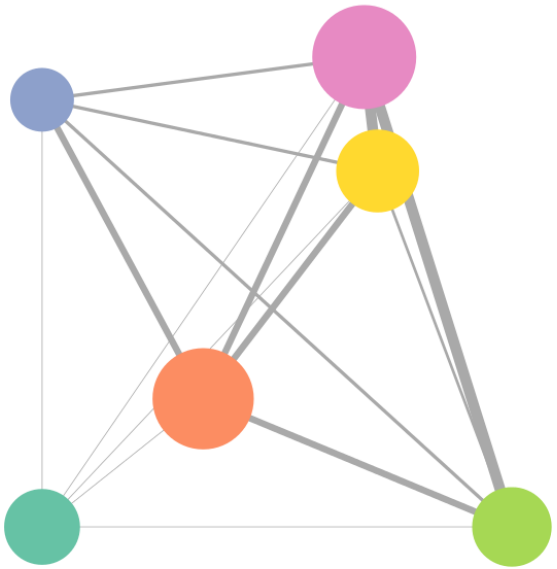
Practical aspects of kernel/relational SOM

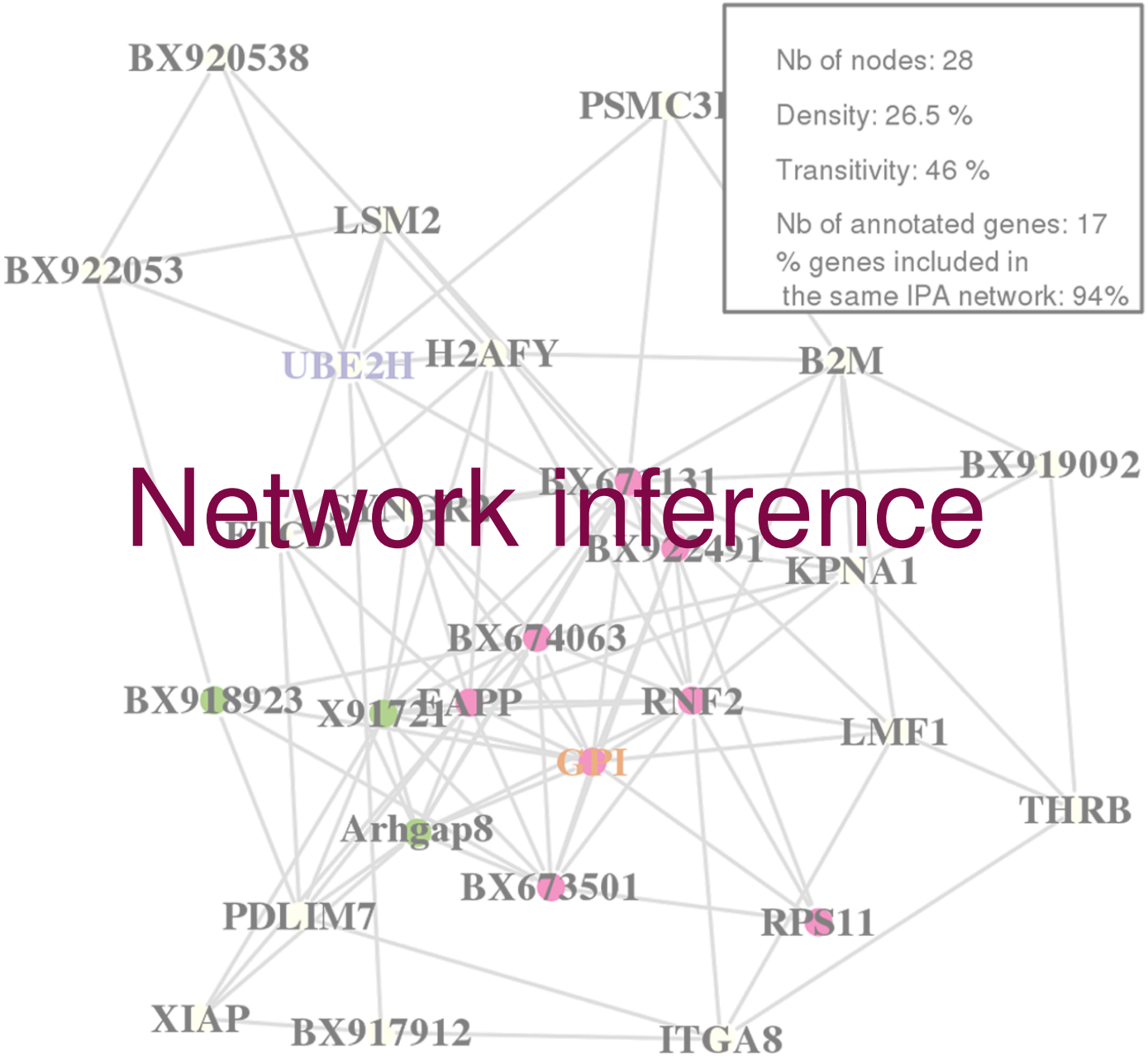
- this approach (and others) is implemented in the R package **SOMbrero** provided with a WUI based on **shiny**;
- it can also handle data described by multiple dissimilarities: the combination is optimized by including a gradient descent-like step in the relational SOM algorithm;

Practical aspects of kernel/relational SOM

- this approach (and others) is implemented in the R package **SOMbrero** provided with a WUI based on **shiny**;
- it can also handle data described by multiple dissimilarities: the combination is optimized by including a gradient descent-like step in the relational SOM algorithm;
- it has been applied to graphs, labeled graphs and to other non vectorial data (school-to-work trajectories)







Network inference

Network inference

Data: large scale gene expression data

individuals
$n \simeq 30/50$

$$\underbrace{\left\{ X = \begin{pmatrix} . & . & . & . & . & . \\ . & . & X_i^j & . & . & . \\ . & . & . & . & . & . \end{pmatrix} \right.}_{\text{variables (genes expression), } p \simeq 10^4}$$

What we want to obtain: a graph/network with

- nodes: (selected) genes;
- edges: strong links between gene expressions.

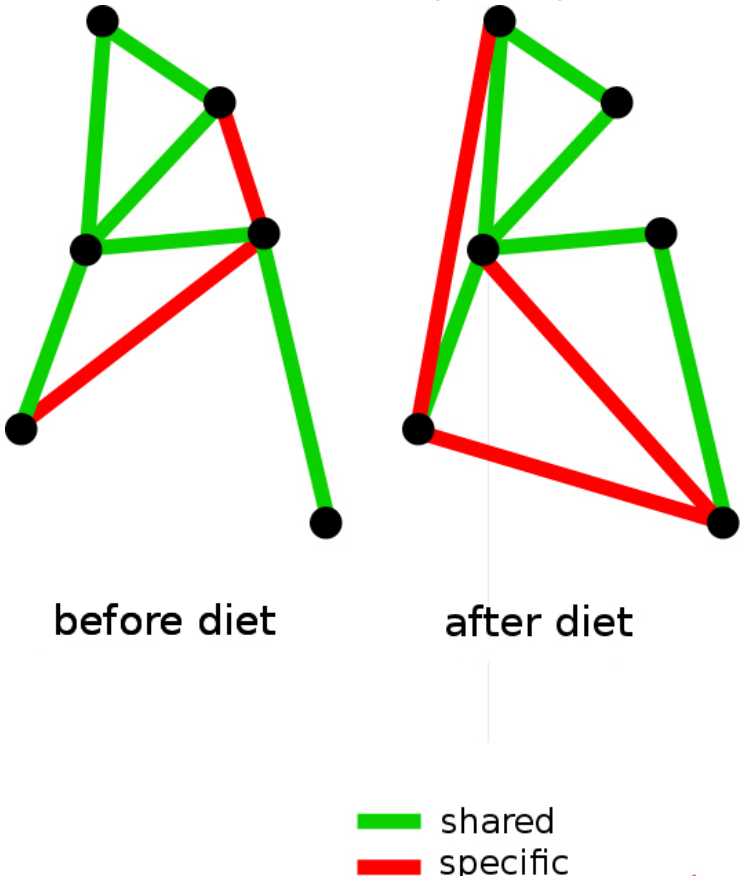
Overview of my contributions

- application of network inference for understanding the determinant of human adipose tissue gene expression [11] (Viguerie, *et al.*, *PLoS Genetics*, 2012)

- network inference with multiple samples [14] (Villa-Vialaneix, *et al.*, *Quality Technology and Quantitative Management*, 2014)

Motivation for multiple networks inference

Pan-European project Diogenes[1] (with Nathalie Viguerie, INSERM): gene expressions (lipid tissues) from 204 obese women before and after a low-calorie diet (LCD).



- Assumption: A common functioning exists regardless the condition;
- Which genes are linked independently from/depending on the condition?

[1] http://www.diogenes-eu.org

Consensus LASSO

GGM with L^1 penalty: one condition

$(X_i)_{i=1,\dots,n}$ are i.i.d. Gaussian random variables $\mathcal{N}(0,\Sigma)$

$$X^j = \beta_j^T X^{-j} + \epsilon \quad ; \quad \arg\min_{(\beta_{jj'})_{j'}} \sum_{i=1}^{n} \left(X_{ij} - \beta_j^T X_i^{-j}\right)^2 + \lambda \|\beta_j\|_{L^1}$$

$$j \longleftrightarrow j' \text{(genes } j \text{ and } j' \text{ are linked)} \Leftrightarrow \beta_{jj'} \neq 0$$

Consensus LASSO: multiple conditions

$$\frac{1}{2}\beta_j^T \widehat{\Sigma}_{\setminus j \setminus j} \beta_j + \beta_j^T \widehat{\Sigma}_{j \setminus j} + \lambda \|\beta_j\|_{L^1} + \mu \sum_c w_c \|\beta_j^c - \beta_j^{\text{cons}}\|_{L^2}^2$$

with $\widehat{\Sigma}_{\setminus j \setminus j}$: block diagonal matrix $\mathbb{D}\text{iag}\left(\widehat{\Sigma}^1_{\setminus j \setminus j}, \dots, \widehat{\Sigma}^k_{\setminus j \setminus j}\right)$ and similarly for $\widehat{\Sigma}_{j \setminus j}$:

- w_c: real number used to weight the conditions;
- μ regularization parameter;
- β_i^{cons} whatever you want...?

Choice of a consensus

Case 1: a priori consensus

Using a fixed β_j^{cons}, the optimization problem is equivalent to minimizing the p following standard quadratic problem in $\mathbb{R}^{k(p-1)}$ with L_1-penalty:

$$\frac{1}{2}\beta_j^T B^1(\mu)\beta_j + \beta_j^T B^2(\mu) + \lambda||\beta_j||_{L^1},$$

Case 2: learn the consensus

Using $\beta_j^{\text{cons}} = \sum_{c=1}^{k} \frac{n_c}{n}\beta_j^c$, the optimization problem is equivalent to minimizing the following standard quadratic problem with L_1-penalty:

$$\frac{1}{2}\beta_j^T S_j(\mu)\beta_j + \beta_j^T \widehat{\Sigma}_{j\backslash j} + \lambda||\beta_j||_{L^1}$$

Optimization by active set, combined with bootstrap approach (BOLASSO type). R package **therese**.

Future work

Research project: data mining and integration with SOM

On-going issues

- stabilize results, improve quality, speed-up training: aggregation, boosting...
- improve interpretability in a multi-kernel/dissimilarity context: visualization, prototype sparsity...
- targeted applications: multi-'omics integration and exploration; ncRNA typology

Collaborations

- methodological aspects: Jérôme Mariette (PhD, MIAT, INRA), Madalina Olteanu (SAMM, Université Paris 1)
- application to multi-'omics data: Nathalie Viguerie (INSERM, Diogenes project)
- application to ncRNA: Christine Gaspin (MIAT, INRA)

Research project: data mining and integration using graphical approaches

On-going issues

- integrating multi-'omics data in network inference and mining, possibly with different numbers of observations
- taking temporal aspects into account in network inference/clustering

Collaborations

- methodological aspects: Valérie Sautron (PhD, GenPhySE, INRA)
- applications: projects SusOStress & PigHeat (GenPhySE, INRA) on systems genetics in pigs (stress & heat resistance)
- application: Nathalie Viguerie (INSERM, Diogenes project) & Ignacio Gonzáles (MIAT, INRA)... *submitted article*

Thank you for your attention...

... questions?

R. Boulet, B. Jouve, F. Rossi, and N. Villa.

Batch kernel SOM and related Laplacian methods for social network analysis.
Neurocomputing, 71(7-9):1257–1273, 2008.

L. Ferré and N. Villa.

Multi-layer perceptron with functional inputs: an inverse regression approach.
Scandinavian Journal of Statistics, 33(4):807–823, 2006.

N. Hernández, R.J. Biscay, N. Villa-Vialaneix, and I. Talavera.

A non parametric approach for calibration with functional data.
Statistica Sinica, 2014.
Forthcoming.

M. Olteanu and N. Villa-Vialaneix.

On-line relational and multiple relational SOM.
Neurocomputing, 147:15–30, 2015.
Forthcoming.

F. Rohart, A. Paris, B. Laurent, C. Canlet, J. Molina, M.J. Mercat, T. Tribout, N. Muller, N. Iannuccelli, N. Villa-Vialaneix, L. Liaubet, D. Milan, and M. San Cristobal.

Phenotypic prediction based on metabolomic data on the growing pig from three main European breeds.
Journal of Animal Science, 90(12), 2012.

F. Rossi and N. Villa.

Support vector machine for functional data classification.
Neurocomputing, 69(7-9):730–742, 2006.

F. Rossi and N. Villa-Vialaneix.

Optimizing an organized modularity measure for topographic graph clustering: a deterministic annealing approach.
Neurocomputing, 73(7-9):1142–1163, 2010.

F. Rossi and N. Villa-Vialaneix.

Consistency of functional learning methods based on derivatives.
Pattern Recognition Letters, 32(8):1197–1209, 2011.

F. Rossi and N. Villa-Vialaneix.
Représentation d'un grand réseau à partir d'une classification hiérarchique de ses sommets.
Journal de la Société Française de Statistique, 152(3):34–65, 2011.

F. Rossi, N. Villa-Vialaneix, and F. Hautefeuille.
Exploration of a large database of French notarial acts with social network methods.
Digital Medievalist, 9, 2013.

N. Viguerie, E. Montastier, J.J. Maoret, B. Roussel, M. Combes, C. Valle, N. Villa-Vialaneix, J.S. Iacovoni, J.A. Martinez, C. Holst, A. Astrup, H. Vidal, K. Clément, J. Hager, W.H.M. Saris, and D. Langin.
Determinants of human adipose tissue gene expression: impact of diet, sex, metabolic status and cis genetic regulation.
PLoS Genetics, 8(9):e1002959, 2012.

N. Villa-Vialaneix, N. Hernandez, A. Paris, C. Domange, N. Priymenko, and P. Besse.
On combining wavelets expansion and sparse linear models for regression on metabolomic data and biomarker selection.
Communications in Statistics - Simulation and Computation, 2014.
Forthcoming.

N. Villa-Vialaneix, L. Liaubet, T. Laurent, P. Cherel, A. Gamot, and M. San Cristobal.
The structure of a gene co-expression network reveals biological functions underlying eQTLs.
PLoS ONE, 8(4):e60045, 2013.

N. Villa-Vialaneix, M. Vignes, N. Viguerie, and M. San Cristobal.
Inferring networks from multiple samples with concensus LASSO.
Quality Technology and Quantitative Management, 11(1):39–60, 2014.