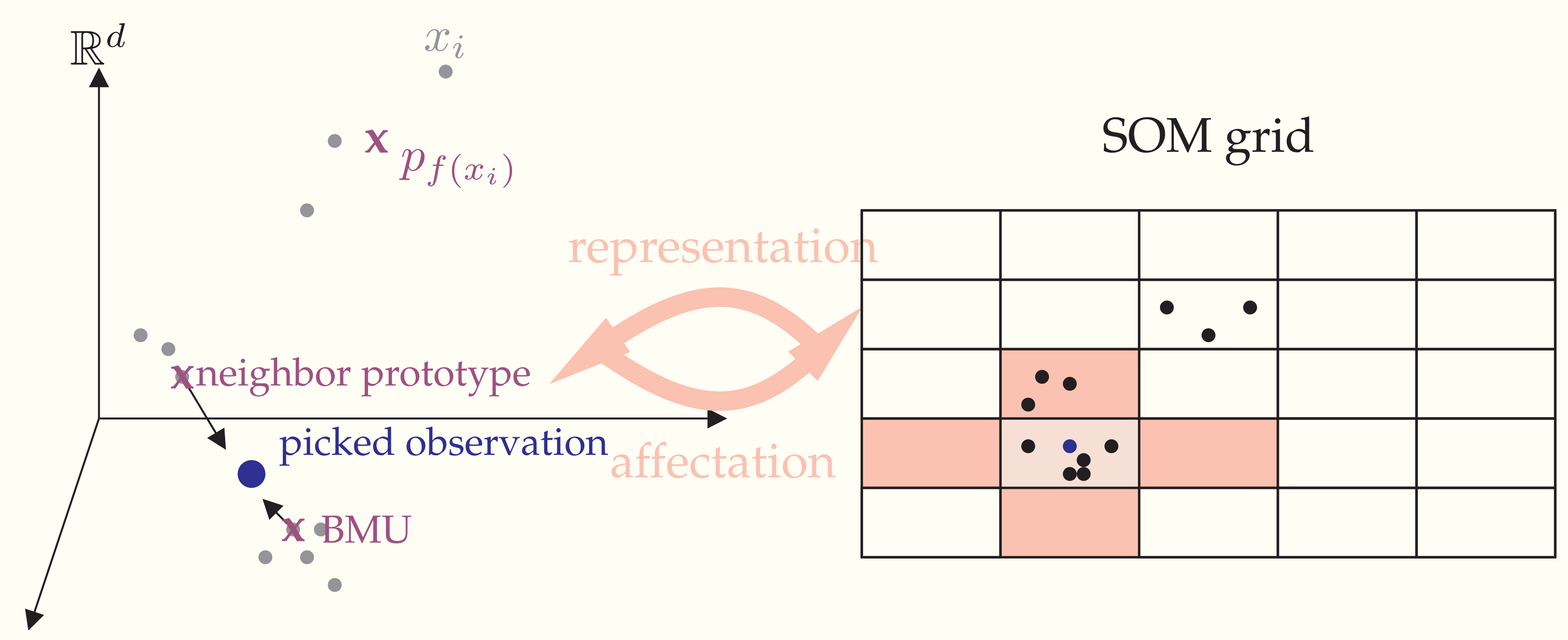
Multiple dissimilarity SOM for clustering and visualizing graphs with node and edge attributes

Nathalie Villa-Vialaneix, INRA, UR 0875 MIAT &
Madalina Olteanu, Université Paris 1, SAMM, France
nathalie.villa@toulouse.inra.fr - madalina.olteanu@univ-paris1.fr

Standard SOM for multidimensional data [4]



Cluster data $(x_i)_{i=1,\dots,n} \in \mathbb{R}^d$ on a grid made of U units and equipped with a distance between units, $d(u,u')$
Units have representers called **prototypes** $(p_u)_u \in \mathbb{R}^d$
Clustering $f : \mathbb{R}^d \rightarrow \{1,\dots,U\}$ and prototypes are **updated iteratively in order to preserve the topology** of the input space

1. *affectation step*: pick a data x_i at random and **find the best matching unit**: $f(x_i) := \arg\min_{u=1,\dots,U} \|x_i - p_u\|^2$
2. *representation step*: update the BMU and its neighbors' prototypes with a **stochastic gradient descent like scheme**: $p_u \leftarrow p_u + \mu H(d(f(x_i),u))\,(x_i - p_u)$

Extension of SOM to data described by a kernel / a dissimilarity

Data: $(x_i)_{i=1,\dots,n} \in \mathcal{G}$ described by pairwise relations with a **kernel** $\mathbf{K} \in \mathcal{M}_{n\times n}$ or a **dissimilarity** $\mathbf{\Delta} \in \mathcal{M}_{n\times n}$ ⇒ **stochastic kernel SOM** [2] and **stochastic relational SOM** [6] implemented in **SOMbrero** (R package)
Prototypes: linear convex combination of the data $p_u = \sum_{i=1}^n \beta_{ui}\phi(x_i)$ (only $(\beta_{ui})_{u=1,\dots,U,i=1,\dots,n}$ are trained. ϕ is implicitely defined by the kernel/dissimilarity)
Updated steps:

1. *affectation step* writes $f(x_i) = \arg\min_u \beta_u^T\mathbf{K}\beta_u - 2\beta_u^T\mathbf{K}_i$ (kernel SOM) or $f(x_i) = \arg\min_u \mathbf{\Delta}_i\beta_u - \frac{1}{2}\beta_u^T\mathbf{\Delta}\beta_u$ (relational SOM)
2. *representation step* writes $\beta_u \leftarrow \beta_u + \mu H(d(f(x_i),u))\,(\mathbf{1}_i - \beta_u)$

Mixing multiple kernels

Data are described by **several pairwise relations** (kernels/dissimilarities) $\mathbf{K}^1, \dots, \mathbf{K}^D$ ⇒ **Multiple kernel**: $\mathbf{K} = \sum_{k=1}^D \alpha_k\mathbf{K}^k$ with $\alpha_k \geq 0$ and $\sum_k \alpha_k = 1$

How to choose $(\alpha_k)_k$?

Similarly to [8], add a stochastic gradient descent step in SOM training:

3. *multiple kernel tuning step* $\alpha_k \leftarrow \alpha_k + \nu\mathcal{D}_{ki}$ with $\mathcal{D}_{ki} = \sum_u H(d(f(x_i),u))\left(\mathbf{K}^k(x_i,x_i) - 2\beta_u^T\mathbf{K}_i^k + \beta_u^T\mathbf{K}^k\beta_u\right)$ (+ reduction & projection to ensure the α_k remain positive and sum to 1)

see [5] (multiple kernels) or [6] (multiple dissimilarities)

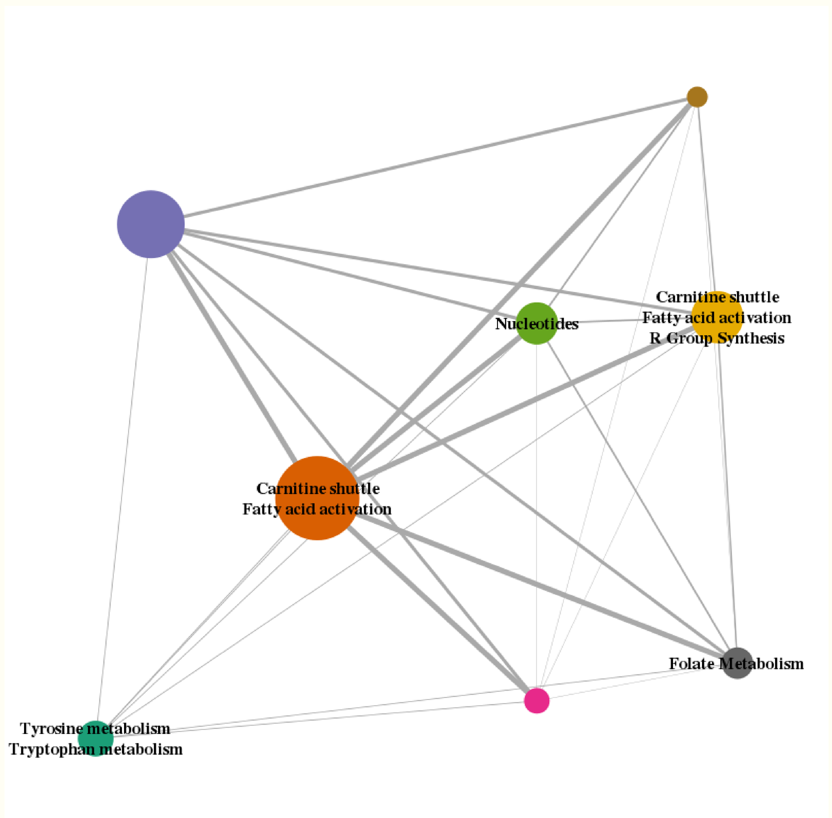
Applications to graphs

Type of data that can be handled:

- **graphs with node attributes** (a kernel for the graph structure - *e.g.*, Laplacian based kernels; kernels for each of the attributes)
- **graphs with different types of edge** (a kernel for each subgraph defined by an edge type)
- both... and can also be used to **combine different kernels with different parameters**

Useful for:

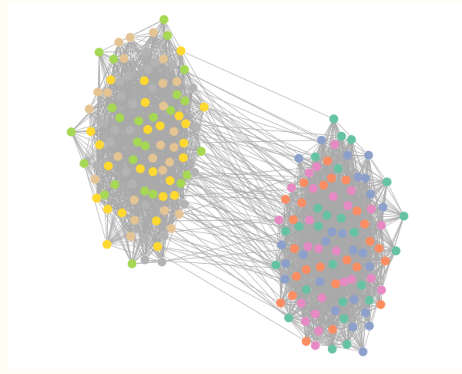
- uncover **communities**...
- ... and **visualize the relations between communnities**
- as shown in [7], the result of the SOM can be combined with clustering of the prototypes to obtain a simplified representation of a graph

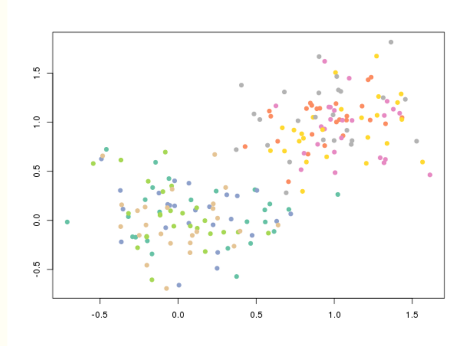


human metabolic network from the BiGG database
http://bigg.ucsd.edu

An example (on simulated data)

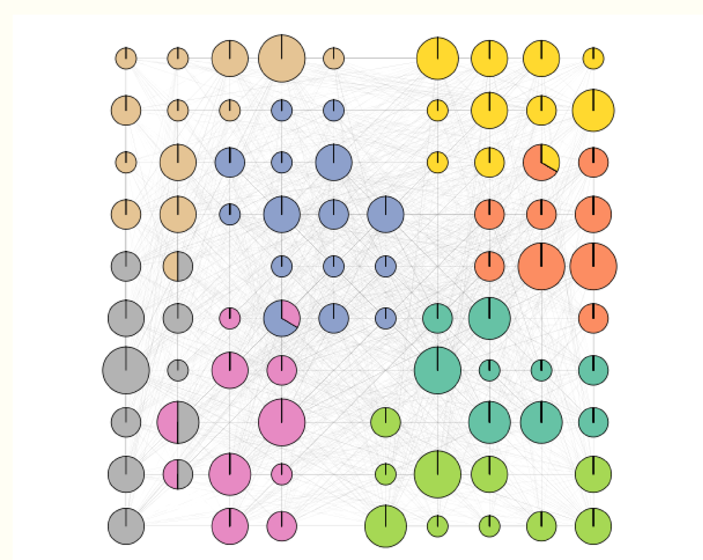
Simulation of 8 groups of observations made from:





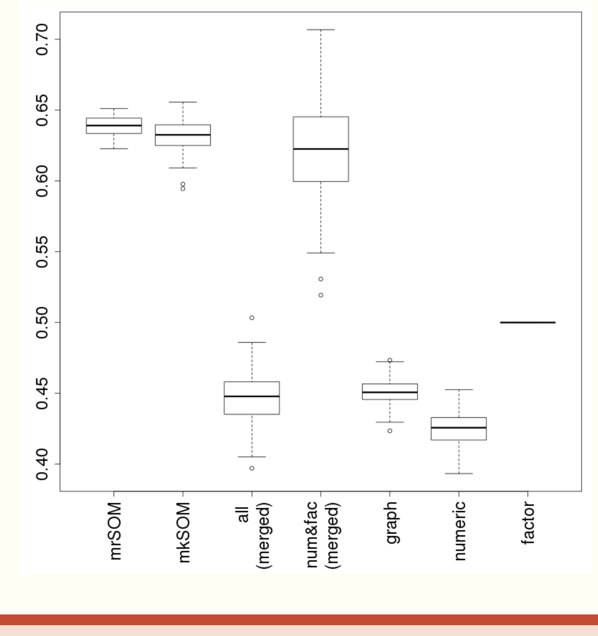
- **unweighted graph** (planted 3-partition graph; see [1]) with two dense groups of nodes: commute time kernel (L^+ with L the Laplacian; see [3])
- **nodes are labelled with numeric data** from a 2D Gaussian mixture: Gaussian kernel;
- ... and **nodes are labelled with a factor** (2-levels): Gaussian kernel on 0/1 recoding.

Resulting Map



Comparison

(100 datasets - NMI with true groups)



References

[1] A. Condon & R.M. Karp (2001) Algorithms for graph partitioning on the planted partition model. *Random Structures and Algorithms*, **18**(2), 116-140.
[2] D. mac Donald & C. Fyfe (2000) The kernel self organising map. In: *Proceedings of 4th International Conference on Knowledge-Based Intelligence Engineering Systems and Applied Technologies*, 317-320.
[3] F. Fouss, A. Pirotte, J.M. Renders & M. Saerens (2007) Random-walk computation of similarities between nodes of a graph, with application to collaborative recommendation. *IEEE Transactions on Knowledge and Data Engineering*, **19**(3), 355-369.
[4] T. Kohonen (1995) *Self-Organizing Maps*, Springer.
[5] M. Olteanu & N. Villa-Vialaneix (2013) Multiple kernel self-organizing maps. In: *Proceedings of XXIst European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN 2013)*, M. Verleysen (ed), 83-88, Bruges, Belgium.
[6] M. Olteanu & N. Villa-Vialaneix (2015) On-line relational and multiple relational SOM. *Neurocomputing*, **147**, 15-30.
[7] M. Olteanu & N. Villa-Vialaneix (2015) Using **SOMbrero** for clustering and visualizing graphs. *Journal de la Société Française de Statistique. Forthcoming*.
[8] A. Rakotomamonjy, F.R. Bach, S. Canu & Y. Grandvalet (2008) SimpleMKL. *Journal of Machine Learning Research*, **9**, 2491-2521.