Noyaux sur des graphes

Cours Master 2005/06

Jean-Philippe Vert

Jean-Philippe.Vert@mines.org

Plan

- Motivation
- Approche par régularisation
- Noyau de diffusion
- Généralisation : analyse harmonique sur un graphe
- Application

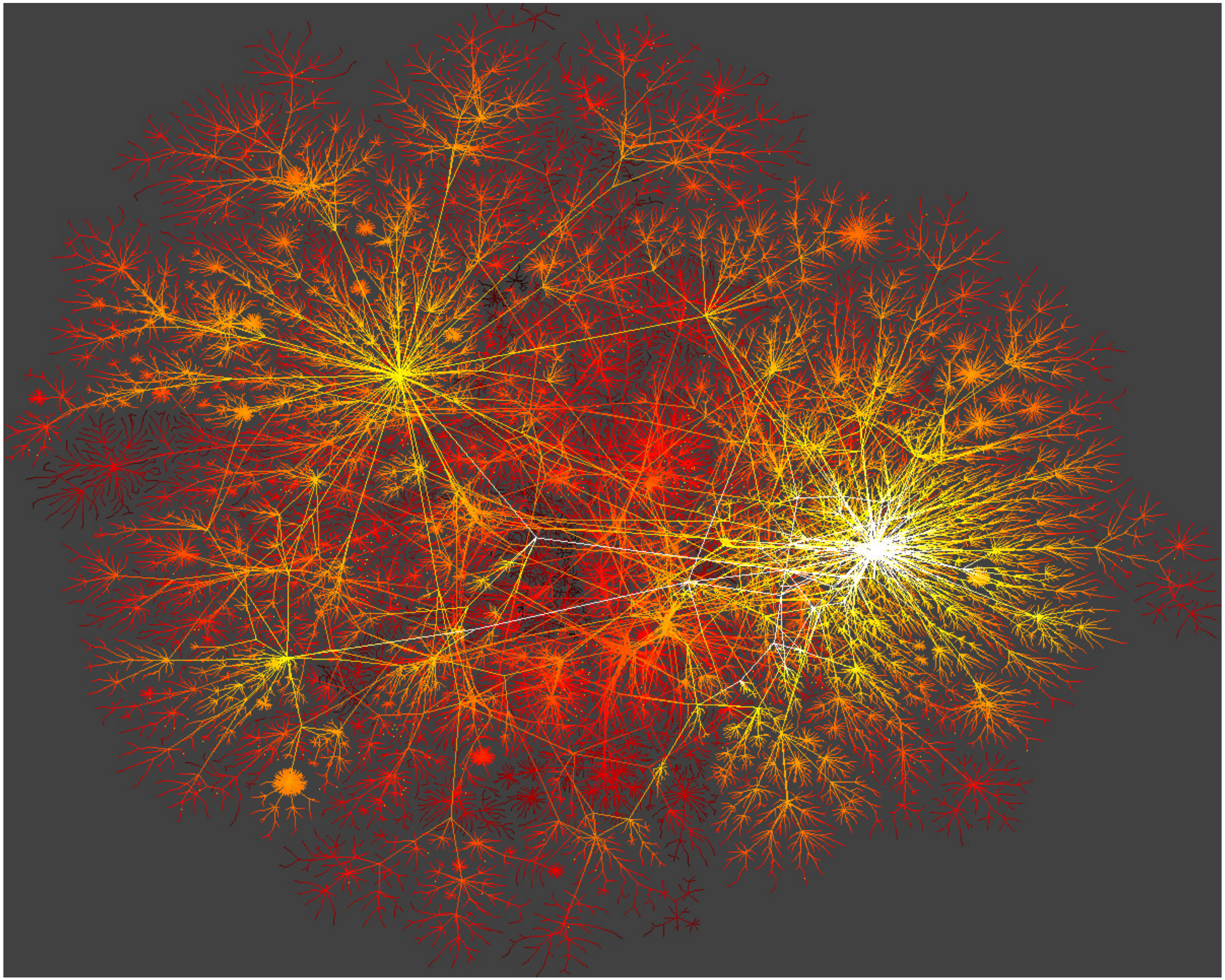
Motivation

Pourquoi des graphes?

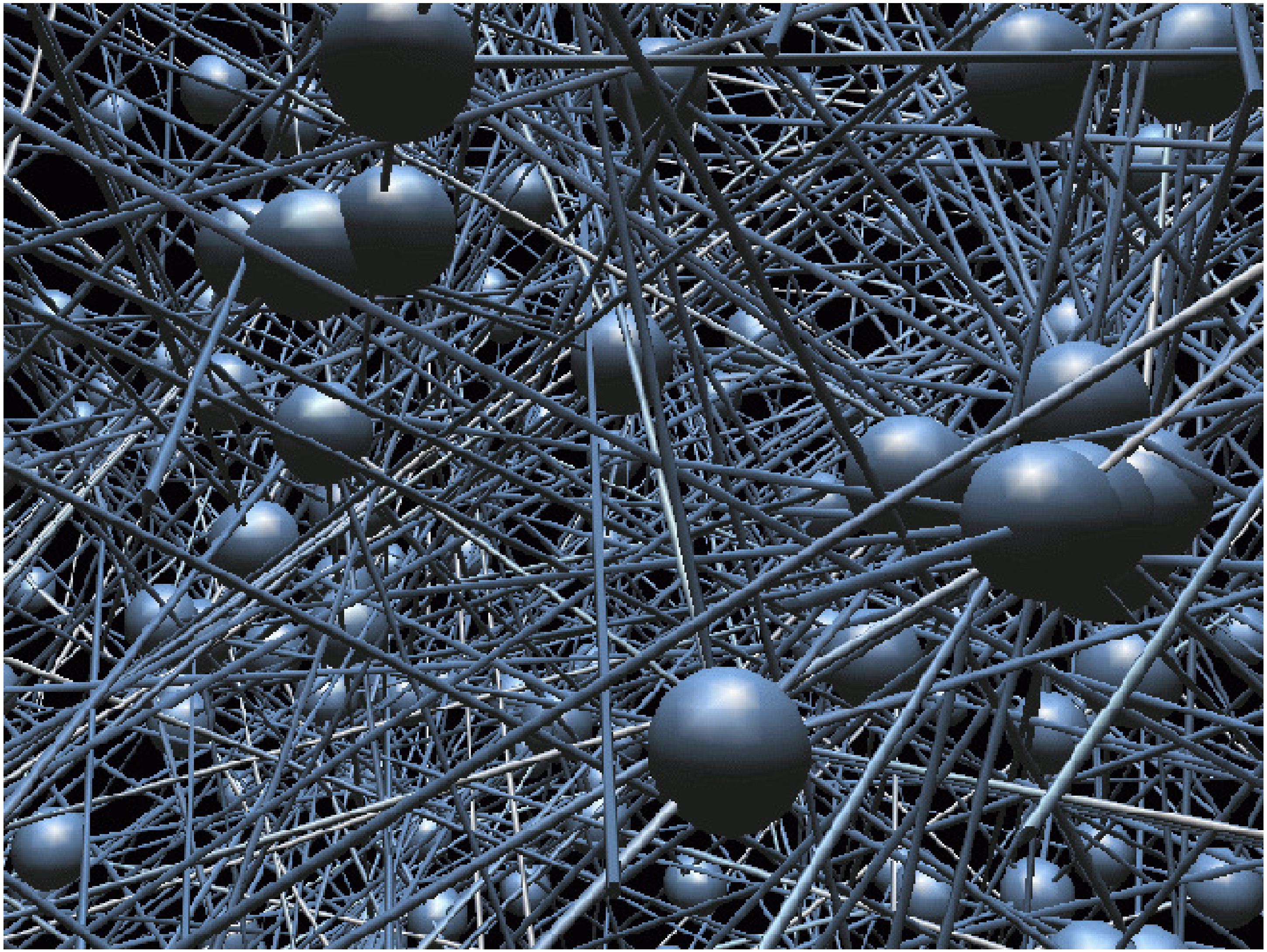
De nombreuses données peuvent se représenter comme les *noeuds d'un graphe*:

- par *nature*,
- par *discrétisation/échantillonage* d'un espace continu,
- par *nécessité*

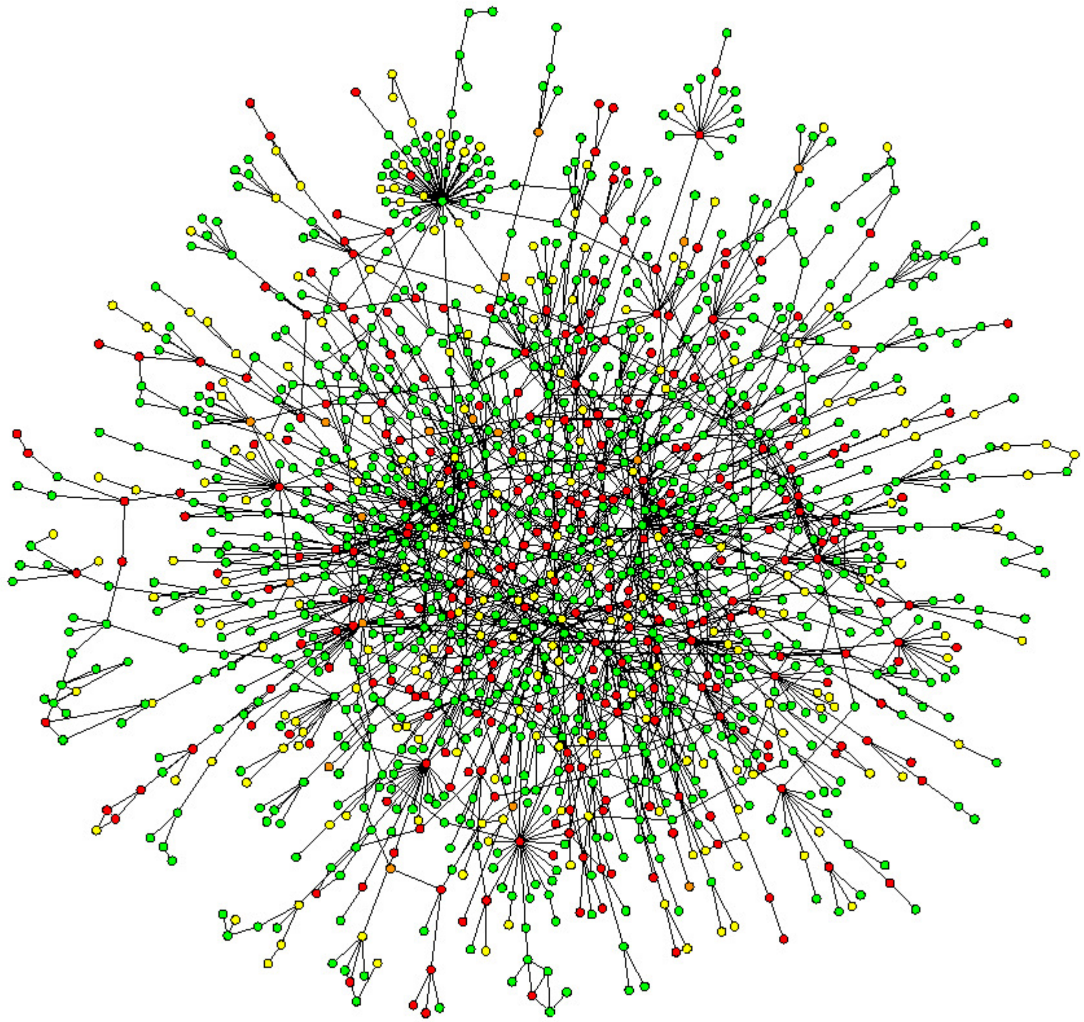
Internet (par nature)



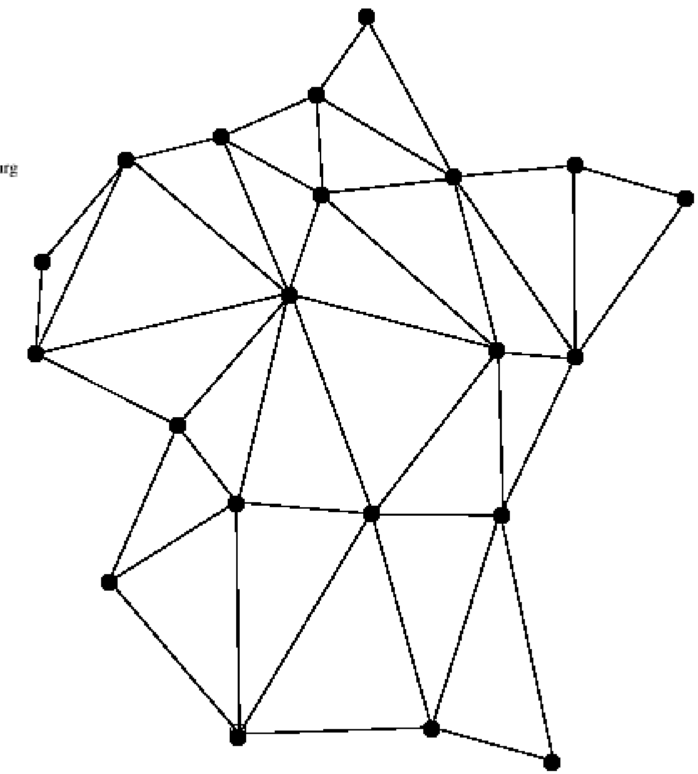
Réseau social (par nature)



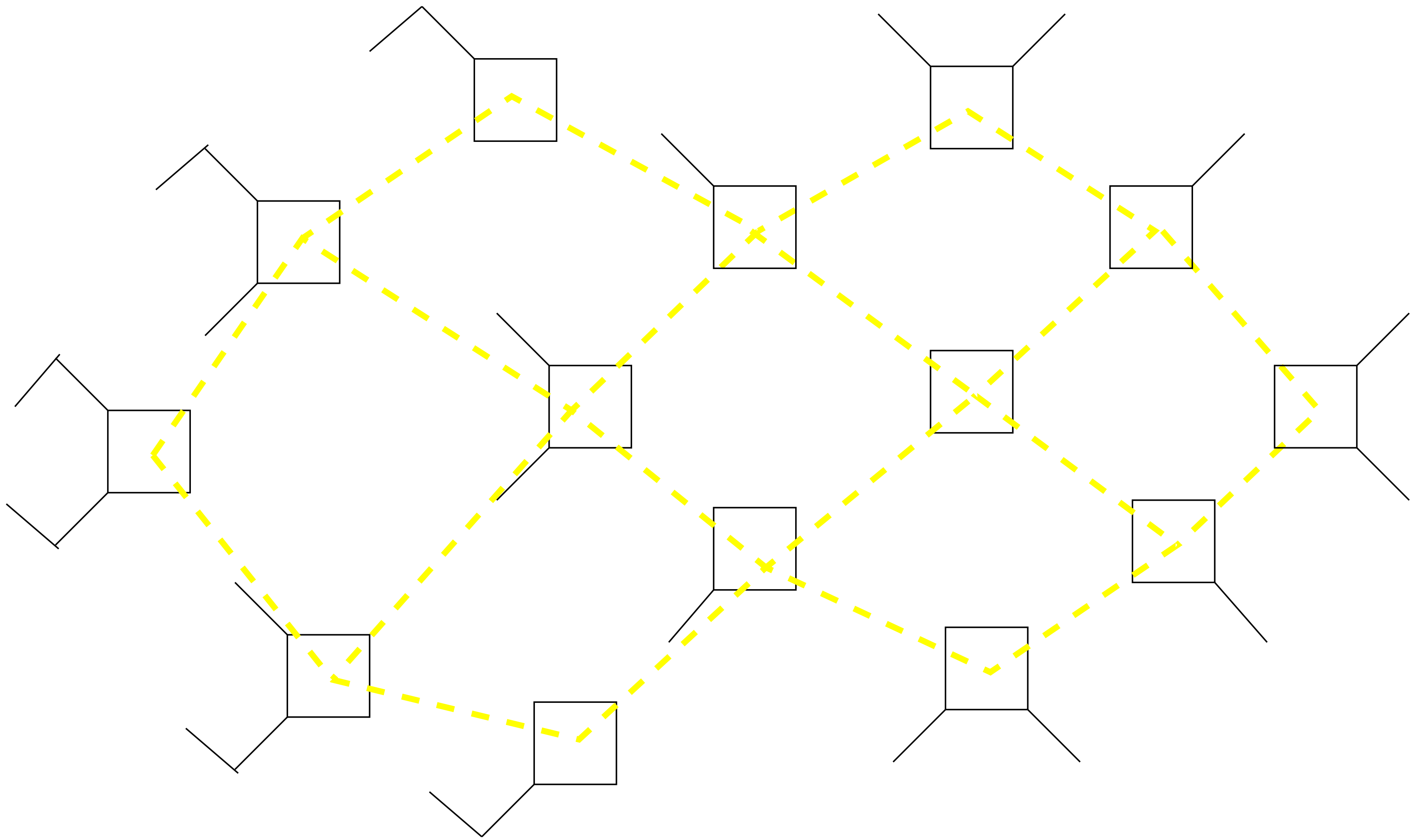
Interaction des protéines (par nature)



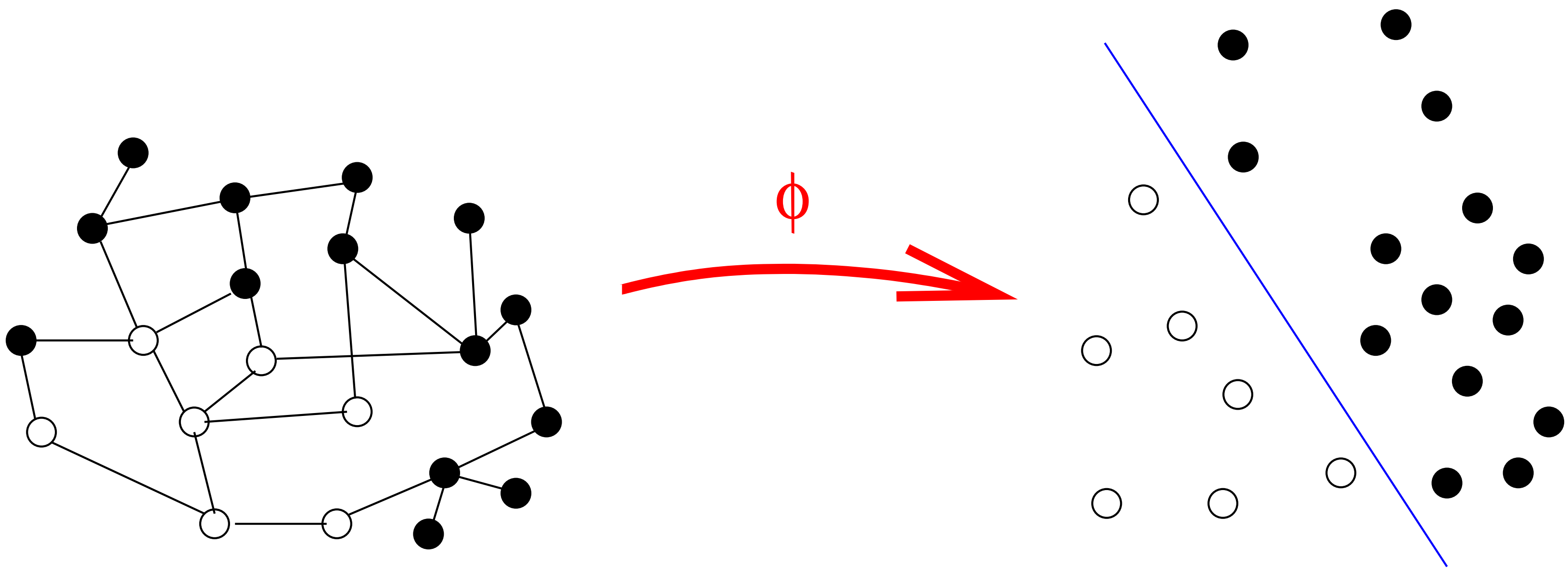
Régions (par discrétisation)



Molécules (par nécessité)



Noyau sur un graphe?



Il faut un *noyau* $K(\mathbf{x}, \mathbf{x}')$ *entre noeuds du graphe*

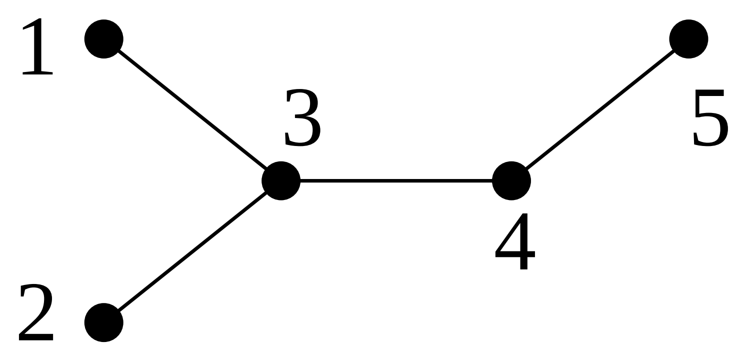
Notations

- $\mathcal{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_m)$ est fini.
- Pour $\mathbf{x}, \mathbf{x}' \in \mathcal{X}$, on note $\mathbf{x} \sim \mathbf{x}'$ pour dénoter une arète
- On suppose qu'il n'y a *pas de boucle* $\mathbf{x} \sim \mathbf{x}$, et qu'il n'y a *qu'une composante connexe*.
- La *matrice d'adjacence* est $A \in \mathbb{R}^{m \times m}$:

$$A_{i,j} = \begin{cases} 1 & \text{si } i \sim j, \\ 0 & \text{sinon.} \end{cases}$$

- D la matrice diagonale où $D_{i,i}$ est le nombre de voisins de $\mathbf{x}_i$ ($D_{i,i} = \sum_{i=1}^{m} A_{i,j}$).

Exemple



$$A = \begin{pmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix}, \qquad D = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

Remarques générales

- $\mathcal{X}$ étant fini, *n'importe quelle matrice symétrique et semi-définie positive K* définit un n.d.p. valide
- Comment "traduire" la topologie du graphe dans un noyau?
  - *Approche géométrique:* $K_{i,j}$ doit être "grand" quand $\mathbf{x}_i$ et $\mathbf{x}_j$ sont "proches" sur le graphe?
  - *Approche fonctionnelle:* $\| f \|_K$ est "petit" quand f est "régulière" sur le graphe?
  - *Lien continu/discret:* Quel est par exemple l'équivalent du noyau Gaussien sur un graphe?

Approche géométrique

- Rappel : pour $\mathcal{X} = \mathbb{R}^n$, le noyau Gaussien est:

$$K(\mathbf{x}, \mathbf{x}') = \exp\left(-d(\mathbf{x}, \mathbf{x}')^2 / 2\sigma^2\right),$$

  où $d(\mathbf{x}, \mathbf{x}')$ est la *distance Euclidienne*.
- Si $\mathcal{X}$ est un *graphe*, soit $d(\mathbf{x}, \mathbf{x}')$ la *longueur du plus court chemin qui relie $\mathbf{x}$ et $\mathbf{x}'$*.
- *Problème:* $\exp\left(-d(\mathbf{x}, \mathbf{x}')^2 / 2\sigma^2\right)$ n'est en général *pas d.p.*
- *Gros problème:* pas de critère simple pour vérifier si $K(\mathbf{x}, \mathbf{x}') = \phi(d(\mathbf{x}, \mathbf{x}'))$ est d.p. ou non...

Approche par régularisation

Approche par régularisation

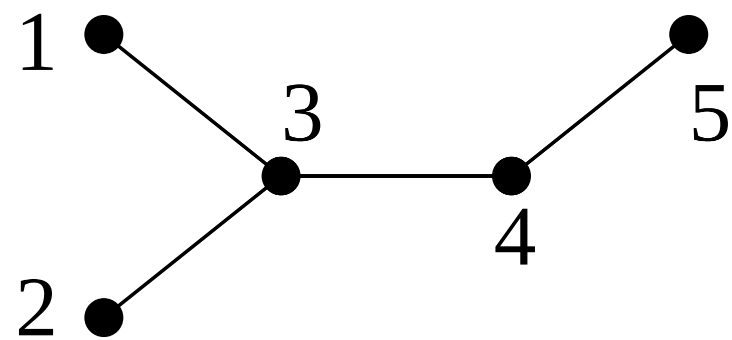
Pour une fonction $f = (f(\mathbf{x}_1), \ldots, f(\mathbf{x}_m)) \in \mathbb{R}^m$, on peut quantifier sa *régularité* par une fonctionnelle qui va définir un rkhs:

Théorème 1 *L'ensemble* $\mathcal{H} = \left\{ f \in \mathbb{R}^m : \sum_{i=1}^m f_i = 0 \right\}$ *muni de la norme:*

$$\Omega(f) = \sum_{i \sim j} \left(f(\mathbf{x}_i) - f(\mathbf{x}_j) \right)^2$$

est un rkhs dont le noyau reproduisant est $(-L)^*$, *la pseudo-inverse de l'opposé du* Laplacien du graphe $L = A - D$.

Laplacien d'un graphe



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

Le Laplacien est une matrice *symétrique*

Propriétés du Laplacien

Lemme 2 *Soit $L = A - D$ le Laplacien du graphe:*

- *Pour tout $f \in \mathbb{R}^m$,*

$$\Omega(f) = \sum_{i \sim j} \left(f\left(\mathbf{x}_i\right) - f\left(\mathbf{x}_j\right) \right)^2 = -f^\top L f$$

- *$-L$ est une matrice* semi-définie positive
- *0 est une valeur propre de multiplicité 1 associé au vecteur propre $\mathbf{1} = (1, \ldots, 1)$*
- *$Im(L) = \mathcal{H}$*

Preuve : lien entre $\Omega(f)$ et L

$$
\begin{aligned}
\Omega\left(f\right) &= \sum_{i\sim j}\left(f\left(\mathbf{x}_i\right)-f\left(\mathbf{x}_j\right)\right)^2 \\
&= \sum_{i\sim j}\left(f\left(\mathbf{x}_i\right)^2+f\left(\mathbf{x}_j\right)^2-2f\left(\mathbf{x}_i\right)f\left(\mathbf{x}_j\right)\right) \\
&= \sum_{i=1}^{m} D_{i,i}f\left(\mathbf{x}_i\right)^2-2\sum_{i\sim j}f\left(\mathbf{x}_i\right)f\left(\mathbf{x}_j\right) \\
&= f^\top Df-f^\top Af \\
&= -f^\top Lf
\end{aligned}
$$

Preuve : structure propre de L

- Pour tout $f \in \mathbb{R}^m$, $-f^\top L f = \Omega(f) \geq 0$, donc les valeurs propres de $-L$ sont ≥ 0 : $-L$ *est semi-définie positive*.
- f est un vecteur propre associé à la valeur propre 0
  ssi $f^\top L f = 0$
  ssi $\sum_{i \sim j} \left(f\left(\mathbf{x}_i\right) - f\left(\mathbf{x}_j\right)\right)^2 = 0$,
  ssi $f\left(\mathbf{x}_i\right) = f\left(\mathbf{x}_j\right)$ quand $i \sim j$,
  ssi f *est constante* (car le graphe est connexe).
- L étant symétrique, $Im(L)$ est le supplémentaire orthogonal de $Ker(L)$, c'est-à-dire $\mathcal{H}$.

Pseudo-inverse : rappel

La pseudo-inverse $(-L)^*$ de $-L$ est l'application qui vaut:

- 0 sur $Ker(-L)$
- L^{-1} sur $Im(-L)$, c'est-à-dire, en écrivant:

$$-L = \sum_{i=1}^{m} \lambda_i u_i u_i^\top$$

  la décomposition propre de $-L$:

$$(-L)^* = \sum_{\lambda_i \neq 0} (\lambda_i)^{-1} u_i u_i^\top .$$

- On a en particulier $(-L)^*(-L) = (-L)(-L)^* = \Pi_{\mathcal{H}}$, la projection sur $Im(-L) = \mathcal{H}$.

Preuve du Théorème ??

- Restreinte à $\mathcal{H}$, la forme bilinéaire symétrique:

$$\langle f, g \rangle = -f^\top L g$$

  est un définie positive (car $-L$ est semi-définie positive, et $\mathcal{H} = Im(-L)$). C'est donc un produit scalaire, ce qui fait de $\mathcal{H}$ un *espace de Hilbert* (en fait Euclidien).
- La norme de $\mathcal{H}$ est donnée par:

$$\| f \|^2 = \langle f, f \rangle = -f^\top L f = \Omega(f)$$

Preuve du Théorème ?? (cont.)

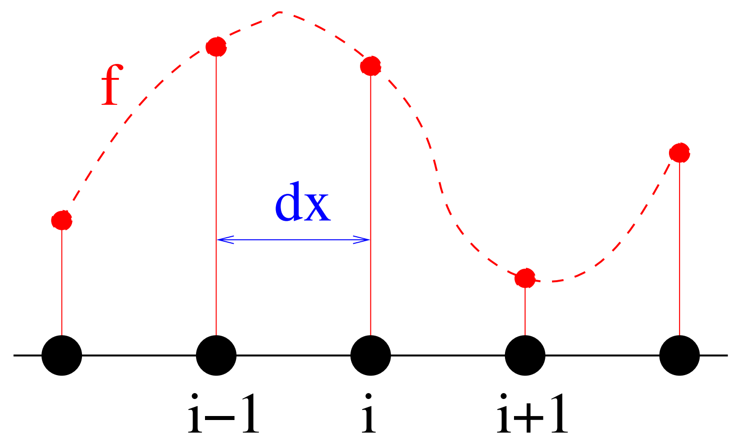
Pour vérifier que $(-L)^*$ est le n.r. de $\mathcal{H}$, il suffit de remarquer que:

- $Ker\left((-L)^*\right) = Ker\left(L\right)$, donc $(-L)^*\mathbf{1} = 0$, donc chaque ligne/colonne de $(-L)^*$ est dans $\mathcal{H}$
- Enfin, pour tout $f \in \mathcal{H}$, en notant $g_i = \langle f, (-L)^*(i, \cdot)\rangle$:

$$g = (-L)^* L f = \Pi_{\mathcal{H}}(f) = f,$$

donc $(-L)^*$ *est bien le n.r. de* $\mathcal{H}$

Interprétation : Laplacien



$$\begin{aligned}
\Delta f(x) &= f''(x) \\
&\sim \frac{f'(x+dx/2) - f'(x-dx/2)}{dx} \\
&\sim \frac{f(x+dx) - f(x) - f(x) + f(x-dx)}{dx^2} \\
&= \frac{f_{i-1} + f_{i+1} - 2f(x)}{dx^2} \\
&= \frac{Lf(i)}{dx^2}.
\end{aligned}$$

Interprétation : régularisation

Pour $f = [0,1] \to \mathbb{R}$ et $x_i = i/m$, on a:

$$\begin{aligned}
\Omega(f) &= \sum_{i=1}^{m} \left(f\left(\frac{i+1}{m}\right) - f\left(\frac{i}{m}\right)\right)^2 \\
&\sim \sum_{i=1}^{m} \left(\frac{1}{m} \times f'\left(\frac{i}{m}\right)\right)^2 \\
&= \frac{1}{m} \times \frac{1}{m} \sum_{i=1}^{m} f'\left(\frac{i}{m}\right)^2 \\
&\sim \frac{1}{m} \int_0^1 f'(t)^2 dt.
\end{aligned}$$

Noyau de diffusion

Motivation

- Soit the noyau Gaussien normalisé sur $\mathbb{R}^d$:

$$K_t\left(\mathbf{x},\mathbf{x}'\right) = \frac{1}{(4\pi t)^{\frac{d}{2}}} \exp\left(-\frac{\parallel \mathbf{x}-\mathbf{x}' \parallel^2}{4t}\right).$$

- La généralisation au graphe en remplaçant la distance Euclidienne par une distance de graphe ne marche pas.
- Nous allons chercher une caractérisation de ce noyau comme solution d'une équation faisant *intervenir le Laplacien*, pour l'étendre au graphe: *l'équation de diffusion*.

Equation de diffusion

Lemme 3 *Pour tout* $\mathbf{x}_0 \in \mathbb{R}^d$*, la fonction:*

$$K_{\mathbf{x}_0}(\mathbf{x}, t) = K_t(\mathbf{x}_0, \mathbf{x}) = \frac{1}{(4\pi t)^{\frac{d}{2}}} \exp\left(-\frac{\| \mathbf{x} - \mathbf{x}_0 \|^2}{4t}\right).$$

est solution de l'équation de diffusion:

$$\frac{\partial}{\partial t} K_{\mathbf{x}_0}(\mathbf{x}, t) = \Delta K_{\mathbf{x}_0}(\mathbf{x}, t).$$

sous la condition initiale $K_{\mathbf{x}_0}(\mathbf{x}, 0) = \delta_{\mathbf{x}_0}(\mathbf{x})$

(preuve = simple calcul laissé en exercice).

Equation de diffusion discrète

Pour $f_t \in \mathbb{R}^m$, l'équation de diffusion devient:

$$\frac{\partial}{\partial t} f_t = L f_t$$

qui admet pour solution:

$$f_t = f_0 e^{tL}$$

avec

$$e^{tL} = I + tL + \frac{t^2}{2!}L^2 + \frac{t^3}{3!}L^3 + \dots$$

Noyau de diffusion

Cela suggère de considérer le noyau:

$$K = e^{tL}$$

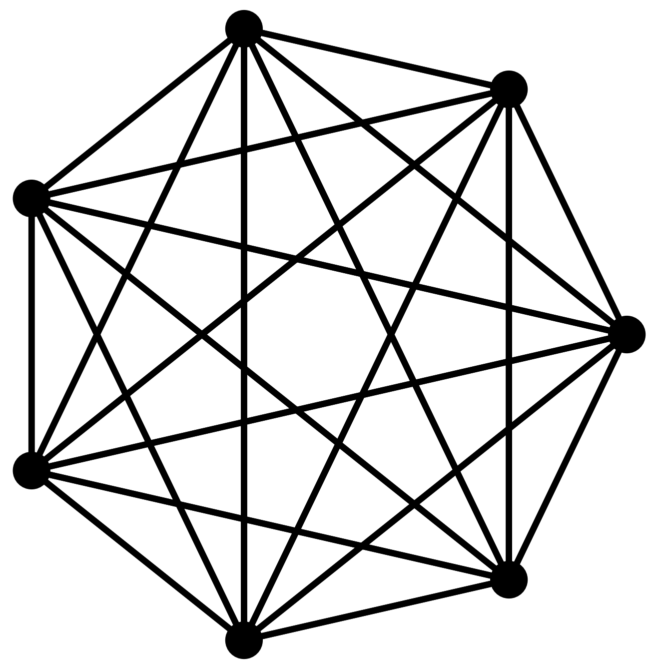
qui est bien symétrique et d.p. car en écrivant:

$$L = \sum_{i=1}^{m} (-\lambda_i) u_i u_i^\top \quad (\lambda_i \geq 0)$$

on a:

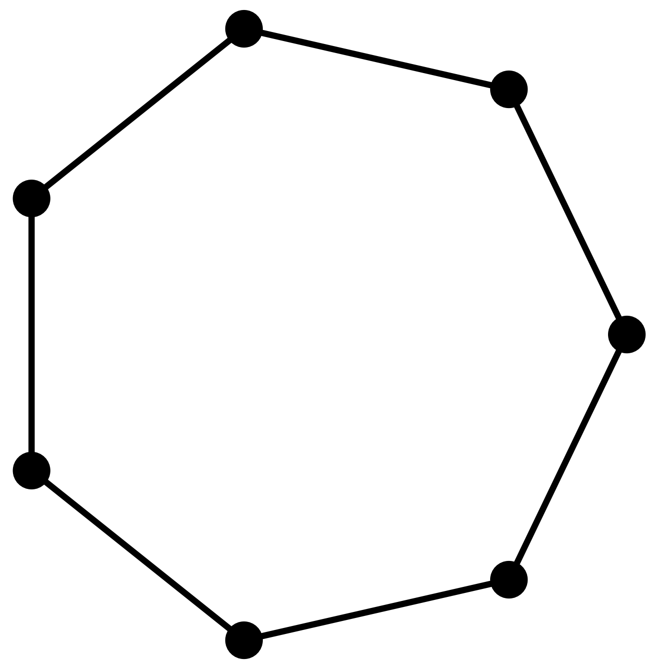
$$K = e^{tL} = \sum_{i=1}^{m} e^{-t\lambda_i} u_i u_i^\top$$

Exemple: graphe complet



$$K_{i,j} = \begin{cases} \frac{1+(m-1)e^{-tm}}{m} & \text{for } i = j, \\ \frac{1-e^{-tm}}{m} & \text{for } i \neq j. \end{cases}$$

Exemple: chaîne fermée



$$K_{i,j} = \frac{1}{m} \sum_{\nu=0}^{m-1} \exp\left[-2t\left(1 - \cos\frac{2\pi\nu}{m}\right)\right] \cos\frac{2\pi\nu(i-j)}{m}.$$

Généralisation par analyse harmonique

Spectre du noyau de diffusion

- Soit $0 = \lambda_1 > -\lambda_2 \geq \ldots \geq -\lambda_m$ les valeurs propres du Laplacien:

$$L = \sum_{i=1}^{m} (-\lambda_i) u_i u_i^\top \quad (\lambda_i \geq 0)$$

- Le noyau de diffusion K_t est une matrice *inversible* car des valeurs propres sont strictement positives:

$$K_t = \sum_{i=1}^{m} e^{-t\lambda_i} u_i u_i^\top$$

Norme dans le rkhs

- Tout fonction $f \in \mathbb{R}^m$ s'écrit $f = K\left(K^{-1}f\right)$, donc sa norme dans le rkhs est:

$$\| f \|_{K_t}^2 = \left(f^\top K^{-1}\right) K \left(K^{-1} f\right) = f^\top K^{-1} f$$

Norme dans le rkhs (cont.)

- Pour $i = 1, \ldots, m$, soit:

$$\hat{f}_i = u_i^\top f$$

  la projection de f sur la base de vecteurs propres.

- On a alors:

$$\| f \|_{K_t}^2 = f^\top K^{-1} f = \sum_{i=1}^{m} e^{t\lambda_i} \hat{f}_i^2 .$$

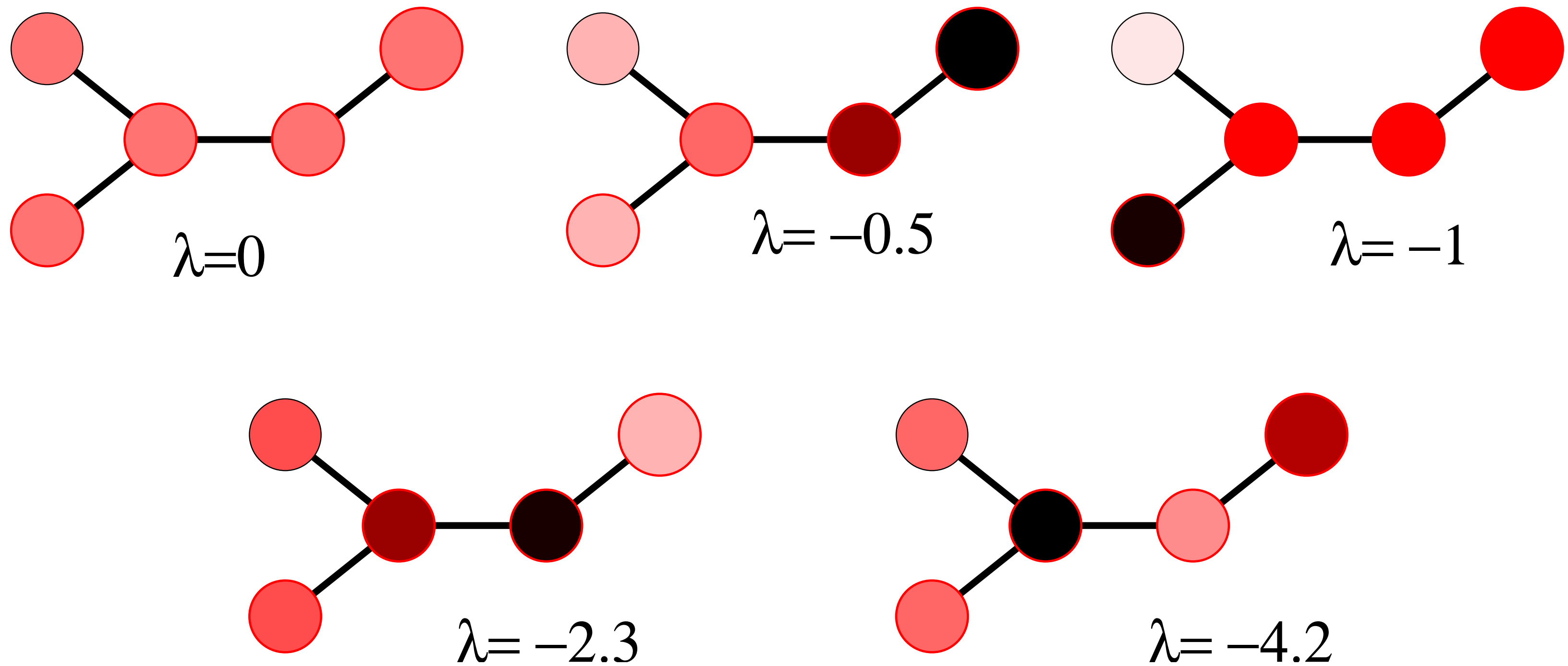
- Ressemblance avec $\int \left| \hat{f}(\omega) \right|^2 e^{\sigma^2 \omega^2} d\omega$

Transformée de Fourier discrète

Définition 4 *Le vecteur* $\hat{f} = \left(\hat{f}_1, \ldots, \hat{f}_m\right)^\top$ *est appelé* transformée de Fourier discrète *du vecteur* $f \in \mathbb{R}^n$

- Les fonctions propres du Laplacien sont l'équivalent discret des sinus/cosinus
- Les valeurs propres λ_i sont les équivalents des fréquences $(i\omega)^2$
- Les fonctions propres successives "oscillent" de plus en plus quand les valeurs propres diminuent

Fonctions propres du Laplacien



Généralisation

Cette représentation suggère de définir une famille de noyaux:

$$K_r = \sum_{i=1}^{m} r(\lambda_i) u_i u_i^\top$$

définissant une norme:

$$\| f \|_{K_r}^2 = \sum_{i=1}^{m} \frac{\hat{f}_i^2}{r(\lambda_i)}$$

avec une fonction $r : \mathbb{R}^+ \to \mathbb{R}_*^+$ *décroissante*.

Exemple : Laplacien régularisé

$$r(\lambda) = \frac{1}{\lambda + \epsilon}, \qquad \epsilon > 0$$

$$K = \sum_{i=1}^{m} \frac{1}{\lambda_i + \epsilon} u_i u_i^\top = (-L + \epsilon I)^{-1}$$

$$\| f \|_K^2 = f^\top K^{-1} f = \sum_{i \sim j} \left(f(\mathbf{x}_i) - f(\mathbf{x}_j) \right)^2 + \epsilon \sum_{i=1}^{m} f(\mathbf{x}_i)^2 .$$

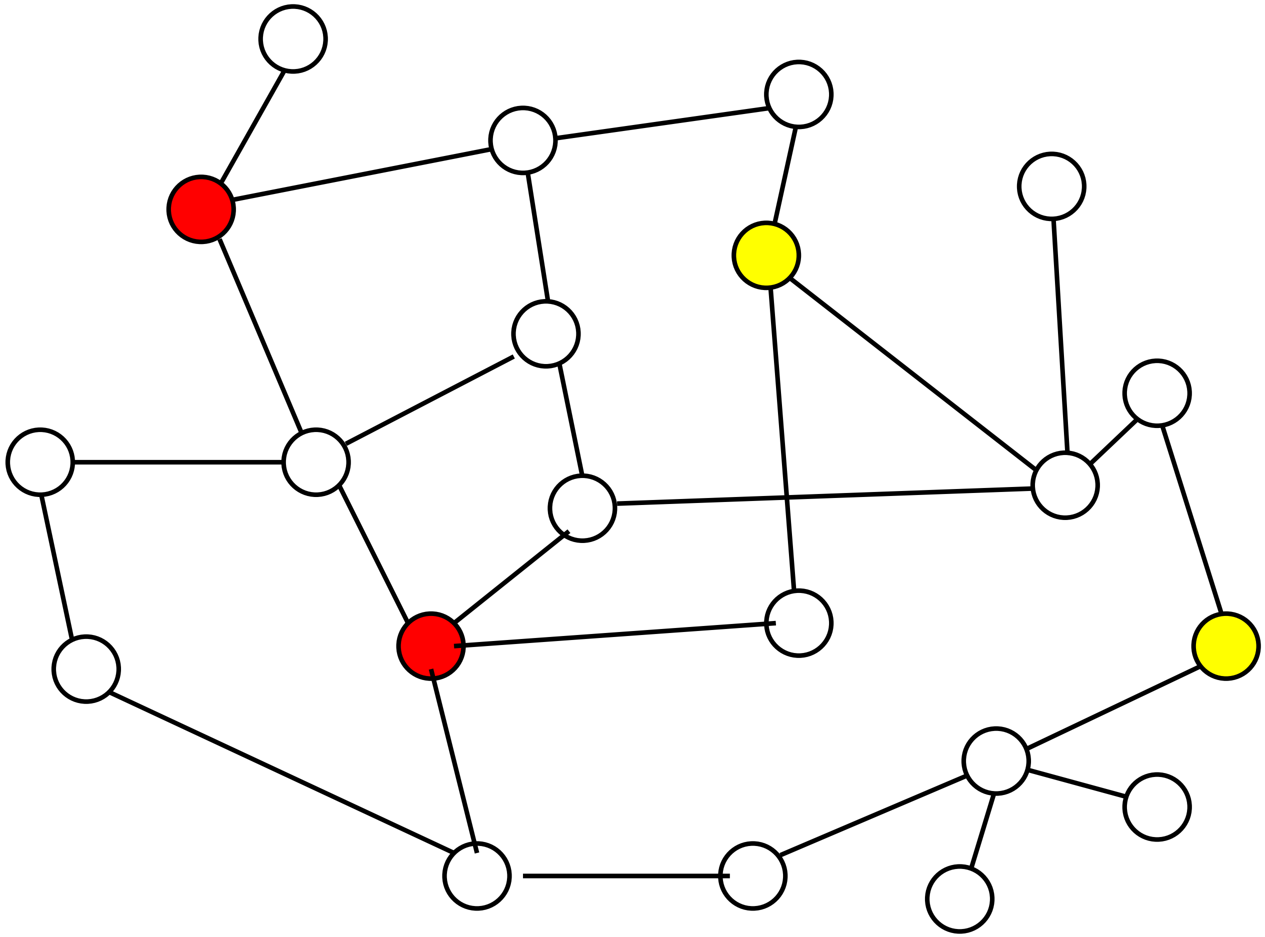
Remarque

Cette manière de construire des noyaux d.p. est applicable à tout espace avec transformée de Fourier / analyse harmonique:

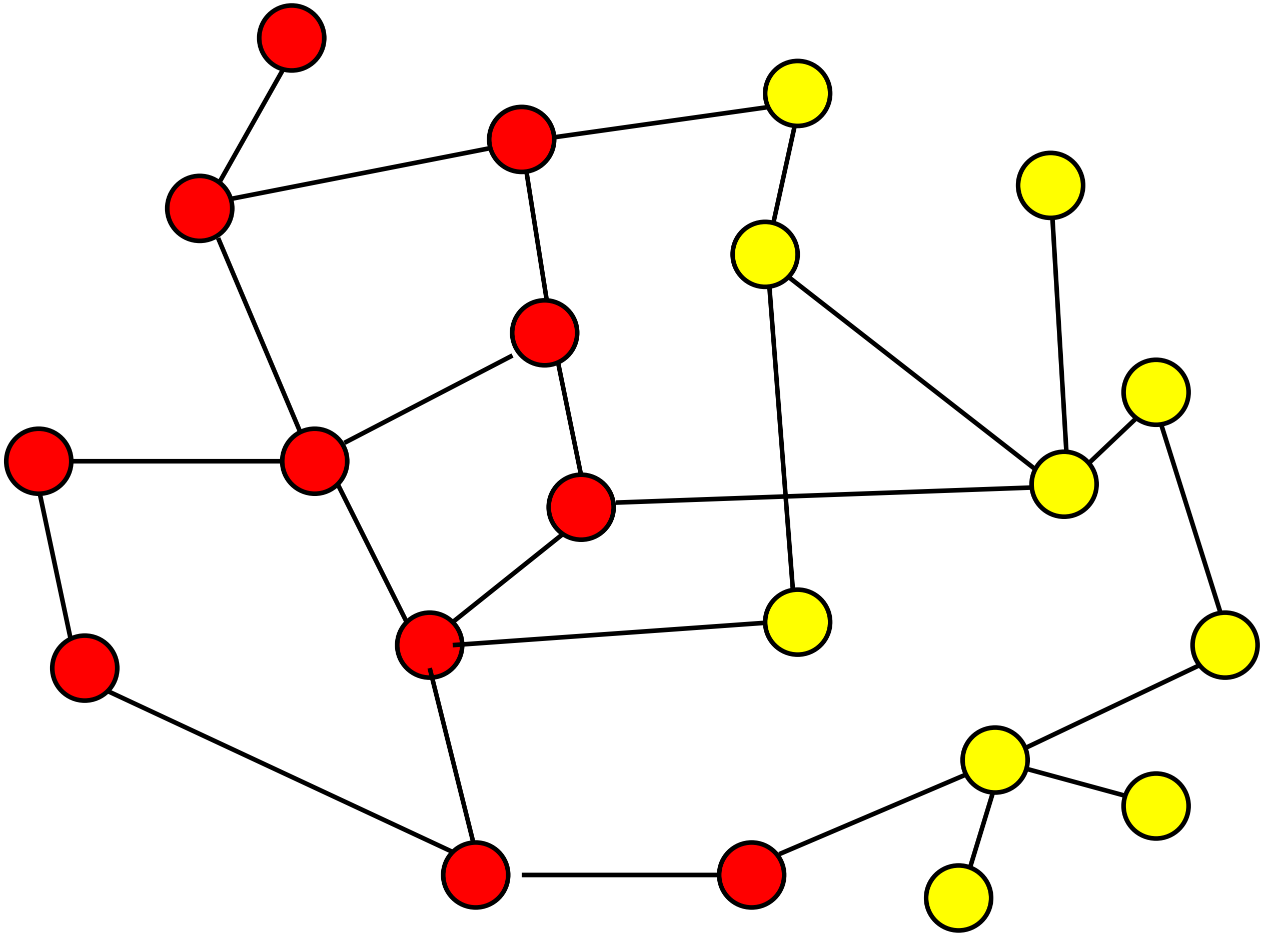
- graphes
- variétés différentielles
- groupes et semi-groupes (etc...)

Application

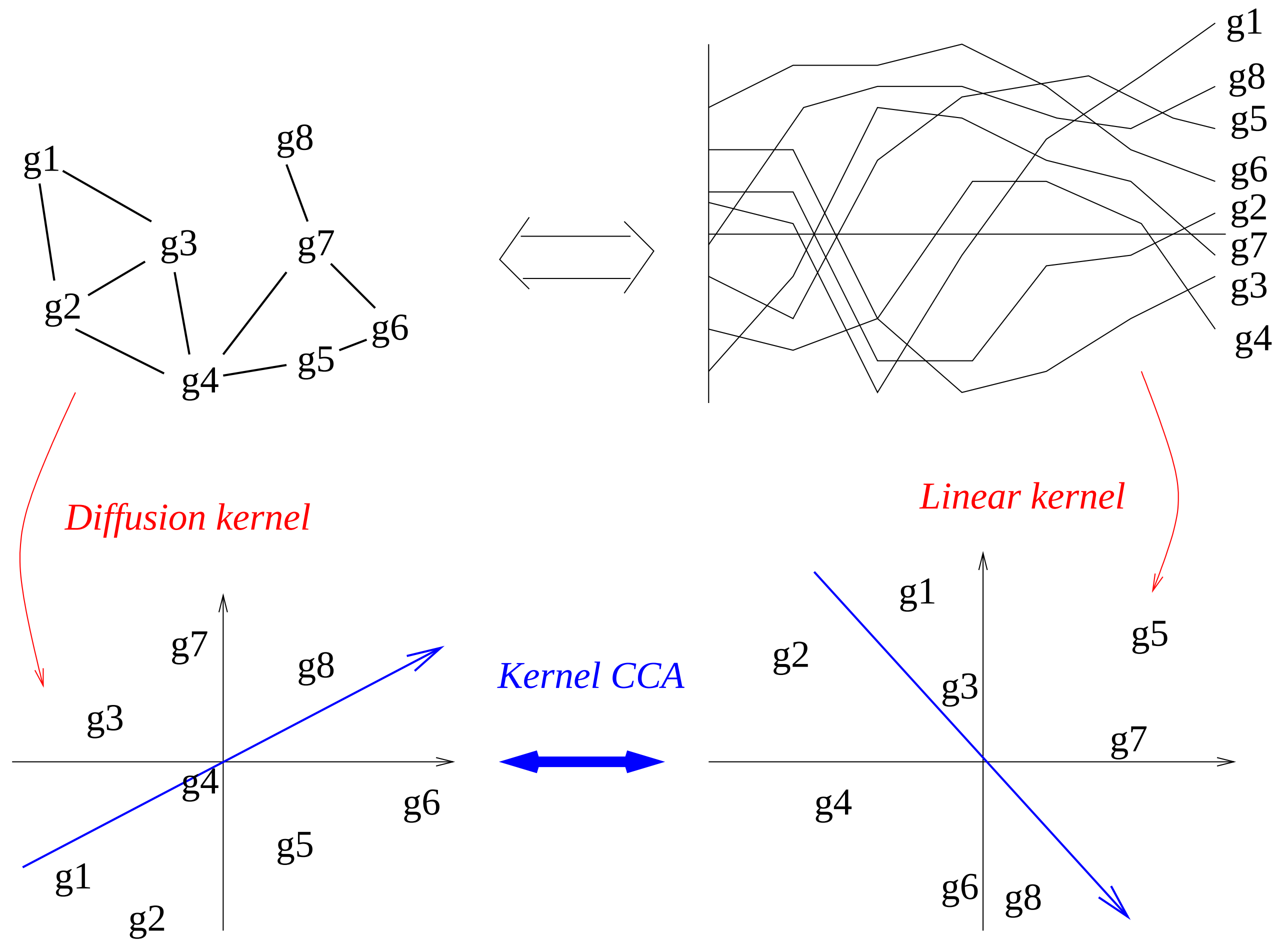
Classification (semi-) supervisée



Classification (semi-) supervisée



Comparaison de données hétérogènes



Références

- R. I. Kondor et J. Lafferty, **Diffusion Kernels on Graphs and Other Discrete Input**, Proceedings of ICML 2002. *-> Introduit le noyau de diffusion et les SVM sur les graphes*
- J.-P. Vert et M. Kanehisa, **Graph-driven features extraction from microarray data using diffusion kernels and kernel CCA**, Proceedings of NIPS 2002. *-> Lien avec transformée de Fourier, application en bioinformatique*
- A. Smola and R. Kondor, **Kernels and regularization on graphs**, Proceedings of COLT 2003. *-> Variantes avec* $r(\lambda)$