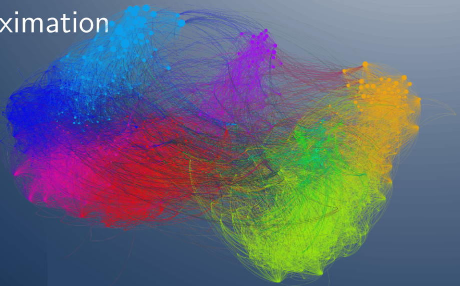
Graphs in Machine Learning

Spectral Clustering: Relaxation

Rayleigh-Ritz and RatioCut Approximation

Michal Valko

Inria & ENS Paris-Saclay, MVA



Partially based on material by: Ulrike von Luxburg, Gary Miller, Doyle & Schnell, Daniel Spielman

Spectral Clustering: Relaxing Balanced Cuts

Relaxation for (simple) balanced cuts for 2 sets

$$\min_{A,B} \mathrm{cut}(A, B) \quad \text{s.t.} \quad |A| = |B|$$

Spectral Clustering: Relaxing Balanced Cuts

Relaxation for (simple) balanced cuts for 2 sets

$$\min_{A,B} \text{cut}(A, B) \quad \text{s.t.} \quad |A| = |B|$$

Graph function $\mathbf{f}$ for cluster membership

Spectral Clustering: Relaxing Balanced Cuts

Relaxation for (simple) balanced cuts for 2 sets

$$\min_{A,B} \operatorname{cut}(A, B) \quad \text{s.t.} \quad |A| = |B|$$

Graph function $\mathbf{f}$ for cluster membership: $f_i = \begin{cases} 1 & \text{if } V_i \in A, \\ -1 & \text{if } V_i \in B. \end{cases}$

Spectral Clustering: Relaxing Balanced Cuts

Relaxation for (simple) balanced cuts for 2 sets

$$\min_{A,B} \operatorname{cut}(A,B) \quad \text{s.t.} \quad |A| = |B|$$

Graph function $\mathbf{f}$ for cluster membership: $f_i = \begin{cases} 1 & \text{if } V_i \in A, \\ -1 & \text{if } V_i \in B. \end{cases}$

What it is the cut value with this definition?

$$\operatorname{cut}(A,B) =$$

Spectral Clustering: Relaxing Balanced Cuts

Relaxation for (simple) balanced cuts for 2 sets

$$\min_{A,B} \operatorname{cut}(A,B) \quad \text{s.t.} \quad |A| = |B|$$

Graph function $\mathbf{f}$ for cluster membership: $f_i = \begin{cases} 1 & \text{if } V_i \in A, \\ -1 & \text{if } V_i \in B. \end{cases}$

What it is the cut value with this definition?

$$\operatorname{cut}(A,B) = \sum_{i \in A, j \in B} w_{i,j} = \frac{1}{4}\sum_{i,j} w_{i,j}(f_i - f_j)^2 = \frac{1}{2}\mathbf{f}^\intercal \mathbf{L}\mathbf{f}$$

Spectral Clustering: Relaxing Balanced Cuts

Relaxation for (simple) balanced cuts for 2 sets

$$\min_{A,B} \operatorname{cut}(A,B) \quad \text{s.t.} \quad |A| = |B|$$

Graph function $\mathbf{f}$ for cluster membership: $f_i = \begin{cases} 1 & \text{if } V_i \in A, \\ -1 & \text{if } V_i \in B. \end{cases}$

What it is the cut value with this definition?

$$\operatorname{cut}(A,B) = \sum_{i \in A, j \in B} w_{i,j} = \frac{1}{4} \sum_{i,j} w_{i,j}(f_i - f_j)^2 = \frac{1}{2}\mathbf{f}^{\mathsf{T}}\mathbf{L}\mathbf{f}$$

What is the relationship with the **smoothness** of a graph function?

Spectral Clustering: Relaxing Balanced Cuts

$$\mathrm{cut}(A, B) = \sum_{i \in A, j \in B} w_{i,j} = \frac{1}{4} \sum_{i,j} w_{i,j} (f_i - f_j)^2 = \frac{1}{2} \mathbf{f}^\mathsf{T} \mathbf{L} \mathbf{f}$$

Spectral Clustering: Relaxing Balanced Cuts

$$\mathrm{cut}(A, B) = \sum_{i \in A, j \in B} w_{i,j} = \frac{1}{4} \sum_{i,j} w_{i,j}(f_i - f_j)^2 = \frac{1}{2}\mathbf{f}^\mathsf{T}\mathbf{L}\mathbf{f}$$

$|A| = |B| \implies \sum_i f_i =$

Spectral Clustering: Relaxing Balanced Cuts

$$\mathrm{cut}(A, B) = \sum_{i \in A, j \in B} w_{i,j} = \frac{1}{4} \sum_{i,j} w_{i,j}(f_i - f_j)^2 = \frac{1}{2}\mathbf{f}^\mathsf{T}\mathbf{L}\mathbf{f}$$

$|A| = |B| \implies \sum_i f_i = 0 \implies$

Spectral Clustering: Relaxing Balanced Cuts

$$\mathrm{cut}(A,B) = \sum_{i\in A, j\in B} w_{i,j} = \frac{1}{4}\sum_{i,j} w_{i,j}(f_i - f_j)^2 = \frac{1}{2}\mathbf{f}^\mathsf{T}\mathbf{L}\mathbf{f}$$

$|A| = |B| \implies \sum_i f_i = 0 \implies \mathbf{f} \perp \mathbf{1}_N$

$\|\mathbf{f}\| =$

Spectral Clustering: Relaxing Balanced Cuts

$$\mathrm{cut}(A,B) = \sum_{i\in A, j\in B} w_{i,j} = \frac{1}{4}\sum_{i,j} w_{i,j}(f_i - f_j)^2 = \frac{1}{2}\mathbf{f}^\mathsf{T}\mathbf{L}\mathbf{f}$$

$|A| = |B| \implies \sum_i f_i = 0 \implies \mathbf{f} \perp \mathbf{1}_N$

$\|\mathbf{f}\| = \sqrt{N}$

Spectral Clustering: Relaxing Balanced Cuts

$$\text{cut}(A, B) = \sum_{i \in A, j \in B} w_{i,j} = \frac{1}{4} \sum_{i,j} w_{i,j}(f_i - f_j)^2 = \frac{1}{2}\mathbf{f}^\intercal \mathbf{L}\mathbf{f}$$

$$|A| = |B| \implies \sum_i f_i = 0 \implies \mathbf{f} \perp \mathbf{1}_N$$

$$\|\mathbf{f}\| = \sqrt{N}$$

objective function of spectral clustering

$$\min_{\mathbf{f}} \mathbf{f}^\intercal \mathbf{L}\mathbf{f} \quad \text{s.t.} \quad f_i = \pm 1, \quad \mathbf{f} \perp \mathbf{1}_N, \quad \|\mathbf{f}\| = \sqrt{N}$$

Spectral Clustering: Relaxing Balanced Cuts

$$\mathrm{cut}(A,B) = \sum_{i\in A, j\in B} w_{i,j} = \frac{1}{4}\sum_{i,j} w_{i,j}(f_i - f_j)^2 = \frac{1}{2}\mathbf{f}^\mathsf{T}\mathbf{L}\mathbf{f}$$

$|A| = |B| \implies \sum_i f_i = 0 \implies \mathbf{f} \perp \mathbf{1}_N$

$\|\mathbf{f}\| = \sqrt{N}$

objective function of spectral clustering

$$\min_{\mathbf{f}} \mathbf{f}^\mathsf{T}\mathbf{L}\mathbf{f} \quad \text{s.t.} \quad f_i = \pm 1, \quad \mathbf{f} \perp \mathbf{1}_N, \quad \|\mathbf{f}\| = \sqrt{N}$$

Still NP hard :(

Spectral Clustering: Relaxing Balanced Cuts

$$\mathrm{cut}(A,B) = \sum_{i\in A, j\in B} w_{i,j} = \frac{1}{4}\sum_{i,j} w_{i,j}(f_i - f_j)^2 = \frac{1}{2}\mathbf{f}^\mathsf{T}\mathbf{L}\mathbf{f}$$

$$|A| = |B| \implies \sum_i f_i = 0 \implies \mathbf{f} \perp \mathbf{1}_N$$

$$\|\mathbf{f}\| = \sqrt{N}$$

objective function of spectral clustering

$$\min_{\mathbf{f}} \mathbf{f}^\mathsf{T}\mathbf{L}\mathbf{f} \quad \text{s.t.} \quad f_i = \pm 1, \quad \mathbf{f} \perp \mathbf{1}_N, \quad \|\mathbf{f}\| = \sqrt{N}$$

Still NP hard :($\rightarrow$ Relax even further!

Spectral Clustering: Relaxing Balanced Cuts

$$\text{cut}(A,B) = \sum_{i\in A, j\in B} w_{i,j} = \frac{1}{4}\sum_{i,j} w_{i,j}(f_i - f_j)^2 = \frac{1}{2}\mathbf{f}^\mathsf{T}\mathbf{L}\mathbf{f}$$

$|A| = |B| \implies \sum_i f_i = 0 \implies \mathbf{f} \perp \mathbf{1}_N$

$\|\mathbf{f}\| = \sqrt{N}$

objective function of spectral clustering

$$\min_{\mathbf{f}} \mathbf{f}^\mathsf{T}\mathbf{L}\mathbf{f} \quad \text{s.t.} \quad f_i = \pm 1, \quad \mathbf{f} \perp \mathbf{1}_N, \quad \|\mathbf{f}\| = \sqrt{N}$$

Still NP hard :($\rightarrow$ Relax even further!

~~$f_i = \pm 1$~~ $\rightarrow$ $f_i \in \mathbb{R}$

Spectral Clustering: Relaxing Balanced Cuts

objective function of spectral clustering

$$\min_{\mathbf{f}} \mathbf{f}^\mathsf{T}\mathbf{L}\mathbf{f} \quad \text{s.t.} \quad f_i \in \mathbb{R}, \quad \mathbf{f} \perp \mathbf{1}_N, \quad \|\mathbf{f}\| = \sqrt{N}$$

Spectral Clustering: Relaxing Balanced Cuts

objective function of spectral clustering

$$\min_{\mathbf{f}} \mathbf{f}^{\mathsf{T}}\mathbf{L}\mathbf{f} \quad \text{s.t.} \quad f_i \in \mathbb{R}, \quad \mathbf{f} \perp \mathbf{1}_N, \quad \|\mathbf{f}\| = \sqrt{N}$$

Rayleigh-Ritz theorem

If $\lambda_1 \leq \cdots \leq \lambda_N$ are the eigenvalues of real symmetric $\mathbf{L}$ then

$$\lambda_1 = \min_{\mathbf{x} \neq 0} \frac{\mathbf{x}^{\mathsf{T}}\mathbf{L}\mathbf{x}}{\mathbf{x}^{\mathsf{T}}\mathbf{x}} = \min_{\mathbf{x}^{\mathsf{T}}\mathbf{x}=1} \mathbf{x}^{\mathsf{T}}\mathbf{L}\mathbf{x}$$

$$\lambda_N = \max_{\mathbf{x} \neq 0} \frac{\mathbf{x}^{\mathsf{T}}\mathbf{L}\mathbf{x}}{\mathbf{x}^{\mathsf{T}}\mathbf{x}} = \max_{\mathbf{x}^{\mathsf{T}}\mathbf{x}=1} \mathbf{x}^{\mathsf{T}}\mathbf{L}\mathbf{x}$$

$\frac{\mathbf{x}^{\mathsf{T}}\mathbf{L}\mathbf{x}}{\mathbf{x}^{\mathsf{T}}\mathbf{x}} \equiv$ **Rayleigh quotient**

Spectral Clustering: Relaxing Balanced Cuts

objective function of spectral clustering

$$\min_{\mathbf{f}} \mathbf{f}^\mathsf{T}\mathbf{L}\mathbf{f} \quad \text{s.t.} \quad f_i \in \mathbb{R}, \quad \mathbf{f} \perp \mathbf{1}_N, \quad \|\mathbf{f}\| = \sqrt{N}$$

Rayleigh-Ritz theorem

If $\lambda_1 \leq \cdots \leq \lambda_N$ are the eigenvalues of real symmetric $\mathbf{L}$ then

$$\lambda_1 = \min_{\mathbf{x} \neq 0} \frac{\mathbf{x}^\mathsf{T}\mathbf{L}\mathbf{x}}{\mathbf{x}^\mathsf{T}\mathbf{x}} = \min_{\mathbf{x}^\mathsf{T}\mathbf{x}=1} \mathbf{x}^\mathsf{T}\mathbf{L}\mathbf{x}$$

$$\lambda_N = \max_{\mathbf{x} \neq 0} \frac{\mathbf{x}^\mathsf{T}\mathbf{L}\mathbf{x}}{\mathbf{x}^\mathsf{T}\mathbf{x}} = \max_{\mathbf{x}^\mathsf{T}\mathbf{x}=1} \mathbf{x}^\mathsf{T}\mathbf{L}\mathbf{x}$$

$\frac{\mathbf{x}^\mathsf{T}\mathbf{L}\mathbf{x}}{\mathbf{x}^\mathsf{T}\mathbf{x}} \equiv$ **Rayleigh quotient**

How can we use it?

Spectral Clustering: Relaxing Balanced Cuts

objective function of spectral clustering

$$\min_{\mathbf{f}} \mathbf{f}^\mathsf{T}\mathbf{L}\mathbf{f} \quad \text{s.t.} \quad f_i \in \mathbb{R}, \quad \mathbf{f} \perp \mathbf{1}_N, \quad \|\mathbf{f}\| = \sqrt{N}$$

Generalized Rayleigh-Ritz theorem (Courant-Fischer-Weyl)

If $\lambda_1 \leq \cdots \leq \lambda_N$ are the eigenvalues of real symmetric $\mathbf{L}$ and $\mathbf{v}_1, \ldots, \mathbf{v}_N$ the corresponding orthogonal eigenvectors, then for $k = 1 : N - 1$

$$\lambda_{k+1} = \min_{\mathbf{x} \neq 0, \mathbf{x} \perp \mathbf{v}_1, \ldots \mathbf{v}_k} \frac{\mathbf{x}^\mathsf{T}\mathbf{L}\mathbf{x}}{\mathbf{x}^\mathsf{T}\mathbf{x}} = \min_{\mathbf{x}^\mathsf{T}\mathbf{x} = 1, \mathbf{x} \perp \mathbf{v}_1, \ldots \mathbf{v}_k} \mathbf{x}^\mathsf{T}\mathbf{L}\mathbf{x}$$

$$\lambda_{N-k} = \max_{\mathbf{x} \neq 0, \mathbf{x} \perp \mathbf{v}_n, \ldots \mathbf{v}_{N-k+1}} \frac{\mathbf{x}^\mathsf{T}\mathbf{L}\mathbf{x}}{\mathbf{x}^\mathsf{T}\mathbf{x}} = \max_{\mathbf{x}^\mathsf{T}\mathbf{x} = 1, \mathbf{x} \perp \mathbf{v}_N, \ldots \mathbf{v}_{N-k+1}} \mathbf{x}^\mathsf{T}\mathbf{L}\mathbf{x}$$

Spectral Clustering: Relaxing Balanced Cuts

objective function of spectral clustering

$$\min_{\mathbf{f}} \mathbf{f}^\mathsf{T}\mathbf{L}\mathbf{f} \quad \text{s.t.} \quad f_i \in \mathbb{R}, \quad \mathbf{f} \perp \mathbf{1}_N, \quad \|\mathbf{f}\| = \sqrt{N}$$

Solution:

Spectral Clustering: Relaxing Balanced Cuts

objective function of spectral clustering

$$\min_{\mathbf{f}} \mathbf{f}^\mathsf{T}\mathbf{L}\mathbf{f} \quad \text{s.t.} \quad f_i \in \mathbb{R}, \quad \mathbf{f} \perp \mathbf{1}_N, \quad \|\mathbf{f}\| = \sqrt{N}$$

Solution: **second eigenvector** How do we get the clustering?

Spectral Clustering: Relaxing Balanced Cuts

objective function of spectral clustering

$$\min_{\mathbf{f}} \mathbf{f}^\mathsf{T}\mathbf{L}\mathbf{f} \quad \text{s.t.} \quad f_i \in \mathbb{R}, \quad \mathbf{f} \perp \mathbf{1}_N, \quad \|\mathbf{f}\| = \sqrt{N}$$

Solution: **second eigenvector** How do we get the clustering?

The solution may not be integral. What to do?

Spectral Clustering: Relaxing Balanced Cuts

objective function of spectral clustering

$$\min_{\mathbf{f}} \mathbf{f}^\mathsf{T}\mathbf{L}\mathbf{f} \quad \text{s.t.} \quad f_i \in \mathbb{R}, \quad \mathbf{f} \perp \mathbf{1}_N, \quad \|\mathbf{f}\| = \sqrt{N}$$

Solution: **second eigenvector** How do we get the clustering?

The solution may not be integral. What to do?

$$\text{cluster}_i = \begin{cases} 1 & \text{if } f_i \geq 0, \\ -1 & \text{if } f_i < 0. \end{cases}$$

Spectral Clustering: Relaxing Balanced Cuts

objective function of spectral clustering

$$\min_{\mathbf{f}} \mathbf{f}^\mathsf{T}\mathbf{L}\mathbf{f} \quad \text{s.t.} \quad f_i \in \mathbb{R}, \quad \mathbf{f} \perp \mathbf{1}_N, \quad \|\mathbf{f}\| = \sqrt{N}$$

Solution: **second eigenvector** How do we get the clustering?

The solution may not be integral. What to do?

$$\text{cluster}_i = \begin{cases} 1 & \text{if } f_i \geq 0, \\ -1 & \text{if } f_i < 0. \end{cases}$$

Works but this heuristics is often too simple.

Spectral Clustering: Relaxing Balanced Cuts

objective function of spectral clustering

$$\min_{\mathbf{f}} \mathbf{f}^\mathsf{T}\mathbf{L}\mathbf{f} \quad \text{s.t.} \quad f_i \in \mathbb{R}, \quad \mathbf{f} \perp \mathbf{1}_N, \quad \|\mathbf{f}\| = \sqrt{N}$$

Solution: **second eigenvector** How do we get the clustering?

The solution may not be integral. What to do?

$$\text{cluster}_i = \begin{cases} 1 & \text{if } f_i \geq 0, \\ -1 & \text{if } f_i < 0. \end{cases}$$

Works but this heuristics is often too simple. In practice, cluster $\mathbf{f}$ using k-means to get $\{C_i\}_i$ and assign:

Spectral Clustering: Relaxing Balanced Cuts

objective function of spectral clustering

$$\min_{\mathbf{f}} \mathbf{f}^{\mathsf{T}}\mathbf{L}\mathbf{f} \quad \text{s.t.} \quad f_i \in \mathbb{R}, \quad \mathbf{f} \perp \mathbf{1}_N, \quad \|\mathbf{f}\| = \sqrt{N}$$

Solution: **second eigenvector** How do we get the clustering?

The solution may not be integral. What to do?

$$\text{cluster}_i = \begin{cases} 1 & \text{if } f_i \geq 0, \\ -1 & \text{if } f_i < 0. \end{cases}$$

Works but this heuristics is often too simple. In practice, cluster $\mathbf{f}$ using k-means to get $\{C_i\}_i$ and assign:

$$\text{cluster}_i = \begin{cases} 1 & \text{if } i \in C_1, \\ -1 & \text{if } i \in C_{-1}. \end{cases}$$

Spectral Clustering: Approximating RatioCut

Wait, but we did not care about approximating mincut!

Spectral Clustering: Approximating RatioCut

Wait, but we did not care about approximating mincut!

RatioCut

$$\mathrm{RatioCut}(A, B) = \sum_{i \in A, j \in B} w_{ij} \left(\frac{1}{|A|} + \frac{1}{|B|} \right)$$

Spectral Clustering: Approximating RatioCut

Wait, but we did not care about approximating mincut!

RatioCut

$$\mathrm{RatioCut}(A, B) = \sum_{i \in A, j \in B} w_{ij} \left(\frac{1}{|A|} + \frac{1}{|B|} \right)$$

Define graph function $\mathbf{f}$ for cluster membership of RatioCut

Spectral Clustering: Approximating RatioCut

Wait, but we did not care about approximating mincut!

RatioCut

$$\mathrm{RatioCut}(A, B) = \sum_{i \in A, j \in B} w_{ij} \left(\frac{1}{|A|} + \frac{1}{|B|} \right)$$

Define graph function $\mathbf{f}$ for cluster membership of RatioCut:

$$f_i = \begin{cases} \sqrt{\frac{|B|}{|A|}} & \text{if } V_i \in A, \\ -\sqrt{\frac{|A|}{|B|}} & \text{if } V_i \in B. \end{cases}$$

Spectral Clustering: Approximating RatioCut

Wait, but we did not care about approximating mincut!

RatioCut

$$\mathrm{RatioCut}(A, B) = \sum_{i \in A, j \in B} w_{ij} \left(\frac{1}{|A|} + \frac{1}{|B|} \right)$$

Define graph function $\mathbf{f}$ for cluster membership of RatioCut:

$$f_i = \begin{cases} \sqrt{\frac{|B|}{|A|}} & \text{if } V_i \in A, \\ -\sqrt{\frac{|A|}{|B|}} & \text{if } V_i \in B. \end{cases}$$

$$\mathbf{f}^{\mathsf{T}} \mathbf{L} \mathbf{f} = \frac{1}{2} \sum_{i,j} w_{i,j} (f_i - f_j)^2 =$$

Spectral Clustering: Approximating RatioCut

Wait, but we did not care about approximating mincut!

RatioCut

$$\mathrm{RatioCut}(A, B) = \sum_{i \in A, j \in B} w_{ij} \left(\frac{1}{|A|} + \frac{1}{|B|} \right)$$

Define graph function $\mathbf{f}$ for cluster membership of RatioCut:

$$f_i = \begin{cases} \sqrt{\frac{|B|}{|A|}} & \text{if } V_i \in A, \\ -\sqrt{\frac{|A|}{|B|}} & \text{if } V_i \in B. \end{cases}$$

$$\mathbf{f}^\mathsf{T}\mathbf{L}\mathbf{f} = \frac{1}{2} \sum_{i,j} w_{i,j} (f_i - f_j)^2 = (|A| + |B|)\mathrm{RatioCut}(A, B)$$

Spectral Clustering: Approximating RatioCut

Define graph function $\mathbf{f}$ for cluster membership of RatioCut

Spectral Clustering: Approximating RatioCut

Define graph function $\mathbf{f}$ for cluster membership of RatioCut:

$$f_i = \begin{cases} \sqrt{\frac{|B|}{|A|}} & \text{if } V_i \in A, \\ -\sqrt{\frac{|A|}{|B|}} & \text{if } V_i \in B. \end{cases}$$

Spectral Clustering: Approximating RatioCut

Define graph function $\mathbf{f}$ for cluster membership of RatioCut:

$$f_i = \begin{cases} \sqrt{\frac{|B|}{|A|}} & \text{if } V_i \in A, \\ -\sqrt{\frac{|A|}{|B|}} & \text{if } V_i \in B. \end{cases}$$

$$\sum_i f_i =$$

Spectral Clustering: Approximating RatioCut

Define graph function $\mathbf{f}$ for cluster membership of RatioCut:

$$f_i = \begin{cases} \sqrt{\frac{|B|}{|A|}} & \text{if } V_i \in A, \\ -\sqrt{\frac{|A|}{|B|}} & \text{if } V_i \in B. \end{cases}$$

$$\sum_i f_i = 0$$

$$\sum_i f_i^2 =$$

Spectral Clustering: Approximating RatioCut

Define graph function $\mathbf{f}$ for cluster membership of RatioCut:

$$f_i = \begin{cases} \sqrt{\frac{|B|}{|A|}} & \text{if } V_i \in A, \\ -\sqrt{\frac{|A|}{|B|}} & \text{if } V_i \in B. \end{cases}$$

$$\sum_i f_i = 0$$

$$\sum_i f_i^2 = N$$

Spectral Clustering: Approximating RatioCut

Define graph function $\mathbf{f}$ for cluster membership of RatioCut:

$$f_i = \begin{cases} \sqrt{\frac{|B|}{|A|}} & \text{if } V_i \in A, \\ -\sqrt{\frac{|A|}{|B|}} & \text{if } V_i \in B. \end{cases}$$

$$\sum_i f_i = 0$$

$$\sum_i f_i^2 = N$$

objective function of spectral clustering (same - it's magic!)

$$\min_{\mathbf{f}} \mathbf{f}^\mathsf{T} \mathbf{L} \mathbf{f} \quad \text{s.t.} \quad f_i \in \mathbb{R}, \quad \mathbf{f} \perp \mathbf{1}_N, \quad \|\mathbf{f}\| = \sqrt{N}$$

Michal Valko

michal.valko@inria.fr

Inria & ENS Paris-Saclay, MVA

https://misovalko.github.io/mva-ml-graphs.html