

Calcul scientifique

CM 5

Uli Fahrenberg

Polytech Paris-Saclay

2025-26

Google PageRank

Google PageRank

(Ceci est une élaboration de ce que vous avez vu en TP 2.)

- Quand on cherche des infos sur l'internet, on aimerait bien que les résultats soient **ordonnés par importance**
- Mais qu'est-ce qu'est une bonne **mesure d'importance** ?

Google PageRank

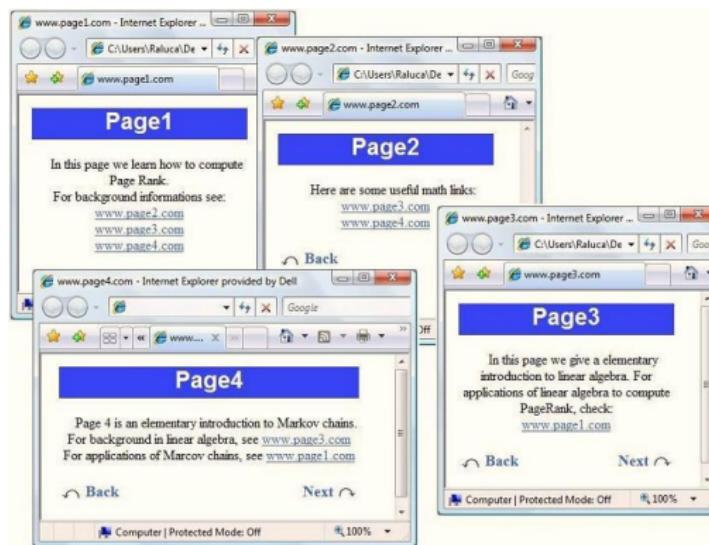
(Ceci est une élaboration de ce que vous avez vu en TP 2.)

- Quand on cherche des infos sur l'internet, on aimerait bien que les résultats soient **ordonnés par importance**
- Mais qu'est-ce qu'est une bonne **mesure d'importance** ?
- Avant google : le nombre de fois que des mots dans notre recherche apparaissent sur la page concernée (!)
- Google : une page est importante si cette importance est reconnue par beaucoup d'autres pages importantes à travers des liens
- Une **révolution** !



Importance

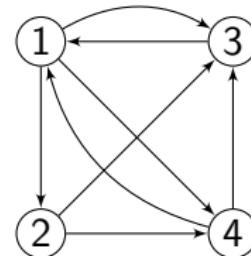
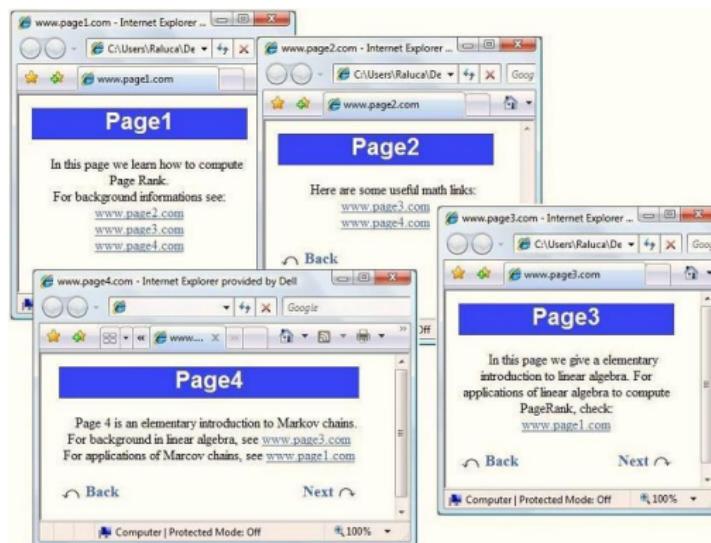
« une page est importante si cette importance est reconnue par beaucoup d'autres pages importantes à travers des liens »



C'est qui le plus important ?

Importance

« une page est importante si cette importance est reconnue par beaucoup d'autres pages importantes à travers des liens »

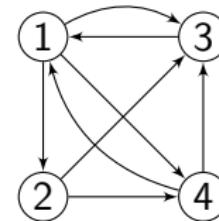
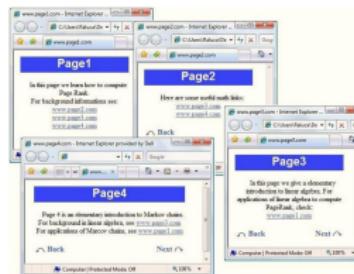


$$\begin{pmatrix} 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 \end{pmatrix}$$

C'est qui le plus important ?

Itération

$$\begin{pmatrix} 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 2 & 1 & 3 & 2 \end{pmatrix} =: A$$



- la page 3 a trois liens entrants ; si toutes pages sont de la même importance, alors la page 3 gagne :

$$(1 \ 1 \ 1 \ 1) A = (1 \ 1 \ 1 \ 1) \begin{pmatrix} 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 \end{pmatrix} = (2 \ 1 \ 3 \ 2)$$

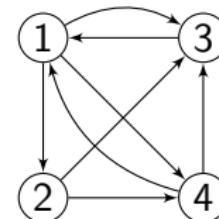
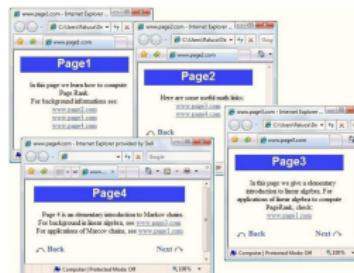
- mais les pages **ne sont pas** de la même importance !

$$(2 \ 1 \ 3 \ 2) A = (5 \ 2 \ 5 \ 3)$$

- moins clair ...

Itération !

$$A = \begin{pmatrix} 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 \end{pmatrix}$$



- Idée : **itérer** le calcul $v := v A$, commençant avec $v = (1 \ 1 \ 1 \ 1)$, et en **normalisant** chaque fois :

$$v := v A \quad v := v * 4 / \|v\|_1$$

- Cela converge assez rapidement :

$$\begin{pmatrix} 1 & 1 & 1 & 1 \end{pmatrix} \rightsquigarrow \begin{pmatrix} 1 & 0.5 & 1.5 & 1 \end{pmatrix} \rightsquigarrow \begin{pmatrix} 1.33 & 0.53 & 1.33 & 0.80 \end{pmatrix}$$

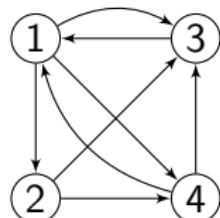
$$\dots \rightsquigarrow \begin{pmatrix} 1.16 & 0.59 & 1.35 & 0.90 \end{pmatrix}$$

- Au final oui, la page 3 est la plus importante

Modélisation

On utilise nos idées pour faire une propre modélisation.

- Normaliser le vecteur v à chaque itération donne l'idée de le voir comme un vecteur de **probabilités**
- Si une page i a plus de liens entrants, alors la probabilité de se trouver ici devrait être plus grande – c'est la composante v_i
- On modélise l'utilisateur comme un **agent aléatoire** qui va cliquer un lien sur une page à l'aléa ("random surfer")



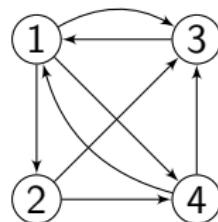
$$\begin{pmatrix} 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 \end{pmatrix}$$

 \rightsquigarrow

$$\begin{pmatrix} 0 & 1/3 & 1/3 & 1/3 \\ 0 & 0 & 1/2 & 1/2 \\ 1/1 & 0 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 \end{pmatrix}$$

- On va aussi **transposer** la matrice, car on aime mieux la multiplication $A v$ que $v A$

Modélisation : Random Surfer



$$A = \begin{pmatrix} 0 & 0 & 1 & 0.5 \\ 0.33 & 0 & 0 & 0 \\ 0.33 & 0.5 & 0 & 0.5 \\ 0.33 & 0.5 & 0 & 0 \end{pmatrix}$$

- A : matrice normalisée (et transposée) des liens
- Chaque colonne correspond à une page et indique les probabilités d'aller vers d'autres pages
- Une matrice stochastique de colonne
- Au début, notre random surfer a une probabilité uniforme d'être sur une page : $v_0 = (0.25 \ 0.25 \ 0.25 \ 0.25)^T$
- Itération $v_{n+1} := A v_n$
- On s'intéresse à la probabilité stationnaire $\lim_{n \rightarrow \infty} A^n v_0$
- Mais comment la calculer ? Qu'en est-il de sa convergence ?

Analyse : Chaînes de Markov

Définition

Une **chaîne de Markov** est une séquence X_0, X_1, \dots de variables aléatoires t.q. $\forall n \geq 1, \forall x_0, x_1, \dots, x_n :$

$$\begin{aligned}\Pr(X_n = x_n \mid X_0 = x_0, X_1 = x_1, \dots, X_{n-1} = x_{n-1}) \\ = \Pr(X_n = x_n \mid X_{n-1} = x_{n-1})\end{aligned}$$

- C. à d. X_n ne dépend que de son **passé immédiat** X_{n-1} (**« pas de mémoire »**)
- Si $\Pr(X_n = y \mid X_{n-1} = x) = \Pr(X_1 = y \mid X_0 = x)$ pour tout n, x, y , la chaîne est dite **homogène**
- Pour une chaîne homogène avec espace d'états fini, on peut définir $P_{ij} = \Pr(X_1 = i \mid X_0 = j)$
- Cela donne la **matrice de transition** $P = (P_{ij})$
- (Attention P_{ij} est la probabilité de transitionner **de j à i**)

Analyse : Probabilité stationnaire

Soit $X = X_0, X_1, \dots$ une chaîne de Markov homogène sur espace d'états fini, avec $P \in \mathbb{R}^{n \times n}$ la matrice de transition.

Définition

Un vecteur $v \in \mathbb{R}^n$ est une **probabilité stationnaire** pour la chaîne si $\sum v_i = 1$ et $Pv = v$.

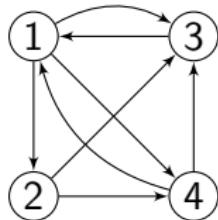
- Mesure la probabilité de se trouver dans des différents états **sur le long terme** / au limite

Théorème

Si X est irréductible et récurrent positif, alors

- *il existe une **unique** probabilité stationnaire v , et*
- *pour tout vecteur v_0 , $v = \lim_{n \rightarrow \infty} P^n v_0$.*

Analyse : Markov Surfer



$$A = \begin{pmatrix} 0 & 0 & 1 & 0.5 \\ 0.33 & 0 & 0 & 0 \\ 0.33 & 0.5 & 0 & 0.5 \\ 0.33 & 0.5 & 0 & 0 \end{pmatrix}$$

- A : matrice normalisée (et transposée) des liens
 - On prend un v_0 et s'intéresse au $\lim_{n \rightarrow \infty} A^n v_0$
 - Une chaîne de Markov homogène avec matrice de transition A
- ⇒ Si A est irréductible et récurrent positif, alors il y a un unique v avec $v = Av$, et $v = \lim_{n \rightarrow \infty} A^n v_0$ pour tout v_0
- **Irréductible** : tout état accessible depuis tout autre
 - **Récurrent positif** : tout état vu infiniment souvent avec probabilité > 0
 - (On va faire plus simple)

Simplification : Perron-Frobenius

Théorème (Perron-Frobenius version stochastique)

Soit A une matrice stochastique de colonne et strictement positive.

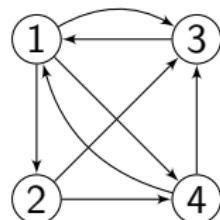
- $\lambda = 1$ est une valeur propre de A , et la valeur propre avec le plus grand module ;
- l'espace propre de $\lambda = 1$ est de dimension 1 ;
- l'unique vecteur stochastique propre correspondant à $\lambda = 1$ est la limite de $A^n v_0$ pour $n \rightarrow \infty$ et tout vecteur stochastique v_0 .

Simplification : Perron-Frobenius

Théorème (Perron-Frobenius version stochastique)

Soit A une matrice stochastique de colonne et strictement positive.

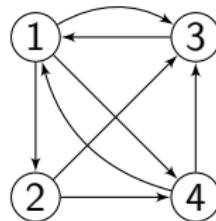
- $\lambda = 1$ est une valeur propre de A , et la valeur propre avec le plus grand module ;
- l'espace propre de $\lambda = 1$ est de dimension 1 ;
- l'unique vecteur stochastique propre correspondant à $\lambda = 1$ est la limite de $A^n v_0$ pour $n \rightarrow \infty$ et tout vecteur stochastique v_0 .



$$A = \begin{pmatrix} 0 & 0 & 1 & 0.5 \\ 0.33 & 0 & 0 & 0 \\ 0.33 & 0.5 & 0 & 0.5 \\ 0.33 & 0.5 & 0 & 0 \end{pmatrix}$$

Mais notre A n'est pas strictement positive !?

The Teleporting Random Surfer



$$A = \begin{pmatrix} 0 & 0 & 1 & 0.5 \\ 0.33 & 0 & 0 & 0 \\ 0.33 & 0.5 & 0 & 0.5 \\ 0.33 & 0.5 & 0 & 0 \end{pmatrix}$$

- On a envie d'appliquer Perron-Frobenius à notre matrice A , mais elle n'est pas strictement positive
- Solution : un **facteur d'amortissement** : soit $0 < p < 1$ et

$$M = (1 - p) A + p * \frac{1}{n} \begin{pmatrix} 1 & 1 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & 1 \end{pmatrix}$$

- On suit les liens avec probabilité $1 - p$, et avec probabilité p on recommence à un endroit aléatoire
- Typiquement $p = 0.15$
- Perron-Frobenius **s'applique** à M !

Conclusion

- Soit A la matrice normalisée (et transposée) des liens, de taille $n \times n$
- Soit $p = 0.15$ et $M = (1 - p)A + p * \frac{1}{n}B$, où B est la matrice avec 1 partout
- Alors l'importance de chaque page est donné par l'unique vecteur stochastique v pour lequel $A v = v$
- Et $v = \lim_{n \rightarrow \infty} A^n v_0$ pour n'importe vecteur stochastique v_0
- Comment **calculer** v , si $n > 10^9$?!

Conclusion

- Soit A la matrice normalisée (et transposée) des liens, de taille $n \times n$
- Soit $p = 0.15$ et $M = (1 - p)A + p * \frac{1}{n}B$, où B est la matrice avec 1 partout
- Alors l'importance de chaque page est donné par l'unique vecteur stochastique v pour lequel $A v = v$
- Et $v = \lim_{n \rightarrow \infty} A^n v_0$ pour n'importe vecteur stochastique v_0
- Comment **calculer** v , si $n > 10^9$?!
- Le calcul vecteur propre $v = A v$ va échouer
- Mais la **méthode des puissances** $v = \lim_{n \rightarrow \infty} A^n v_0$ converge assez vite
- “The Billion Dollar Eigenvector Calculation”

Jupyter notebooks

Jupyter notebooks

- ➊ Changement repère
- ➋ Vecteurs propres
- ➌ Systèmes d'équations et décomposition LU



That's all Folks!