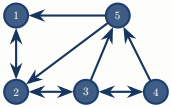
Chapitre 4

Les chaînes de Markov et l'algorithme PageRank

MAT-2901 — Mathématiques et technologie



Jérôme Soucy

Université Laval

Plan de l’exposé

1. Un internaute qui navigue au hasard

2. Rappels sur les probabilités

3. Un premier modèle : la météo à deux états

4. Les chaînes de Markov

5. Matrices stochastiques et états d’équilibre

6. L’algorithme PageRank

7. Calculer le PageRank : la méthode des puissances

8. Synthèse

Un internaute qui navigue au hasard

Quelles pages sont les plus importantes ?

- Une requête : des millions de pages contiennent les mots cherchés. Dans quel ordre les présenter ?
- Brin et Page (Stanford, fin des années 1990) : imaginons un internaute qui clique au hasard sur les liens, et mesurons la proportion du temps qu'il passe sur chaque page.
- Une page souvent visitée est une page importante : c'est l'idée de PageRank, et de Google.

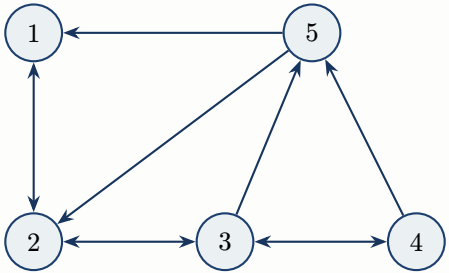


Figure 4.1 — Un mini-web de cinq pages. La page 2 reçoit quatre liens, la page 1 deux : où l'internaute passera-t-il le plus de temps ?

Une chaîne de Markov

Le déplacement de l'internaute est une chaîne de Markov : un phénomène aléatoire qui évolue par étapes, dont l'état à l'étape $n+1$ ne dépend que de l'état à l'étape n, et non de tout le passé.

Usages : météorologie, démographie, génétique, finance, reconnaissance de la parole, génération de texte.

Le chemin du chapitre :

1. rappels de probabilités ;
2. deux modèles : la météo à deux états, la migration ville–banlieue–campagne ;
3. chaînes de Markov, matrices de transition ;
4. le long terme : un vecteur propre de valeur propre 1 ;
5. PageRank, avec les difficultés du Web réel ;
6. le calcul par la méthode des puissances — et le retour du théorème de Banach.

Note historique 4.1 — Andreï Markov (1856–1922)



Vers 1875. Domaine public, Wikimedia Commons.

- Né à Riazan ; enfant de santé fragile, il se déplace avec des béquilles jusqu'à une dizaine d'années.
- Élève de Tchebychev à Saint-Pétersbourg, où il enseigne toute sa carrière.
- 1906 : ses chaînes, contre Nekrassov, qui soutenait que la loi des grands nombres exige l'indépendance (et en tirait des conclusions sur le libre arbitre). Markov montre qu'une dépendance « au précédent seulement » suffit.

Note historique 4.1 (suite)

- Ses chaînes étaient d'abord, pour lui, un contre-exemple théorique ; il ne les appliqua qu'une fois à des données réelles (note historique 4.2).
- Un esprit indépendant : en 1902, il proteste quand l'Académie, à la demande du tsar, annule l'élection de Maxime Gorki ; en 1912, il demande à être excommunié, par solidarité avec Tolstoï.
- En 1913, pendant que l'Empire fête le tricentenaire des Romanov, il organise plutôt le bicentenaire de la loi des grands nombres (Bernoulli, Ars conjectandi, 1713).
- Son fils, Andreï Andreïevitch Markov (1903–1979), sera lui aussi mathématicien, spécialiste de logique.

Rappels sur les probabilités

Définition 4.1 — Probabilité conditionnelle

Expérience aléatoire, univers (ensemble des résultats), événement (sous-ensemble de l'univers), probabilité $\mathbb{P}(A) \in [0,1]$. Une information partielle modifie une probabilité : s'il pleut ce matin, l'autobus est plus souvent en retard.

Définition 4.1 — Probabilité conditionnelle

Soient A, B deux événements, avec $\mathbb{P}(B) > 0$. La probabilité conditionnelle de A sachant B est

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$

Autrement dit $\mathbb{P}(A \cap B) = \mathbb{P}(A \mid B)\,\mathbb{P}(B)$: la règle « multiplier le long d'une branche » d'un diagramme en arbre.

Proposition 4.1 — Formule des probabilités totales

Proposition 4.1 — Formule des probabilités totales

Soient $B_1, \dots, B_N$ deux à deux incompatibles, de probabilités non nulles, de réunion l'univers. Pour tout événement A,

$$\mathbb{P}(A) = \mathbb{P}(A \mid B_1)\mathbb{P}(B_1) + \mathbb{P}(A \mid B_2)\mathbb{P}(B_2) + \cdots + \mathbb{P}(A \mid B_N)\mathbb{P}(B_N).$$

DÉMONSTRATION AU TABLEAU

Les $A \cap B_j$ sont deux à deux incompatibles, de réunion A ; additionner, puis appliquer la définition 4.1.

C'est la règle « additionner les branches qui mènent au même résultat ».

Exemple 4.1 — L'autobus du matin

Il pleut un matin sur trois. S'il pleut, l'autobus est en retard avec probabilité $0{,}4$; sinon, $0{,}1$. P : « il pleut » ; R : « en retard ».

$$\begin{aligned}\mathbb{P}(R) &= \mathbb{P}(R \mid P)\mathbb{P}(P) + \mathbb{P}\big(R \mid \overline{P}\big)\mathbb{P}\big(\overline{P}\big) \\ &= 0{,}4 \cdot \tfrac{1}{3} + 0{,}1 \cdot \tfrac{2}{3} = \tfrac{0{,}6}{3} = 0{,}2.\end{aligned}$$

L'autobus est en retard un matin sur cinq.

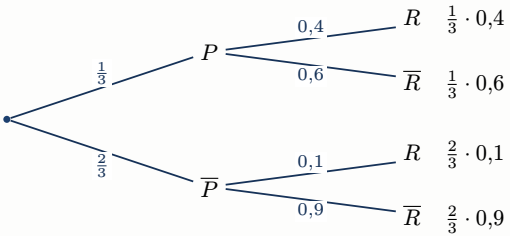


Figure 4.2

Définition 4.2 — Variable aléatoire discrète, espérance

Définition 4.2 — Variable aléatoire discrète

Une variable aléatoire discrète X associe un nombre à chaque résultat, et ne prend qu'un nombre fini (ou une suite) de valeurs $x_1, x_2, \ldots$ Sa loi : les $p_i = \mathbb{P}(X = x_i)$, positifs et de somme 1. Son espérance : $\mathbb{E}(X) = \sum_i x_i\, p_i$.

- Un dé : $\mathbb{E}(X) = \frac{1}{6}(1 + 2 + \cdots + 6) = 3{,}5$, qui n'est pas une valeur possible : c'est une moyenne à long terme.
- Linéarité : $\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y)$, même sans indépendance.
- Indicatrice $\mathbb{1}_A$ (vaut 1 si A se produit, 0 sinon) : $\mathbb{E}(\mathbb{1}_A) = \mathbb{P}(A)$.

Ces deux propriétés permettent de compter en moyenne sans connaître la loi d'une somme (exemple 4.2).

Définition 4.3 — Vecteur de probabilités

Ici, une variable aléatoire décrit l'état d'un système (le temps, un lieu de résidence, une page web), numéroté $1, \dots, N$: sa loi est un vecteur de $\mathbb{R}^N$.

Définition 4.3 — Vecteur de probabilités

Un vecteur $\mathbf{p} = (p_1, \dots, p_N)$ de $\mathbb{R}^N$ est un vecteur de probabilités si

$$p_i \geq 0 \text{ pour tout } i, \qquad p_1 + p_2 + \cdots + p_N = 1.$$

- Dans $\mathbb{R}^2$: le segment de $(1,0)$ à $(0,1)$; dans $\mathbb{R}^3$: le triangle de sommets $(1,0,0)$, $(0,1,0)$, $(0,0,1)$. C'est le simplexe de $\mathbb{R}^N$.
- Avec $\mathbf{e} = (1, 1, \dots, 1)$: la condition de somme s'écrit $\mathbf{e}^\mathsf{T}\mathbf{p} = 1$.

Pour la classe 4.1 — Les probabilités conditionnelles au secondaire

Pour la classe — Les probabilités conditionnelles au secondaire

- Diagrammes en arbre dès le primaire ; au deuxième cycle, espérance (moyenne pondérée, équité d'un jeu) et décisions à partir de probabilités conditionnelles.
- L'arbre joue le rôle que jouera la matrice de transition : les branches du second niveau portent des probabilités conditionnelles, et « multiplier le long d'une branche, additionner les branches » est la formule des probabilités totales.
- Mais un arbre devient vite illisible : avec deux états, $2^{10} = 1024$ branches après dix étapes. D'où la notation matricielle.

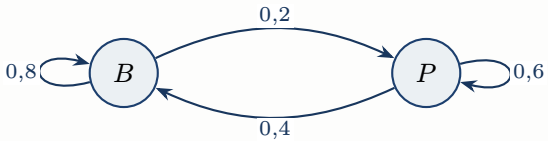
Un premier modèle : la météo à deux états

Figure 4.3 — Le diagramme de transition

Dans une ville imaginaire :

- après un beau jour : beau le lendemain avec probabilité $0{,}8$, pluie avec probabilité $0{,}2$;
- après un jour de pluie : beau avec probabilité $0{,}4$, pluie avec probabilité $0{,}6$.

Seul compte le temps de la veille (ni la saison, ni les jours précédents).



Les probabilités des flèches qui partent d'un même état ont une somme égale à 1.

Prévoir le temps des jours suivants

Il fait beau le jour 0 ; X_n : le temps dans n jours. $\mathbb{P}(X_1 = B) = 0{,}8$. Probabilités totales :

$$\mathbb{P}(X_2 = B) = 0{,}8 \cdot 0{,}8 + 0{,}4 \cdot 0{,}2 = 0{,}72.$$

En général :

$$\begin{cases} \mathbb{P}(X_{n+1} = B) = 0{,}8\,\mathbb{P}(X_n = B) + 0{,}4\,\mathbb{P}(X_n = P), \\ \mathbb{P}(X_{n+1} = P) = 0{,}2\,\mathbb{P}(X_n = B) + 0{,}6\,\mathbb{P}(X_n = P). \end{cases} \tag{4.1}$$

Un produit matriciel : avec $\mathbf{x}_n = (\mathbb{P}(X_n = B)\,;\, \mathbb{P}(X_n = P))$,

$$\mathbf{x}_{n+1} = M\mathbf{x}_n, \qquad M = \begin{pmatrix} 0{,}8 & 0{,}4 \\ 0{,}2 & 0{,}6 \end{pmatrix}, \qquad \mathbf{x}_n = M^n \mathbf{x}_0.$$

M est la matrice de transition ; m_{ij} est la probabilité de passer de l'état j à l'état i. Prévoir le temps : calculer une puissance de matrice.

Remarque 4.1 — Lignes ou colonnes ?

Remarque 4.1

- Ici, les probabilités qui partent d'un même état forment une colonne : $\mathbf{x}_{n+1} = M\mathbf{x}_n$, comme les vecteurs de l'annexe D. La somme de chaque colonne de M vaut 1.
- Beaucoup d'ouvrages de probabilités (et Wikipédia) font l'inverse : une ligne, une loi en vecteur ligne , et $_{n+1} = {}_n P$.
- Équivalent, puisque $P = M^{\mathsf{T}}$ — mais il faut repérer la convention d'une source avant d'en reprendre les calculs.

Tableau 4.1 — L'évolution de la loi

$\mathbb{P}(X_n = B)$, selon le temps du jour 0 :

n	0	1	2	3	4	5	6	7
beau le jour 0	1	$0{,}8$	$0{,}72$	$0{,}688$	$0{,}675\,2$	$0{,}670\,1$	$0{,}668\,0$	$0{,}667\,2$
pluie le jour 0	0	$0{,}4$	$0{,}56$	$0{,}624$	$0{,}649\,6$	$0{,}659\,8$	$0{,}663\,9$	$0{,}665\,6$

Dans les deux cas, la loi se rapproche rapidement de $\left(\frac{2}{3}\,;\ \frac{1}{3}\right) \approx (0{,}666\,7\,;\ 0{,}333\,3)$: après une semaine, le temps du jour 0 n'a presque plus d'influence.

L'état d'équilibre

Si $\mathbf{x}_n \to \mathbf{q}$, alors $\mathbf{q} = M\mathbf{q}$: un état d'équilibre (ou loi stationnaire).

- Algèbre linéaire : un vecteur propre de M associé à la valeur propre 1.
- Chapitre 2 : un point fixe de $\mathbf{x} \mapsto M\mathbf{x}$.

$(M - I)\mathbf{q} = \mathbf{0}$: $-0{,}2\,q_1 + 0{,}4\,q_2 = 0$ (deux fois), d'où $q_1 = 2q_2$; avec $q_1 + q_2 = 1$:

$$\mathbf{q} = \left(\tfrac{2}{3}\,;\ \tfrac{1}{3}\right).$$

À long terme, il fait beau deux jours sur trois.

Pourquoi la convergence est-elle si rapide ?

Polynôme caractéristique : $\lambda^2 - 1{,}4\,\lambda + 0{,}4 = (\lambda - 1)(\lambda - 0{,}4)$. Seconde valeur propre $\lambda_2 = 0{,}4$, vecteur propre $\mathbf{v}_2 = (1, -1)$.

Pour $\mathbf{x}_0 = (1, 0)$: $\mathbf{x}_0 = \mathbf{q} + \frac{1}{3}\mathbf{v}_2$, donc

$$\mathbf{x}_n = M^n \mathbf{q} + \tfrac{1}{3} M^n \mathbf{v}_2 = \mathbf{q} + \tfrac{1}{3}(0{,}4)^n\, \mathbf{v}_2.$$

- L'écart à l'équilibre est multiplié par $0{,}4$ chaque jour : environ $0{,}000\,5$ après une semaine (tableau 4.1).
- La seconde valeur propre gouverne la vitesse de convergence (on y revient en section 7).

Exemple 4.2 — Combien de beaux jours dans la semaine ?

Il fait beau aujourd'hui. Y : le nombre de beaux jours parmi les sept suivants. $Y = \mathbb{1}_1 + \cdots + \mathbb{1}_7$, où $\mathbb{1}_n$ indique « beau le jour n ». Pas d'indépendance, mais la linéarité suffit :

$$\mathbb{E}(Y) = \sum_{n=1}^{7} \mathbb{P}(X_n = B) \approx 0{,}8 + 0{,}72 + 0{,}688 + \cdots + 0{,}667\,2 \approx 4{,}89.$$

Un peu plus que les $7 \cdot \frac{2}{3} \approx 4{,}67$ jours de l'équilibre, puisque la semaine commence par un beau jour.

Programme 4.1 — Le calcul avec Python

```python
import numpy as np
np.set_printoptions(precision=4,
                    suppress=True)

# colonnes : départ d'un beau jour,
#            puis d'un jour de pluie
M = np.array([[0.8, 0.4],
              [0.2, 0.6]])
x0 = np.array([1.0, 0.0])   # il fait beau

for n in [1, 2, 5, 10, 20]:
    print(n, np.linalg.matrix_power(M, n) @ x0)

valeurs, vecteurs = np.linalg.eig(M)
print("valeurs propres :", valeurs)
v = vecteurs[:, np.argmax(valeurs)]  # associé à 1
print("équilibre :", v / v.sum())    # somme = 1
```

- `np.linalg.eig` renvoie des vecteurs propres de norme 1 : on divise par la somme des composantes pour obtenir un vecteur de probabilités.
- Sortie : équilibre $(0{,}666\,7\,;\ 0{,}333\,3)$, valeurs propres 1 et $0{,}4$.
- Simulation sur dix mille jours (programme 4.3 des notes) : proportion de beaux jours $0{,}664\,4$, proche de $\frac{2}{3}$.

Deux lectures de l'équilibre : limite des probabilités, et proportion du temps passé dans chaque état — au cœur de PageRank.

Outil 4.1 — Puissances de matrices sans programmer

Outil — Puissances de matrices sans programmer

- WolframAlpha : `{{0.8, 0.4}, {0.2, 0.6}}^10` calcule M^{10} ; `eigenvectors {{0.8, 0.4}, {0.2, 0.6}}` donne les vecteurs propres.
- GeoGebra : `M = {{0.8, 0.4}, {0.2, 0.6}}`, puis `M^n` avec un curseur n : les deux colonnes de M^n deviennent presque égales à $\left(\frac{2}{3}\,;\,\frac{1}{3}\right)$.
- Si toutes les colonnes de M^n sont proches de $\mathbf{q}$, alors $M^n\mathbf{x}_0$ est proche de $\mathbf{q}$ pour tout vecteur de probabilités $\mathbf{x}_0$: le point de départ est oublié.
- Un tableur convient aussi : une ligne par jour, calculée à partir de la précédente par (4.1).

Les chaînes de Markov

Définition 4.4 — Chaîne de Markov

Définition 4.4 — Chaîne de Markov

Une suite de variables aléatoires $X_0, X_1, X_2, \dots$ à valeurs dans un ensemble fini d'états $\{1, \dots, N\}$ est une chaîne de Markov si, pour tout n et tous états $i, j, i_0, \dots, i_{n-1}$,

$$\mathbb{P}(X_{n+1} = i \mid X_n = j,\ X_{n-1} = i_{n-1},\ \dots,\ X_0 = i_0) = \mathbb{P}(X_{n+1} = i \mid X_n = j)$$

(propriété de Markov). Elle est homogène si cette probabilité ne dépend pas de n ; on la note m_{ij}, et $M = (m_{ij})$ est la matrice de transition.

- Toutes nos chaînes sont homogènes.
- Le hasard peut dépendre du passé, mais toute l'information utile est contenue dans l'état présent.
- Pile ou face : un cas extrême, chaque colonne vaut $\left(\frac{1}{2}, \frac{1}{2}\right)$.

Proposition 4.2 — Évolution de la loi

Soit (X_n) une chaîne de Markov homogène de matrice M, et $\mathbf{x}_n$ la loi de X_n. Alors $\mathbf{x}_{n+1} = M\mathbf{x}_n$ pour tout n, et donc $\mathbf{x}_n = M^n\mathbf{x}_0$.

DÉMONSTRATION AU TABLEAU

Probabilités totales avec $B_j = \{X_n = j\}$: $\mathbb{P}(X_{n+1} = i) = \sum_j m_{ij}\,\mathbb{P}(X_n = j)$, la i^{e} composante de $M\mathbf{x}_n$; puis récurrence.

Avec $\mathbf{x}_0 = \mathbf{e}_j$: la j^{e} colonne de M^n est la loi de X_n partant de j. Le coefficient $(M^n)_{ij}$ est la probabilité de passer de j à i en n étapes.

Pour la classe 4.2 — Graphes et matrices au secondaire

Pour la classe — Graphes et matrices au secondaire

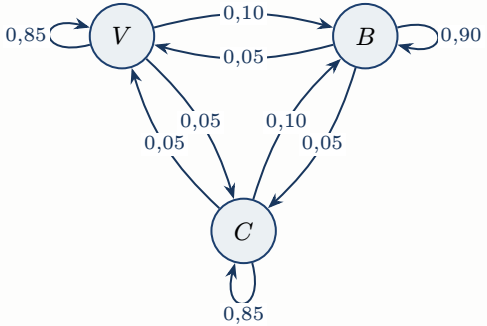
- Le diagramme de transition est un graphe orienté valué : sommets = états, arêtes orientées portant un poids.
- La progression des apprentissages prévoit une introduction à la théorie des graphes au deuxième cycle (graphes orientés et valués).
- En Technico-sciences, les matrices sont un registre de représentation ; on y fait des multiplications matricielles.
- Une chaîne de Markov réunit les deux registres : le graphe pour décrire, la matrice pour calculer. Passer de l'un à l'autre oblige à préciser ce que représente chaque ligne et chaque colonne.

Les chaînes de Markov

Ville, banlieue et campagne

Figure 4.4 — Le modèle de migration

Chaque année : en ville, $85\,\%$ restent, $10\,\%$ vont en banlieue, $5\,\%$ à la campagne ; en banlieue, $90\,\%$ restent, $5\,\%$ vont en ville et $5\,\%$ à la campagne ; à la campagne, $85\,\%$ restent, $10\,\%$ vont en banlieue, $5\,\%$ en ville. Population constante, taux fixes.



$$M = \begin{pmatrix} 0{,}85 & 0{,}05 & 0{,}05 \\ 0{,}10 & 0{,}90 & 0{,}10 \\ 0{,}05 & 0{,}05 & 0{,}85 \end{pmatrix}$$

colonnes : départ de V, de B, de C
lignes : arrivée en V, en B, en C

Tableau 4.2 — L'évolution de la répartition

$\mathbf{x}_0 = (0{,}5\,;\ 0{,}3\,;\ 0{,}2)$, et $\mathbf{x}_1 = M\mathbf{x}_0 = (0{,}45\,;\ 0{,}34\,;\ 0{,}21)$.

n (années)	0	1	2	5	10	20	50
ville	0,50	0,45	0,41	0,331 9	0,276 8	0,252 9	0,250 0
banlieue	0,30	0,34	0,372	0,434 5	0,478 5	0,497 7	0,500 0
campagne	0,20	0,21	0,218	0,233 6	0,244 6	0,249 4	0,250 0

Gauss–Jordan sur $M - I$: le noyau est engendré par $(1, 2, 1)$, d'où $\mathbf{q} = \left(\frac{1}{4}, \frac{1}{2}, \frac{1}{4}\right)$. À long terme, la moitié de la population vit en banlieue.

Exemple 4.3 — Une formule exacte pour $\mathbf{x}_n$

Soit $Q = \mathbf{q}\,\mathbf{e}^{\mathsf{T}}$ (trois colonnes égales à $\mathbf{q}$). On vérifie que

$$M = 0{,}8\,I + 0{,}2\,Q$$

(première colonne : $(0{,}8 + 0{,}05\,;\ 0{,}1\,;\ 0{,}05)$). Chaque année, on garde son lieu avec probabilité $0{,}8$; avec probabilité $0{,}2$, on « tire au sort » un milieu selon $\mathbf{q}$.

- Si $\mathbf{e}^{\mathsf{T}}\mathbf{v} = 0$: $Q\mathbf{v} = \mathbf{0}$, donc $M\mathbf{v} = 0{,}8\,\mathbf{v}$ (tout le plan $v_1 + v_2 + v_3 = 0$ est propre).
- $\mathbf{x}_0 - \mathbf{q}$ est dans ce plan, d'où
$$\mathbf{x}_n = \mathbf{q} + (0{,}8)^n(\mathbf{x}_0 - \mathbf{q}).$$
- L'écart diminue de $20\,\%$ par année ; $(0{,}8)^{10} \approx 0{,}107$. Ce « tirage au sort » reviendra dans la matrice Google.

Les chaînes de Markov

Et si l'avenir dépendait de plusieurs jours ?

Une mémoire de deux jours

hier	aujourd'hui	$\mathbb{P}$(beau demain)
B	B	0,9
P	B	0,7
B	P	0,5
P	P	0,3

(X_n) n'est plus une chaîne de Markov. Mais le couple $Y_n = (X_{n-1}, X_n)$ en est une, à quatre états (BB, PB, BP, PP) :

$$M_2 = \begin{pmatrix} 0{,}9 & 0{,}7 & 0 & 0 \\ 0 & 0 & 0{,}5 & 0{,}3 \\ 0{,}1 & 0{,}3 & 0 & 0 \\ 0 & 0 & 0{,}5 & 0{,}7 \end{pmatrix}.$$

- Équilibre $\mathbf{q} = (\frac{21}{32}, \frac{3}{32}, \frac{3}{32}, \frac{5}{32})$; beau temps à long terme : $\frac{24}{32} = \frac{3}{4}$.
- En général, une mémoire de k états donne N^k états. Les premiers générateurs de texte : l'état est formé des k derniers mots ; plus k est grand, plus le texte est cohérent, mais le nombre d'états explose.

Note historique 4.2 — Les voyelles d'Eugène Onéguine

En 1913, Markov applique son modèle aux $20\,000$ premières lettres du roman en vers de Pouchkine, avec deux états : voyelle ou consonne.

- $8\,638$ voyelles ; une voyelle est suivie d'une voyelle $1\,104$ fois, soit une fréquence de $0{,}128$; une consonne est suivie d'une voyelle avec une fréquence de $0{,}663$.
- Matrice de transition $\begin{pmatrix} 0{,}128 & 0{,}663 \\ 0{,}872 & 0{,}337 \end{pmatrix}$.
- Équilibre des voyelles : $0{,}663/(0{,}872 + 0{,}663) \approx 0{,}432$, exactement la proportion $8\,638/20\,000$ du texte.

Ce calcul, fait à la main, est la première application d'une chaîne de Markov à des données réelles.

Matrices stochastiques et états d’équilibre

Définition 4.5 et proposition 4.3 — Matrices stochastiques

Dans les deux modèles, la loi converge vers un unique équilibre, quel que soit le départ. Est-ce toujours le cas ?

Définition 4.5 — Matrice stochastique

Une matrice carrée M est stochastique si ses coefficients sont positifs ou nuls et si chaque colonne a une somme égale à 1 : chaque colonne est un vecteur de probabilités, et $\mathbf{e}^\mathsf{T} M = \mathbf{e}^\mathsf{T}$.

Toute matrice de transition est stochastique, et réciproquement.

Proposition 4.3 — Une matrice stochastique conserve les lois

Si M est stochastique et $\mathbf{v}$ un vecteur de probabilités, alors $M\mathbf{v}$ (et donc $M^n\mathbf{v}$) est un vecteur de probabilités.

Lecture matricielle : $\mathbf{e}^\mathsf{T}(M\mathbf{v}) = (\mathbf{e}^\mathsf{T}M)\mathbf{v} = \mathbf{e}^\mathsf{T}\mathbf{v} = 1$ (le reste : un exercice du chapitre).

Théorèmes 4.4 et 4.5 — La valeur propre 1

Théorème 4.4 — Valeur propre d'une matrice stochastique

Si M est stochastique, alors 1 est une valeur propre de M.

DÉMONSTRATION AU TABLEAU

M et $M^\intercal$ ont le même polynôme caractéristique ; et $M^\intercal \mathbf{e} = \mathbf{e}$ (transposer $\mathbf{e}^\intercal M = \mathbf{e}^\intercal$).

Ce vecteur propre a-t-il des composantes positives ? Perron–Frobenius (admis) :

Théorème 4.5 — Perron–Frobenius, cas stochastique

Soit M stochastique. (i) Toute valeur propre λ (réelle ou complexe) vérifie $|\lambda| \leq 1$. (ii) Il existe un vecteur de probabilités $\mathbf{q}$ tel que $M\mathbf{q} = \mathbf{q}$.

Toute chaîne à un nombre fini d'états a au moins un équilibre.

Exemples 4.4 et 4.5 — Deux chaînes qui ne convergent pas

- I (taille 2) : chaque état est conservé. Tout vecteur de probabilités est un équilibre, et $\mathbf{x}_n = \mathbf{x}_0$: équilibre non unique, limite qui dépend du départ. Cela arrive dès que deux groupes d'états ne communiquent pas.
- $S = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$: l'état change à chaque étape. Unique équilibre $\left(\frac{1}{2}, \frac{1}{2}\right)$, mais depuis l'état 1 la loi alterne entre $(1, 0)$ et $(0, 1)$.
  - Valeurs propres 1 et -1 : la composante selon -1 change de signe au lieu de disparaître. Chaîne périodique.
  - La proportion du temps passé dans chaque état converge, elle, vers $\frac{1}{2}$.

Dans les deux cas, 1 n'est pas la seule valeur propre de module 1.

Définition 4.6 — Matrice stochastique régulière

Définition 4.6 — Matrice stochastique régulière

Une matrice stochastique M est régulière s'il existe un entier $k \geq 1$ tel que tous les coefficients de M^k soient strictement positifs.

- Après k étapes, tous les états sont accessibles depuis tous les états.
- Météo et migration : régulières dès $k = 1$.
- I et S ne le sont pas : leurs puissances valent I ou S.

Théorème 4.6 — Convergence vers l'équilibre

Théorème 4.6 — Convergence vers l'équilibre

Soit M une matrice stochastique régulière de taille N.

(i) 1 est une valeur propre simple de M, et toute autre valeur propre vérifie $|\lambda| < 1$;

(ii) il existe un unique vecteur de probabilités $\mathbf{q}$ tel que $M\mathbf{q} = \mathbf{q}$, à composantes strictement positives ;

(iii) pour tout vecteur de probabilités $\mathbf{x}_0$, $\mathbf{x}_n = M^n \mathbf{x}_0 \to \mathbf{q}$.

DÉMONSTRATION AU TABLEAU

(i), (ii) : Perron–Frobenius, admis. (iii) si M est diagonalisable : $\mathbf{x}_0 = \alpha_1 \mathbf{q} + \sum \alpha_i \mathbf{v}_i$, donc $\mathbf{x}_n = \alpha_1 \mathbf{q} + \sum \alpha_i \lambda_i^n \mathbf{v}_i \to \alpha_1 \mathbf{q}$; la somme des composantes reste 1 (prop. 4.3), d'où $\alpha_1 = 1$.

La vitesse : le plus grand $|\lambda_i|$, $i \geq 2$. Météo : $0{,}4$; migration : $0{,}8$, plus lent (tableaux 4.1 et 4.2).

Remarque 4.2 — Proportion du temps et temps de retour

Remarque 4.2

- Première lecture (théorème 4.6) : $\mathbb{P}(X_n = i) \to q_i$.
- Deuxième lecture, sur une seule trajectoire : la proportion des étapes passées dans l'état i tend vers q_i, avec probabilité 1 — le théorème ergodique (admis), une loi des grands nombres pour des étapes dépendantes. Vrai aussi pour les chaînes périodiques, si l'équilibre est unique.
- Troisième lecture : la chaîne revient en moyenne dans l'état i toutes les $1/q_i$ étapes. Météo : un jour de pluie tous les 3 jours, un beau jour tous les $1{,}5$ jour.

Pour la classe 4.3 — Simuler une chaîne de Markov en classe

Pour la classe — Simuler une chaîne de Markov en classe

- Probabilités fréquentielles et théoriques : au cœur de l'enseignement dès le premier cycle.
- Dé à dix faces : s'il fait beau, « pluie » sur 1 et 2 ; s'il pleut, « beau » sur 1 à 4. Chaque équipe simule un mois ; on met en commun.
- La proportion de beaux jours se rapproche de $\frac{2}{3}$. Question non évidente : pourquoi ne dépend-elle pas du premier jour ?
- Les jours de pluie arrivent en séquences : l'occasion de discuter la conception erronée d'un hasard « bien mélangé ».

L’algorithme PageRank

Contenu et popularité

Un moteur de recherche classe les pages selon deux types de critères :

- le contenu : présence des mots cherchés, position dans le titre ou le texte, proximité ;
- la popularité, mesurée par la structure des liens du Web.

PageRank (Brin et Page, 1998) relève du second type.

L'internaute aléatoire : sur une page qui contient k liens, il clique sur l'un d'eux, chacun avec probabilité $1/k$. La proportion du temps qu'il passe sur une page mesure son importance.

L’algorithme PageRank

L’internaute aléatoire

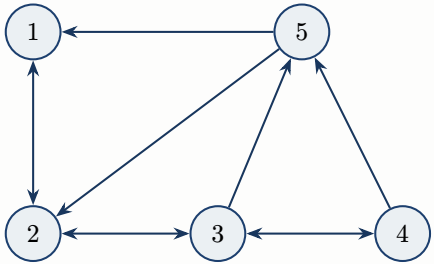
La matrice de transition du mini-web

Départ sur la page 3 (trois liens : vers 2, 4, 5). Pour être sur la page 1 après deux clics, il faut passer par 2 ou par 5 :

$$\mathbb{P}(X_2 = 1) = \tfrac{1}{2} \cdot \tfrac{1}{3} + \tfrac{1}{2} \cdot \tfrac{1}{3} = \tfrac{1}{3}.$$

La page visitée est une chaîne de Markov; $a_{ij} = 1/k_j$ si la page j (avec k_j liens) pointe vers i, 0 sinon :

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



Colonnes : liens qui sortent d'une page.
Lignes : liens qui y entrent.

Programme 4.5 — Le PageRank du mini-web

```python
def matrice_transition(liens, N):
    """liens[j] : pages vers lesquelles pointe j."""
    A = np.zeros((N, N))
    for j, sorties in liens.items():
        for i in sorties:
            A[i - 1, j - 1] = 1 / len(sorties)
    return A

liens = {1: [2], 2: [1, 3], 3: [2, 4, 5],
         4: [3, 5], 5: [1, 2]}
A = matrice_transition(liens, 5)
valeurs, vecteurs = np.linalg.eig(A)
k = np.argmin(abs(valeurs - 1))
q = np.real(vecteurs[:, k])
q = q / q.sum()
print("fois 55 :", 55 * q)  # [13. 20. 12. 4. 6.]
```

- `eig` peut renvoyer des complexes (matrice non symétrique) : `np.real`.
- On ne sait pas où se trouve la valeur propre 1 dans la liste : on cherche la plus proche.
- Avec $\mathbf{p}_0 = (0, 0, 1, 0, 0)$:
  $\mathbf{p}_1 = (0, \frac{1}{3}, 0, \frac{1}{3}, \frac{1}{3})$,
  $\mathbf{p}_2 = (\frac{1}{3}, \frac{1}{6}, \frac{1}{3}, 0, \frac{1}{6})$.

Le PageRank du mini-web

$$\mathbf{q} = \tfrac{1}{55}\,(13, 20, 12, 4, 6).$$

- A est régulière (coefficients de A^5 tous > 0) : le théorème 4.6 s'applique, quel que soit le départ.
- Classement : page 2, page 1, page 3, page 5, page 4.

La deuxième ligne de $\mathbf{q} = A\mathbf{q}$:

$$q_2 = 1 \cdot q_1 + \tfrac{1}{3}\, q_3 + \tfrac{1}{2}\, q_5 = \tfrac{13}{55} + \tfrac{1}{3} \cdot \tfrac{12}{55} + \tfrac{1}{2} \cdot \tfrac{6}{55} = \tfrac{20}{55}.$$

- L'importance d'une page est la somme des « votes » des pages qui pointent vers elle ; chaque page partage son PageRank également entre ses liens sortants.
- La page 1 (deux liens entrants) devance la page 3 (deux aussi) : l'un de ses liens vient de la page 2, la plus importante.
- Circulaire en apparence ; l'algèbre linéaire en fait la solution d'un système.

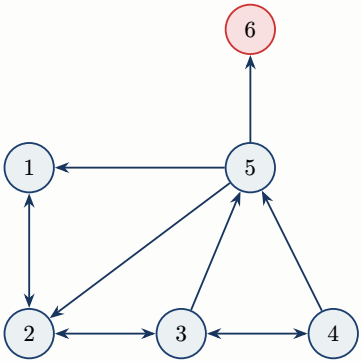
Note historique 4.3 — Le gogol et un vecteur propre à 25 milliards

- Le projet de Page et Brin sur les liens du Web donne naissance à Google (article de 1998).
- Google vient de googol (gogol en français) : le nombre 10^{100}. Selon Edward Kasner (Mathematics and the Imagination, 1940), le mot aurait été inventé par son neveu de neuf ans, Milton Sirotta.
- Kurt Bryan et Tanya Leise ont popularisé l'analyse mathématique de l'algorithme sous le titre « The $25,000,000,000 eigenvector ». Tout repose sur les notions de ce chapitre.

L’algorithme PageRank

Les pages sans liens sortants

Figure 4.5 — Le mini-web 2



PDF, images, pages en construction : la page 6 n'a aucun lien. La sixième colonne de A est nulle : A n'est plus stochastique, la probabilité « s'échappe ».

Correction : depuis une page sans lien, l'internaute choisit une page au hasard parmi les N (probabilité $1/N$ chacune). Avec $\mathbf{E}$, l'indicatrice des pages sans lien ($\mathbf{E} = (0,0,0,0,0,1)$ ici) :

$$B = A + \frac{1}{N}\,\mathbf{e}\,\mathbf{E}^{\mathsf{T}}. \tag{4.2}$$

Le PageRank du mini-web 2

$$B = \begin{pmatrix} 0 & \frac{1}{2} & 0 & 0 & \frac{1}{3} & \frac{1}{6} \\ 1 & 0 & \frac{1}{3} & 0 & \frac{1}{3} & \frac{1}{6} \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 & \frac{1}{6} \\ 0 & 0 & \frac{1}{3} & 0 & 0 & \frac{1}{6} \\ 0 & 0 & \frac{1}{3} & \frac{1}{2} & 0 & \frac{1}{6} \\ 0 & 0 & 0 & 0 & \frac{1}{3} & \frac{1}{6} \end{pmatrix}$$

stochastique et régulière.

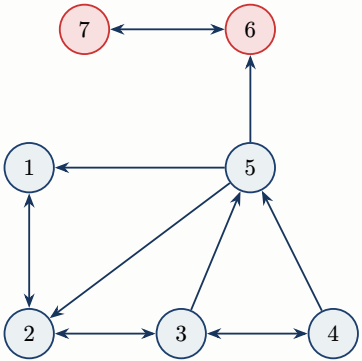
$$\mathbf{q} = \tfrac{1}{127}\,(27, 42, 27, 10, 15, 6) \approx (0{,}213\,;\ 0{,}331\,;\ 0{,}213\,;\ 0{,}079\,;\ 0{,}118\,;\ 0{,}047).$$

Classement : page 2, pages 1 et 3 à égalité, page 5, page 4, page 6. Peu de changements : la page 1 rejoint la page 3, en partie parce que la page 5 partage maintenant son vote entre trois liens.

L’algorithme PageRank

Les régions piégeantes et la téléportation

Figure 4.6 — Le mini-web 3 : une région piégeante



Une page 7, et des liens $6 \leftrightarrow 7$. On entre dans la région $\{6, 7\}$ par la page 5, mais on ne peut plus en sortir.

- A est stochastique mais pas régulière.
- Unique équilibre $\mathbf{q} = (0, 0, 0, 0, 0, \frac{1}{2}, \frac{1}{2})$: les pages 1 à 5 ont un PageRank nul. Inutilisable.
- La loi ne converge même pas (aller-retour entre 6 et 7, comme l'exemple 4.5) ; -1 est valeur propre de A.
- Sur le Web réel, de telles régions se forment spontanément, ou délibérément pour accaparer le PageRank.

La téléportation

À chaque étape, l'internaute suit un lien avec probabilité β, et avec probabilité $1-\beta$ saute vers une page choisie au hasard parmi les N (il tape une nouvelle adresse). $\beta \in]0,1[$ est le facteur d'amortissement ; Brin et Page utilisaient $\beta = 0{,}85$.

$$G = \beta B + (1-\beta)\, U, \qquad U = \frac{1}{N}\,\mathbf{e}\,\mathbf{e}^{\mathsf{T}} \quad \text{(tous les coefficients valent } 1/N\text{)}. \tag{4.3}$$

C'est le « tirage au sort » de l'exemple 4.3, avec la loi uniforme.

Proposition 4.7 — Propriétés de la matrice G

Si $0 < \beta < 1$, G est stochastique à coefficients strictement positifs ; elle est donc régulière, et a un unique équilibre $\mathbf{q}$, à composantes strictement positives.

DÉMONSTRATION AU TABLEAU

Somme d'une colonne : $\beta \cdot 1 + (1-\beta)\cdot 1 = 1$; chaque coefficient est au moins $(1-\beta)/N > 0$; $k = 1$.

Programme 4.7 — La matrice Google

```python
def matrice_google(liens, N, beta=0.85):
    A = matrice_transition(liens, N)
    E = (A.sum(axis=0) == 0).astype(float)
    e = np.ones(N)
    B = A + np.outer(e, E) / N    # pages sans sortie
    U = np.outer(e, e) / N        # téléportation
    return beta * B + (1 - beta) * U

def equilibre(M):
    valeurs, vecteurs = np.linalg.eig(M)
    k = np.argmin(abs(valeurs - 1))
    q = np.real(vecteurs[:, k])
    return q / q.sum()

liens3 = {1: [2], 2: [1, 3], 3: [2, 4, 5],
          4: [3, 5], 5: [1, 2, 6],
          6: [7], 7: [6]}
G3 = matrice_google(liens3, 7)
print(equilibre(G3))
```

- `np.outer(u, v)` : le produit $\mathbf{u}\,\mathbf{v}^{\mathsf{T}}$.
- `beta=1` : seulement la correction des pages sans lien (mini-web 2).
- Avec $\beta = 0{,}85$, classement : 6, 7, 2, 3, 1, 5, 4.
- La région piégeante reste en tête, sans monopoliser le classement.

Tableau 4.3 — Le choix de β

PageRank du mini-web 3 ; dernière colonne : la part de la région piégeante.

β	q_1	q_2	q_3	q_4	q_5	q_6	q_7	$q_6 + q_7$
0,5	0,137	0,183	0,141	0,095	0,119	0,169	0,156	0,325
0,85	0,122	0,182	0,123	0,056	0,080	0,225	0,212	0,437
0,99	0,031	0,048	0,031	0,012	0,017	0,432	0,429	0,861

- β proche de 1 : fidèle à la structure des liens, mais la région piégeante accapare le PageRank.
- β petit : le classement se rapproche de la loi uniforme et perd son information.

L'algorithme PageRank

La matrice Google

La matrice Google

Les deux corrections réunies :

$$G = \beta B + (1-\beta)U = \beta A + \frac{1}{N}\,\mathbf{e}\,(\beta \mathbf{E}^{\intercal} + (1-\beta)\,\mathbf{e}^{\intercal}),$$

et le PageRank est l'unique vecteur de probabilités tel que $G\mathbf{q} = \mathbf{q}$.

- Web réel : N en milliards ; G aurait plus de 10^{18} coefficients, tous non nuls. Impossible à stocker.
- Inutile : il suffit de A, $\mathbf{E}$ et β.
- A est creuse : quelques dizaines de liens par page au plus. On stocke les coefficients non nuls et leurs positions, c'est-à-dire la liste des liens.
- Reste à calculer $\mathbf{q}$ sans jamais former G.

Calculer le PageRank : la méthode des puissances

Algorithme 4.1 — Méthode des puissances

`np.linalg.eig` calcule toutes les valeurs et tous les vecteurs propres d'une matrice pleine : inutilisable. Le théorème 4.6 suggère plutôt d'appliquer G de façon répétée.

Algorithme 4.1 — Méthode des puissances pour le PageRank

Données : A creuse, $\mathbf{E}$, β, une tolérance $\varepsilon > 0$.

1. $\mathbf{w}_0 = \frac{1}{N}\,\mathbf{e}$ (la loi uniforme).
2. Pour $k = 1, 2, \ldots$: $\mathbf{w}_k = \beta A \mathbf{w}_{k-1} + \dfrac{\beta s_{k-1} + 1 - \beta}{N}\,\mathbf{e}$, où $s_{k-1} = \mathbf{E}^{\mathsf{T}}\mathbf{w}_{k-1}$.
3. S'arrêter si $d_1(\mathbf{w}_k, \mathbf{w}_{k-1}) < \varepsilon$; renvoyer $\mathbf{w}_k$.

- Étape 2 = $G\mathbf{w}_{k-1}$, puisque $\mathbf{e}^{\mathsf{T}}\mathbf{w} = 1$; s est la masse sur les pages sans lien.
- Coût : le produit $A\mathbf{w}$, proportionnel au nombre de liens.
- $d_1(\mathbf{x}, \mathbf{y}) = \sum |x_i - y_i|$: la taxi-distance du chapitre 2, dans $\mathbb{R}^N$.

Programme 4.9 — La méthode des puissances

```python
def pagerank(liens, N, beta=0.85, tol=1e-8):
    """Méthode des puissances, sans construire G."""
    A = matrice_transition(liens, N)  # creuse
    sans_sortie = (A.sum(axis=0) == 0)  # le vecteur E
    w = np.ones(N) / N                  # départ : loi uniforme
    for k in range(1, 1000):
        s = w[sans_sortie].sum()        # le produit E^T w
        w_suiv = beta * (A @ w) + (beta*s + 1 - beta) / N
        if np.abs(w_suiv - w).sum() < tol:  # d_1
            return w_suiv, k
        w = w_suiv

q, k = pagerank(liens3, 7)     # 93 itérations
```

On retrouve le PageRank du mini-web 3. Ajouter un nombre à un vecteur l'ajoute à chaque composante : le terme en $\mathbf{e}$, sans construire $\mathbf{e}$.

Théorème 4.8 — Retour au théorème de Banach

Le théorème 4.6 ne dit pas combien d'itérations faire, et n'a été démontré que dans le cas diagonalisable. Banach donne une preuve complète, avec une estimation de l'erreur.

Théorème 4.8 — La matrice Google est une contraction

Soit Δ l'ensemble des vecteurs de probabilités de $\mathbb{R}^N$, muni de d_1. Pour tous $\mathbf{x}, \mathbf{y} \in \Delta$,

$$d_1(G\mathbf{x}, G\mathbf{y}) \leq \beta\, d_1(\mathbf{x}, \mathbf{y}).$$

DÉMONSTRATION AU TABLEAU

$\mathbf{z} = \mathbf{x} - \mathbf{y}$ vérifie $\mathbf{e}^{\mathsf{T}}\mathbf{z} = 0$, donc $G\mathbf{z} = \beta B\mathbf{z}$. Puis $\sum_i |\sum_j b_{ij} z_j| \leq \sum_j |z_j| \sum_i b_{ij} = \sum_j |z_j|$ (coefficients positifs, colonnes de somme 1).

Le PageRank, point fixe d'une contraction

- (Δ, d_1) est complet : une suite de Cauchy de Δ a des composantes de Cauchy, donc convergentes, et la limite a encore des composantes positives de somme 1.
- Banach (forme générale) : un unique point fixe $\mathbf{q}$, et la méthode converge quel que soit le départ — sans Perron–Frobenius.
- Majoration de l'erreur : $d_1(\mathbf{w}_k, \mathbf{q}) \leq \frac{\beta^k}{1-\beta}\, d_1(\mathbf{w}_0, \mathbf{w}_1)$; en partant de $\mathbf{w}_{k-1}$: $d_1(\mathbf{w}_k, \mathbf{q}) \leq \frac{\beta}{1-\beta}\, d_1(\mathbf{w}_{k-1}, \mathbf{w}_k)$. Au critère d'arrêt, l'erreur est inférieure à $\frac{0{,}85}{0{,}15}\,\varepsilon \approx 5{,}7\,\varepsilon$.
- Le nombre d'itérations ne dépend pas de la taille du Web. Comme $d_1 \leq 2$ sur Δ : $d_1(\mathbf{w}_k, \mathbf{q}) \leq 2\beta^k$; pour 10^{-8} avec $\beta = 0{,}85$, il suffit de $k \geq 118$.
- Majoration pessimiste : le programme s'arrête après 93 itérations ; Page rapportait qu'une cinquantaine suffisaient sur des centaines de millions de liens.

Remarque 4.3 — Le rôle de la seconde valeur propre

Remarque 4.3

- La vitesse est gouvernée par $|\lambda_2|$, le plus grand module des valeurs propres de G autres que 1 ; on démontre que $|\lambda_2| \leq \beta$ (cohérent avec le théorème 4.8).
- Mini-web 3 : égalité, G a la valeur propre $-0{,}85$, héritée du -1 de A (l'aller-retour $6 \leftrightarrow 7$).
- β joue un double rôle : compromis fidélité / téléportation (tableau 4.3), et vitesse de convergence. Avec $\beta = 0{,}99$: environ $1\,900$ itérations pour la même précision.

PageRank n'est qu'un critère parmi d'autres, et sa mise en œuvre chez Google n'a jamais été entièrement publiée. Mais l'idée sert ailleurs : citations scientifiques, équipes sportives, réseaux alimentaires — chaque fois, un vecteur propre de valeur propre 1.

Pour la classe 4.4 — Classer, une activité pour la classe

Pour la classe — Classer, une activité pour la classe

- Un petit réseau de quatre ou cinq pages au tableau ; les élèves prédisent un classement (la plupart comptent les liens entrants).
- On simule l'internaute aléatoire avec un dé : une trentaine de clics par équipe ; on compare les fréquences à la prédiction.
- Un réseau bien choisi, comme celui de la figure 4.1, où une page qui reçoit peu de liens devance une page qui en reçoit autant, lance une discussion : que veut dire « être important » dans un réseau ?
- En Technico-sciences : calculer $A\mathbf{p}$ au tableur (multiplication matricielle).

Note historique 4.4 — Larry Page (né en 1973)



En 2014. Photo : Steve Jurvetson, CC BY 2.0, Wikimedia Commons.

- Né à East Lansing (Michigan) ; ses deux parents enseignent l'informatique. Doctorat à Stanford en 1995 (Terry Winograd) ; il y rencontre Sergey Brin.
- Analogie avec les citations : un article souvent cité, surtout par des articles importants, est important. Le robot du projet BackRub télécharge les liens du Web dès 1996.
- « The PageRank Citation Ranking » (1999) ; jeu de mots entre son nom et la page web. Brevet de Stanford, licence exclusive à Google.
- Septembre 1998 : Google, dans un garage de Menlo Park ; il abandonne son doctorat. PDG jusqu'en 2001, puis de 2011 à 2015 ; Alphabet jusqu'en 2019.

Synthèse

Ce qu'il faut retenir

- Probabilités totales : « multiplier le long d'une branche, additionner les branches » — une ligne d'un produit matriciel.
- Chaîne de Markov : l'avenir ne dépend que du présent ; la loi évolue par $\mathbf{x}_{n+1} = M\mathbf{x}_n$, avec M stochastique (colonnes de somme 1).
- L'équilibre $\mathbf{q} = M\mathbf{q}$: un vecteur propre de valeur propre 1, et un point fixe ; limite des probabilités et proportion du temps passé dans chaque état.
- Si M est régulière : un unique équilibre, atteint depuis n'importe quel départ ; la vitesse dépend de $|\lambda_2|$.
- PageRank : corriger les pages sans lien (B), puis téléporter ($G = \beta B + (1-\beta)U$).
- Méthode des puissances : G est une contraction de rapport β pour d_1 ; Banach garantit la convergence et borne l'erreur.

Un même point fixe, trois chapitres

Chapitre	Espace complet	Point fixe
2	$I \subset \mathbb{R}$ fermé	le zéro cherché par Newton
3	$(\mathcal{K}(\mathbb{R}^2), d_H)$	une fractale
4	(Δ, d_1)	le PageRank du Web

À l'origine de l'une des plus grandes entreprises du monde : un projet de doctorat dont le cœur est le calcul d'un vecteur propre associé à la valeur propre 1.

Et maintenant

- Les démonstrations laissées « au tableau » aujourd'hui sont toutes dans les notes de cours, avec tous les détails.
- Les exercices du chapitre (avec solutions complètes) couvrent chacun des résultats vus : à travailler avant le prochain cours.
- Les programmes Python des notes (météo, simulation, mini-webs, méthode des puissances) sont à exécuter et à modifier.
- Les rappels d'algèbre linéaire (valeurs propres, diagonalisation) sont à l'annexe D.