Théorie des graphes (5)
Théorie algébrique des graphes

Michel Rigo

http://www.discmath.ulg.ac.be/

Année 2015–2016

Université de Liège

Comportement asymptotique
du nombre de chemins de longueur n

Rappel du cas primitif

Corollaire du thm. de Perron

Si A est une matrice primitive,

$$\boxed{A^k = \lambda_A^k \, v_A \, \widetilde{w_A} + o(\lambda_A^k)}$$

où v_A et $\widetilde{w_A}$ sont des vecteurs propres choisis t.q. $\widetilde{w_A}.v_A = 1$.

$$A = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}, \ \lambda \simeq 1.839$$

$$v_A = \begin{pmatrix} 0.543689 \\ 0.160713 \\ 0.295598 \end{pmatrix}, \quad \widetilde{w_A} = \begin{pmatrix} 0.419643 & 0.228155 & 0.352201 \end{pmatrix}$$

$$\widetilde{w_A}.v_A \simeq 0.368933 \neq 1$$

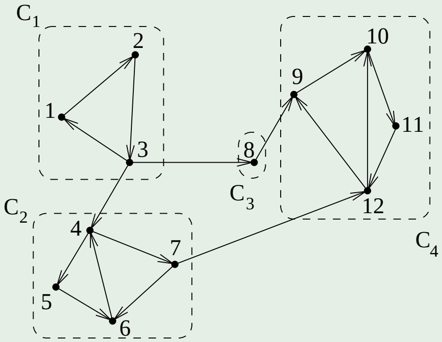
$\lambda \simeq 1.839$

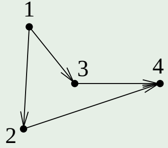
$$\lim_{k\to+\infty} \frac{A^k}{\lambda^k} = \frac{v_A\,\widetilde{w_A}}{\widetilde{w_A}\,.\,v_A} = \begin{pmatrix} 0.61842 & 0.336228 & 0.519032 \\ 0.182804 & 0.0993883 & 0.153425 \\ 0.336228 & 0.182804 & 0.282192 \end{pmatrix}.$$

On sait donc comment se comporte le nombre de chemins de longueur k entre deux sommets quelconques quand $k \to +\infty$.

Graphe ayant plusieurs composantes fortement connexes

Exemple





On considère le condensé $\mathcal{C}$ d'un graphe G (ou graphe acyclique des composantes).

On peut ordonner les sommets de $\mathcal{C}$ par tri topologique.

On obtient une matrice bloc triangulaire supérieure :

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

Remarque

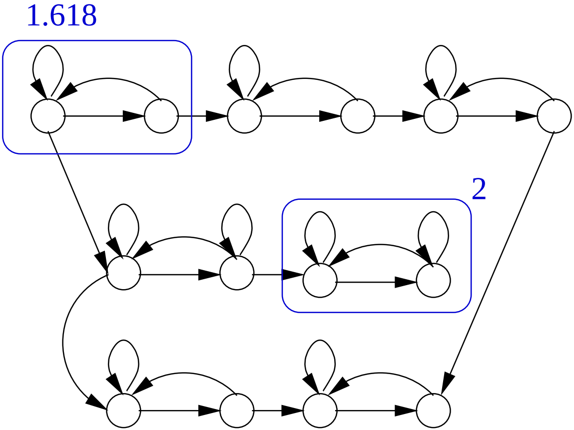
Le spectre d'un graphe est l'union des spectres de ses composantes connexes.

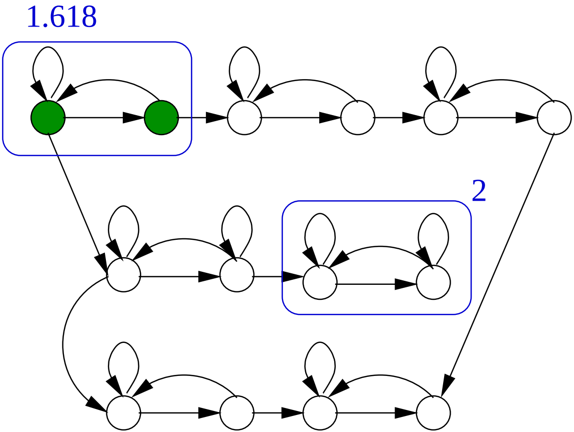
Si toutes les composantes f. connexes sont **primitives**

On peut montrer que

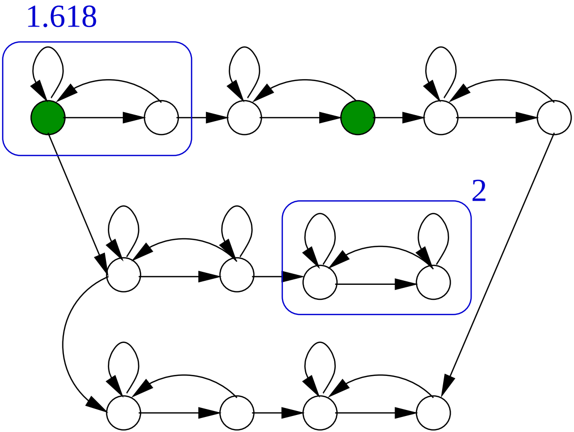
Pour estimer le nombre de chemins de longueur n entre deux sommets, il suffit de :

- Détecter la plus grande valeur de Perron λ des différentes composantes connexes par lesquelles passent les chemins d'intérêt
- Compter le nombre k de composantes ayant cette valeur propre comme valeur dominante.
- Le nombre de chemins de longueur n se comporte alors asymptotiquement comme $n^{k-1}\lambda^n$.

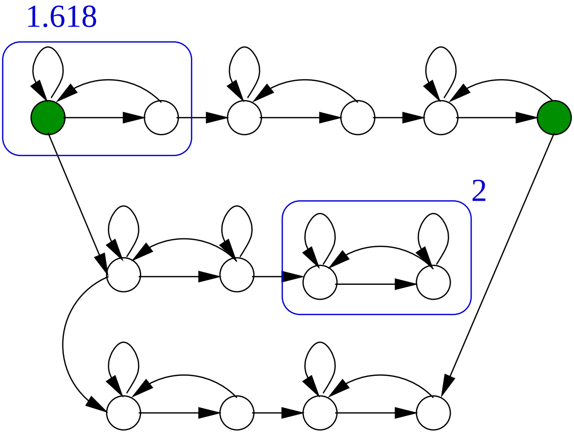

1.618
2



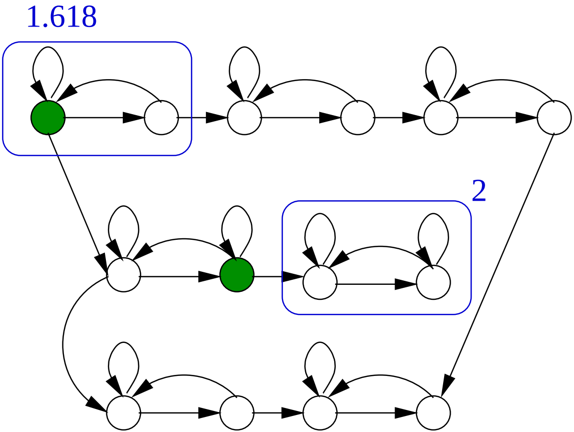
$\#$chemins de lg. $n \simeq \varphi^n$



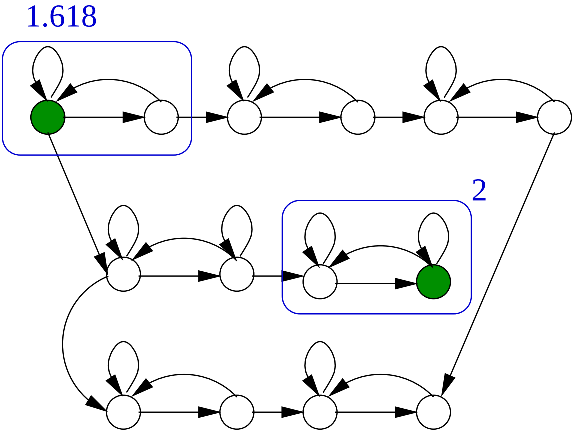
$\#$chemins de lg. $n \simeq n\,\varphi^n$



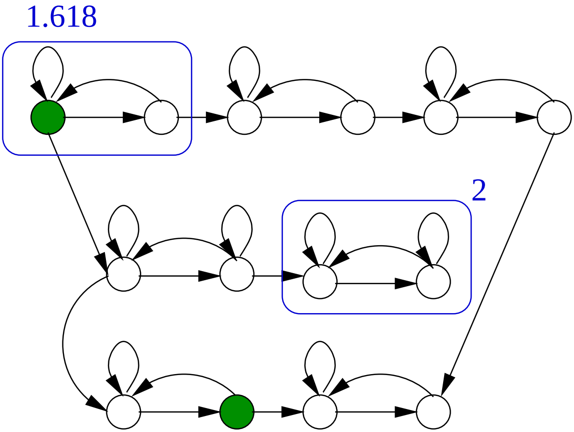
#chemins de lg. $n \simeq n^2\,\varphi^n$



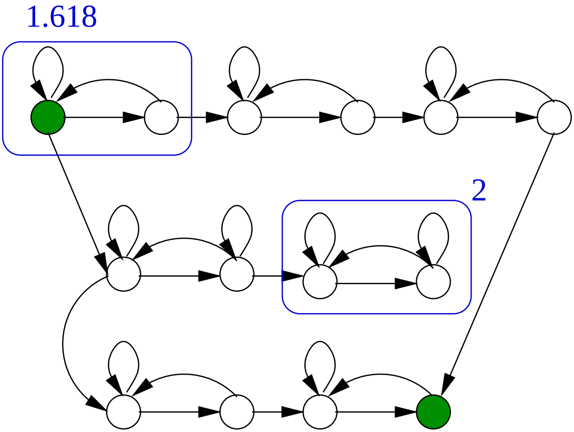
#chemins de lg. $n \simeq \color{red}{2^n}$



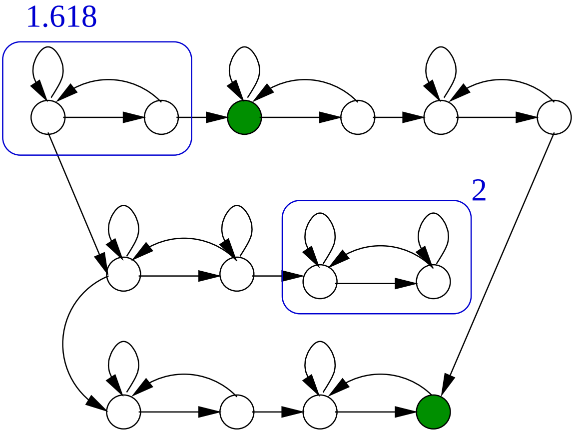
#chemins de lg. $n \simeq n\,2^n$



#chemins de lg. $n \simeq 2^n$



#chemins de lg. $n \simeq 2^n$



#chemins de lg. $n \simeq n^2\,\varphi^n$

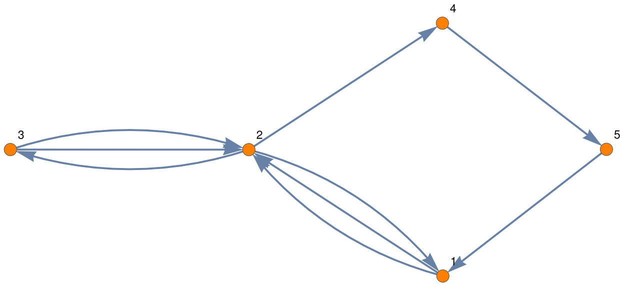
Si on connaît la forme normal de Jordan

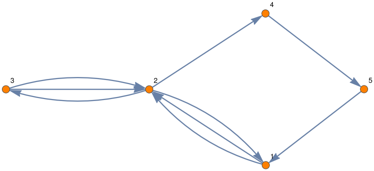
Soient $\lambda_1, \ldots, \lambda_p$ les valeurs propres distinctes d'une $M \in \mathbb{C}_d^d$ zéros du polynôme minimum de M de multiplicité $m_1, \ldots, m_p$, alors $\forall k$,

$$(M^k)_{i,j} = \sum_{t=1}^{p} P_{i,j}^{(t)} \, \lambda_t^k$$

où $P_{i,j}^{(t)}$ est un polynôme de degré $< m_t$.

Structure des matrices irréductibles





$$\begin{pmatrix} 0 & 2 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 \\ 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 2 & 0 & 2 & 2 & 0 \\ 0 & 4 & 0 & 0 & 1 \\ 2 & 0 & 2 & 2 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 8 & 0 & 0 & 2 \\ 5 & 0 & 4 & 4 & 0 \\ 0 & 8 & 0 & 0 & 2 \\ 0 & 2 & 0 & 0 & 0 \\ 2 & 0 & 2 & 2 & 0 \end{pmatrix},$$

$$\begin{pmatrix} 10 & 0 & 8 & 8 & 0 \\ 0 & 18 & 0 & 0 & 4 \\ 10 & 0 & 8 & 8 & 0 \\ 2 & 0 & 2 & 2 & 0 \\ 0 & 8 & 0 & 0 & 2 \end{pmatrix}, \begin{pmatrix} 0 & 36 & 0 & 0 & 8 \\ 22 & 0 & 18 & 18 & 0 \\ 0 & 36 & 0 & 0 & 8 \\ 0 & 8 & 0 & 0 & 2 \\ 10 & 0 & 8 & 8 & 0 \end{pmatrix}$$

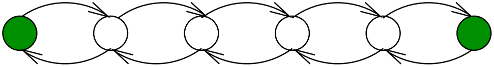
On va introduire une notion de *période* et le but est d'obtenir

Théorème fondamental (cas irréductible)

Soit $A \geq 0$ une matrice carrée irréductible de période $p \geq 1$.
Pour tout couple (i, j) d'indices,
il existe un unique entier $r_{i,j} \in \{0, \dots, p-1\}$ tel que

- $[A^n]_{i,j} > 0$ entraîne $n \equiv r_{i,j} \pmod p$ et
- il existe $N_{i,j}$ tel que

$$[A^{np+r_{i,j}}]_{i,j} > 0 \quad \text{pour tout } n \geq N_{i,j}.$$



on a des chemins de longueur $2n+1$ pour $n \geq 2$.

Définition (en termes de matrice ou de graphe)

S'il existe $N > 0$ tel que $[A^N]_{i,i} > 0$, la période de l'indice i est le p.g.c.d. de l'ensemble des entiers $n > 0$ pour lesquels

$$[A^n]_{i,i} > 0.$$

On la note $p(i)$.

S'il existe un circuit de longueur N passant par i, la période du sommet i est le p.g.c.d. de l'ensemble des longueurs des circuits passant par i.

Le p.g.c.d. d'un ensemble (infini) $X = \{x_1 < x_2 < \cdots\} \subseteq \mathbb{N}$ est le plus grand entier p appartenant à l'ensemble fini $\{1, 2, \ldots, x_1\}$ tel que pour tout $k \geq 1$, p divise x_k.

Lemme 1

Soient i, j deux indices de $A \geq 0$. S'il existe m, n tels que $[A^m]_{i,j} > 0$ et $[A^n]_{j,i} > 0$, alors $p(i) = p(j)$.

$\rightsquigarrow$ Deux sommets quelconques d'une même composante f. connexe ont même la période.

Pour tout s tel que $[A^s]_{j,j} > 0$, on a

$$\begin{aligned}
[A^{m+s+n}]_{i,i} &= \sum_{k=1}^{d} [A^{m+s}]_{i,k}[A^n]_{k,i} \\
&\geq [A^{m+s}]_{i,j}[A^n]_{j,i} = \sum_{k=1}^{d} [A^m]_{i,k}[A^s]_{k,j}[A^n]_{j,i} \\
&\geq [A^m]_{i,j}[A^s]_{j,j}[A^n]_{j,i} > 0
\end{aligned}$$

$\Rightarrow$ $p(i)$ divise $m + n + s$.

Pour un tel s, on a $[A^{2s}]_{j,j} > 0$
(en effet, $[A^s.A^s]_{j,j} \geq [A^s]_{j,j}.[A^s]_{j,j}$).
Dès lors, on a aussi

$$[A^{m+2s+n}]_{i,i} > 0.$$

$\Rightarrow$ $p(i)$ divise $m + 2s + n$

$p(i)$ divise s $(= m + 2s + n - (m + s + n))$.

Pour tout s tel que $[A^s]_{j,j} > 0$, $p(i)$ divise s donc $p(i) \leq p(j)$.

Par symétrie, on a aussi que $p(j) \leq p(i)$ et $p(i) = p(j)$.

Conclusion : on peut définir *la période* d'une composante f. connexe ou d'un graphe f. connexe (ou d'une matrice irréductible).

Définition

Une matrice irréductible $A \in \mathbb{R}_n^n$ est cyclique de période p, si tous les indices de A sont de période $p > 1$.

Si la période $p = 1$, alors A est dite acyclique.

Lemme 2

Soit $A \geq 0$ une matrice carrée irréductible de période $p \geq 1$.
Soit i un indice,

$$\exists N_i \geq 0 : \ \forall n \geq N_i, \ [A^{np}]_{i,i} > 0.$$

Supposons tout d'abord que $[A^{kp}]_{i,i} > 0$ et $[A^{\ell p}]_{i,i} > 0$. Dès lors

$$[A^{(k+\ell)p}]_{i,i} \geq [A^{kp}]_{i,i} \, [A^{\ell p}]_{i,i} > 0.$$

L'ensemble $\mathcal{S}$ des multiples np de p qui sont tels que $[A^{np}]_{i,i} > 0$ est

- stable pour l'addition
- contient au moins un multiple de p
- le p.g.c.d. des éléments de $\mathcal{S}$ vaut p.

La conclusion découle alors du lemme 'arithmétique' suivant.

Lemme

Soit $X \subseteq \mathbb{N}$ un ensemble d'entiers stable pour l'addition.
Alors X contient tous les multiples du p.g.c.d. des éléments de X à l'exception d'un nombre fini d'entre eux.

Exemple (pgcd=1)

$2, 7 \in X$ et X stable par addition

$$2, 4, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, \ldots$$

Exemple (pgcd=2)

$4, 14 \in X$ et X stable par addition

$$4, 8, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, \ldots$$

$$X = \{x_1 < x_2 < x_3 < \cdots\}$$

Soit p le p.g.c.d. des éléments de X.
Quitte à diviser par p, on peut supposer que $p = 1$.

Il existe un ensemble fini $\{x_1, \ldots, x_k\} \subseteq X$ tel que

$$\text{p.g.c.d.}\,\{x_1, \ldots, x_k\} = 1$$

Nous savons que le p.g.c.d. de X vaut 1.
$\{x_1\}$, le p.g.c.d. potentiel x_1.
$\{x_1, x_2\}$, le p.g.c.d. potentiel $\leq x_1$
$\{x_1, x_2, x_3\}$... à chaque étape, le p.g.c.d. décroît.

Il existe k tel que le p.g.c.d. de $\{x_1, \ldots, x_k\}$ soit 1.
Sinon, le p.g.c.d. de X serait > 1 !

k peut être > 2, exemple : $\{6, 10, 15\}$ dont le p.g.c.d. vaut 1 mais dont les éléments 2 à 2 ne sont pas premiers entre eux.

$$\text{p.g.c.d.}\,\{x_1,\ldots,x_k\}=1$$

thm. de Bezout : $\exists\lambda_1,\ldots,\lambda_k\in\mathbb{Z}$ t.q. $\lambda_1\,x_1+\cdots+\lambda_k\,x_k=1$.
Si on regroupe tous les termes dont les coefficients λ_i sont positifs (resp. négatifs), cette somme se réécrit

$$m-n=1$$

avec $m,n\in X$ car X est stable pour l'addition.
Soit q un entier tel que $q\geq n(n-1)$.
Par division euclidienne (par n),

$$q=a\,n+b,\quad 0\leq b<n.$$

De plus, $a\geq n-1$. Puisque $m-n=1$, il vient

$$q=a\,n+b\,(m-n)=(a-b)\,n+b\,m$$

avec $a-b\geq 0$. On en conclut q appartient à X (car $m,n\in X$).
Donc tout $q\geq n(n-1)$ appartient à X.

Théorème fondamental (cas irréductible)

Soit $A \geq 0$ une matrice carrée irréductible de période $p \geq 1$.
Pour tout couple (i, j) d'indices,
il existe un unique entier $r_{i,j} \in \{0, \ldots, p-1\}$ tel que

(I) $[A^n]_{i,j} > 0$ entraîne $n \equiv r_{i,j} \pmod{p}$ et

(II) il existe $N_{i,j}$ tel que

$$[A^{np+r_{i,j}}]_{i,j} > 0 \quad \text{pour tout } n \geq N_{i,j}.$$

(i) Supposons $[A^m]_{i,j} > 0$ et $[A^n]_{i,j} > 0$.
Thèse : $m \equiv n \pmod{p}$.

Puisque A est irréductible, il existe ℓ tel que $[A^\ell]_{j,i} > 0$. Dès lors,

$$[A^{m+\ell}]_{i,i} \geq [A^m]_{i,j}[A^\ell]_{j,i} > 0 \quad \text{et} \quad [A^{n+\ell}]_{i,i} > 0.$$

La période p divise donc $m + \ell$ et $n + \ell$ donc leur différence.
Autrement dit, $m - n \equiv 0 \pmod{p}$.

(ii) Puisque A est irréductible, il existe ℓ tel que $[A^\ell]_{i,j} > 0$ et au vu de la première partie,

$$\ell = mp + r_{i,j}.$$

Posons $N_{i,j} = N_i + m$ (avec N_i donné dans le Lemme 2).

Par définition de N_i, on a

$$\forall n \geq N_i,\ [A^{np}]_{i,i} > 0.$$

De là, si $k \geq N_{i,j}$, alors

$$kp + r_{i,j} = (n+m)p + r_{i,j} \quad \text{avec } n \geq N_i.$$

et

$$[A^{kp+r_{i,j}}]_{i,j} \geq [A^{np}]_{i,i}[A^{mp+r_{i,j}}]_{i,j} > 0.$$

Proposition

Une matrice irréductible est acyclique SSI elle est primitive.

$\Rightarrow$ Si la matrice est acyclique (i.e., de période $p = 1$), avec les notations du thm. de structure, $r_{i,j} = 0 \ \forall i, j$. Donc

$$[A^n]_{i,j} > 0 \quad \text{si } n \geq N_{i,j}.$$

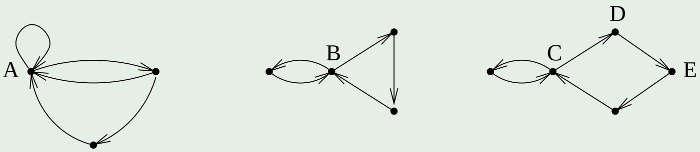
On pose $\mathcal{N} = \sup_{i,j} N_{i,j}$ et $A^{\mathcal{N}} > 0$, i.e. A est primitive.

$\Leftarrow$ Si A est primitive, (en particulier, A est irréductible) pour k suffisamment grand et pour tout indice i de A,

$$[A^k]_{i,i} > 0 \text{ et } [A^{k+1}]_{i,i} > 0.$$

Le p.g.c.d. de k et de $k + 1$ étant 1, la conclusion en découle.

Applications



Lien (sans preuve) entre le théorème de Perron–Frobenius

Rappel (théorème de Perron–Frobenius)

Soit $A \geq 0$ matrice irréductible

- ...
- Il existe $d \geq 1$ tel que si μ est une valeur propre de A telle que $|\mu| = \lambda_A$, alors $\mu = \lambda_A \, e^{2ik\pi/d}$ et pour tout $k \in \{0, \ldots, d-1\}$, $\lambda_A \, e^{2ik\pi/d}$ est une valeur propre de A.

et la notion de période $p \geq 1$ d'une matrice irréductible

$$\boxed{d = p}$$

Corollaire

Si $A \geq 0$ est une matrice irréductible possédant une valeur propre dominante λ (i.e., $d = 1$: pour toute valeur propre $\mu \neq \lambda$ de A, $|\mu| < \lambda$), alors A est primitive.

Algorithme du PageRank

Encore un peu de théorie...

Proposition

Soit $G = (V, E)$ un multi-graphe non orienté k-régulier. Alors

- k est une valeur propre de G,
- pour toute valeur propre λ de G, on a $|\lambda| \leq k$,
- si G est connexe, k est valeur propre simple
  (i.e., les multiplicités géométrique et algébrique valent 1).

Remarque

Proposition OK dans le cas orienté.
Remplacer “connexe” par f. connexe.

1. $(1, \ldots, 1)^{\sim}$ est un vecteur propre de $A(G)$ de valeur propre k.

2. considérons une valeur propre λ de $A(G)$ ayant $y \neq 0$ comme vecteur propre.

Soit y_j une composante de y de module maximum

$$|\lambda|\, |y_j| = |[A(G)y]_j| \leq \sum_{i=1}^{n} [A(G)]_{j,i} |y_i| \leq |y_j| \sum_{i=1}^{n} [A(G)]_{j,i} = k\, |y_j|$$

donc $|\lambda| \leq k$.

3. G est connexe, $A(G)$ est irréductible.
Par le thm. de Perron–Frobenius, la matrice $A(G)$ possède une unique valeur propre réelle dominante et vu 2, il s'agit de k.

Une matrice $M \geq 0$ est stochastique,
si la somme des éléments de chaque ligne vaut 1.

Corollaire

Si $M \in \mathbb{Q}_r^r$ est une matrice stochastique,
alors 1 est valeur propre dominante de M.

Soit $M \in \mathbb{Q}_r^r$. En multipliant tous les éléments de M par le p.p.c.m. γ des dénominateurs des éléments de M, la matrice

$$M' = \gamma\, M$$

est telle que la somme des éléments de chaque ligne vaut $\gamma \in \mathbb{N}$.

Il s'agit donc de la matrice d'adjacence d'un digraphe γ-régulier.
Vu la prop. précédente, $\gamma\, M$ possède γ comme valeur propre dominante (i.e., toute autre valeur propre μ est telle que $|\mu| \leq \gamma$).

La conclusion suit en divisant par γ.

Google...

Algorithme de PageRank : S. Brin, L. Page

Google attribue à chaque page une mesure, appelée "PageRank", destinée à déterminer si elles font ou non autorité.

Lorsqu'on effectue une recherche sur un mot clé donné, Google extrait les pages contenant ce mot clé et les classe en se basant sur ce PageRank.

Larry Page, Sergey Brin

On voudrait implémenter deux règles simples :

- on accorde plus d'importance, i.e., un score de "PageRank" plus élevé, aux *pages référencées par des pages qui font elles-mêmes autorité* , càd dont le PageRank est élevé ;
- on accorde d'autant moins de crédit à un lien, si il provient d'une *page qui dispose de nombreux liens*.

Le PageRank $\pi_j \geq 0$ de la page $j \in \{1, \ldots, n\}$ serait donné par

$$\pi_j = \sum_{i \in \text{pred}(j)} \frac{\pi_i}{d^+(i)} \tag{1}$$

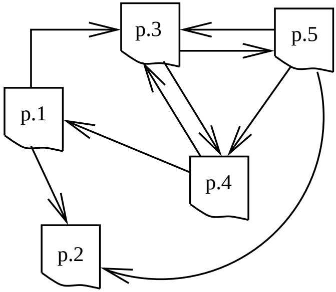
formule récursive, on ne dispose pas *a priori* de méthode assurant

- l'existence,
- l'unicité,
- le calcul efficace

d'une solution $\pi = (\pi_1, \ldots, \pi_n)$ non triviale.

On peut supposer que les scores recherchés sont *normalisés*,

$$\sum_{i=1}^{n} \pi_i = 1.$$



$$\begin{pmatrix}\pi_1 & \pi_2 & \pi_3 & \pi_4 & \pi_5\end{pmatrix}\begin{pmatrix}0 & 1/2 & 1/2 & 0 & 0\\ 0 & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 1/2 & 1/2\\ 1/2 & 0 & 1/2 & 0 & 0\\ 0 & 1/3 & 1/3 & 1/3 & 0\end{pmatrix} = \begin{pmatrix}\pi_1 & \pi_2 & \pi_3 & \pi_4 & \pi_5\end{pmatrix}$$

Remarque

La matrice n'est ni primitive, ni irréductible.
On cherche un vecteur propre de valeur propre 1.

$$\begin{pmatrix}\pi_1 & \pi_2 & \pi_3 & \pi_4 & \pi_5\end{pmatrix}\begin{pmatrix} 0 & 1/2 & 1/2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1/2 & 1/2 \\ 1/2 & 0 & 1/2 & 0 & 0 \\ 0 & 1/3 & 1/3 & 1/3 & 0 \end{pmatrix} = \begin{pmatrix}\pi_1 & \pi_2 & \pi_3 & \pi_4 & \pi_5\end{pmatrix}$$

$$\begin{cases} \pi_1 = \pi_4/2 \\ \pi_2 = \pi_1/2 + \pi_5/3 \\ \pi_3 = \pi_1/2 + \pi_4/2 + \pi_5/3 \\ \pi_4 = \pi_3/2 + \pi_5/3 \\ \pi_5 = \pi_3/2 \end{cases}$$

$$\sum_{i=1}^{n} \pi_i = 1.$$

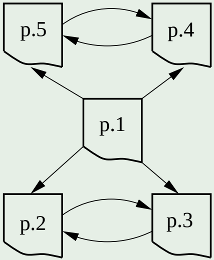
Continuons l'exemple...

$$\begin{cases} \pi_1 = \pi_4/2 \\ \pi_2 = \pi_1/2 + \pi_5/3 \\ \pi_3 = \pi_1/2 + \pi_4/2 + \pi_5/3 \\ \pi_4 = \pi_3/2 + \pi_5/3 \\ \pi_5 = \pi_3/2 \end{cases} \Leftrightarrow \begin{cases} \pi_1 = \pi_4/2 \\ \pi_5 = \pi_3/2 \\ \pi_2 = \pi_4/4 + \pi_3/6 \\ \pi_3 = \pi_4/4 + \pi_4/2 + \pi_3/6 \\ \pi_4 = \pi_3/2 + \pi_3/6 \end{cases}$$

$$\begin{cases} \pi_4 = 10\pi_3/9 \\ \pi_4 = 6\pi_3/9 \end{cases}$$

La seule solution est $\pi_1 = \pi_2 = \pi_3 = \pi_4 = \pi_5 = 0$.

Un second exemple...



$$\begin{cases} \pi_2 = \pi_1/4 + \pi_3 \\ \pi_3 = \pi_1/4 + \pi_2 \\ \pi_4 = \pi_1/4 + \pi_5 \\ \pi_5 = \pi_1/4 + \pi_4. \end{cases}$$

on trouve par exemple,

- $\pi_1 = \pi_2 = \pi_3 = 0$, $\pi_4 = \pi_5 = 1/2$ ou bien,
- $\pi_1 = \pi_4 = \pi_5 = 0$, $\pi_2 = \pi_3 = 1/2$.

Rappel : modèle proposé

$$\pi_j = \sum_{i \in \text{pred}(j)} \frac{\pi_i}{d^+(i)}$$

Réécriture matricielle ("H" comme "hyperlien"),

$$\pi = \pi H \tag{2}$$

où

$$H_{ij} = \begin{cases} A(G)_{ij} / d^+(i) & \text{si } d^+(i) > 0 \\ 0 & \text{si } d^+(i) = 0 \end{cases}$$

avec $A(G)$ la matrice d'adjacence du graphe G

Remarque

La matrice H est stochastique (sauf pour les lignes de 0).

Au vu des deux exemples

On ne peut *a priori*

- ni garantir l'existence d'une solution $\neq 0$ **:-(**
- ni garantir l'unicité de la solution **:-(**

Solution

$\rightarrow$ Perturber légèrement le modèle initial
pour obtenir un système "proche" mais avec de "belles" propriétés
$\rightsquigarrow$ pouvoir appliquer le thm. de Perron.

On pertube le modèle

1. Pour se débarrasser des “puits”, i.e., des pages ne pointant vers aucune autre page et pour obtenir une matrice stochastique, on introduit une matrice S (“S” comme “stochastique”) définie par

$$S_{ij} = \begin{cases} A(G)_{ij}/d^+(i) & \text{si } d^+(i) > 0 \\ 1/n & \text{si } d^+(i) = 0. \end{cases}$$

On pertube le modèle

2. Pour assurer la forte connexité du graphe, on construit une matrice G ("G" comme Google) donnée par la combinaison affine (et même convexe) suivante avec un réel $\alpha \in [0,1]$ fixé

$$G = \alpha\, S + (1-\alpha)\, J/n$$

où $J = (1)_{1 \leq i,j \leq n}$. L'équation initiale (2) est remplacée par

$$\boxed{\pi = \pi\, G.}$$

(La matrice J/n est parfois appelée *matrice de téléportation*)

$$\overbrace{\begin{pmatrix} 0 & 1/2 & 1/2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1/2 & 1/2 \\ 1/2 & 0 & 1/2 & 0 & 0 \\ 0 & 1/3 & 1/3 & 1/3 & 0 \end{pmatrix}}^{H}, \quad \overbrace{\begin{pmatrix} 0 & 1/2 & 1/2 & 0 & 0 \\ 1/5 & 1/5 & 1/5 & 1/5 & 1/5 \\ 0 & 0 & 0 & 1/2 & 1/2 \\ 1/2 & 0 & 1/2 & 0 & 0 \\ 0 & 1/3 & 1/3 & 1/3 & 0 \end{pmatrix}}^{S}.$$

$$0,85.\begin{pmatrix} 0 & 1/2 & 1/2 & 0 & 0 \\ 1/5 & 1/5 & 1/5 & 1/5 & 1/5 \\ 0 & 0 & 0 & 1/2 & 1/2 \\ 1/2 & 0 & 1/2 & 0 & 0 \\ 0 & 1/3 & 1/3 & 1/3 & 0 \end{pmatrix} + 0,15.\begin{pmatrix} 1/5 & 1/5 & 1/5 & 1/5 & 1/5 \\ 1/5 & 1/5 & 1/5 & 1/5 & 1/5 \\ 1/5 & 1/5 & 1/5 & 1/5 & 1/5 \\ 1/5 & 1/5 & 1/5 & 1/5 & 1/5 \\ 1/5 & 1/5 & 1/5 & 1/5 & 1/5 \end{pmatrix}$$

$$= \begin{pmatrix} 3/100 & 91/200 & 91/200 & 3/100 & 3/100 \\ 1/5 & 1/5 & 1/5 & 1/5 & 1/5 \\ 3/100 & 3/100 & 3/100 & 91/200 & 91/200 \\ 91/200 & 3/100 & 91/200 & 3/100 & 3/100 \\ 3/100 & 47/150 & 47/150 & 47/150 & 3/100 \end{pmatrix} = \mathbf{G}.$$

Choix heuristique de α

Remarque

Google attribue à α une valeur de $0,85$.
Ce choix n'est pas arbitraire.

Au plus α est proche de 1 :

- au mieux on approche le modèle "naturel" (2) proposé initialement
- on diminue le rôle artificiel de la matrice de téléportation.

Cependant, on peut montrer que ce paramètre α contrôle la vitesse de convergence de la méthode de calcul développée et donc le nombre d'itérations à effectuer pour obtenir une estimation du vecteur π

Choix heuristique de α

Quand α tend vers 1, le nombre d'itérations devient prohibitif (cf. C. Meyer et A. Langville).

α	nombre d'itérations
$0,5$	34
$0,75$	81
$0,8$	104
$0,85$	142
$0,9$	219
$0,95$	449
$0,99$	2292
$0,999$	23015

Choix heuristique de α

Compromis

$\alpha = 0,85$ semble un bon compromis entre le caractère artificiel introduit par la matrice de téléportation et la masse de calculs à réaliser.

Par une discussion plus fine sur les valeurs propres : au plus α est proche de 1, au plus π est sensible aux petites pertubations de la matrice H (gênant vu la grande volatilité du web et de sa structure)

Le modèle probabiliste du surfeur

un surfeur se trouvant sur une page quelconque a deux choix possibles :

- avec une probabilité α, il clique avec une probabilité uniforme sur l'un des liens de la page pour changer de page.
- Soit, avec une probabilité $1-\alpha$, il se déplace avec une probabilité uniforme sur l'une des n pages de l'Internet tout entier.

G_{ij} représente la probabilité de transition lorsque le surfeur se trouve sur la page i de passer à la page j.

$\rightsquigarrow$ G_{ij}^k représente la probabilité de transition lorsque le surfeur se trouve sur la page i de passer à la page j en k clics (chemins de longueur k).

Modèles initial et perturbé

Par rapport à l'équation initiale (1), l'emploi de la matrice G donne la formule suivante pour la détermination des "nouveaux" π_j qui seront effectivement calculés

$$\pi_j = \sum_{i=1}^{n} \pi_i \; (\alpha\, S_{ij} + (1-\alpha)\, \frac{1}{n})$$

$$= \alpha \sum_{i \in \text{pred}(j)} \frac{\pi_i}{d^+(i)} + \frac{1}{n} \left(1 - \alpha + \alpha \sum_{i : d^+(i)=0} \pi_i \right) .$$

Nous nous sommes donc éloignés quelque peu du modèle initialement proposé mais ces modifications vont permettre un calcul efficace (et assurant l'existence et l'unicité d'une solution) !

Les matrices S, J/n et G sont stochastiques,
$\rightsquigarrow 1$ est valeur propre dominante de G (corollaire).

Par construction, la matrice G est primitive car $G > 0$.

On peut appliquer le théorème de Perron, la valeur propre dominante 1 est simple et il existe un unique vecteur colonne $x > 0$ (resp. ligne $y > 0$) tel que

$$\sum_{i=1}^{n} x_i = 1 \text{ (resp. } \sum_{i=1}^{n} y_i = 1) \quad \text{et} \quad Gx = x \text{ (resp. } yG = y).$$

Remarque

valeur propre simple $\rightsquigarrow$ unicité de la solution “normalisée”

Conclusion

Déterminer le vecteur des "PageRanks" π revient à chercher le vecteur propre y de Perron à gauche (normalisé) de G.

En appliquant le résultat asymptotique (A primitive)

$$\boxed{A^k = \lambda_A^k\, v_A\, \widetilde{w_A} + o(\lambda_A^k), \quad \widetilde{w_A} v_A = 1}$$

e joue le rôle de v_A, π celui de w_A :

- $e = (1 \cdots 1)\tilde{}$ est un vecteur propre à droite de G de valeur propre 1 (G est stochastique)
- π est un vecteur propre à gauche de G de valeur propre 1
- $\pi e = 1$ (scores sont normalisés)

$$G^k = e\pi + o(1) \text{ i.e., } \lim_{k\to\infty} G^k = e\pi = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \begin{pmatrix} \pi_1 & \cdots & \pi_n \end{pmatrix}.$$

Méthode itérative pour estimer π

Soit

$$p^{(0)} = \begin{pmatrix} p_1^{(0)} & \cdots & p_n^{(0)} \end{pmatrix} > 0 \text{ vecteur t.q. } \sum_i p_i^{(0)} = 1.$$

$\forall k \geq 1$, on pose $p^{(k)} = p^{(0)} G^k = p^{(k-1)} G$.

Thèse : Montrer que

$$\lim_{k \to \infty} p^{(k)} = \pi$$

$\leadsto$ il suffira de

- partir d'une distribution initiale, e.g. $\begin{pmatrix} 1/n & \cdots & 1/n \end{pmatrix}$
- d'appliquer G de manière itérative
- jusqu'à la précision voulue mesurée par $||p^{(k)} - p^{(k-1)}||$

Méthode itérative pour estimer π

Thèse : $\lim_{k\to\infty} p^{(k)} = \pi$

$$\lim_{k\to\infty} G^k = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \begin{pmatrix} \pi_1 & \cdots & \pi_n \end{pmatrix} = \begin{pmatrix} \pi_1 & \pi_2 & \cdots & \pi_n \\ \vdots & \vdots & & \vdots \\ \pi_1 & \pi_2 & \cdots & \pi_n \end{pmatrix} =: P$$

et

$$[p^{(0)}P]_j = \sum_{i=1}^{n} p_i^{(0)} \pi_j = \pi_j \underbrace{\sum_{i=1}^{n} p_i^{(0)}}_{=1} = \pi_j .$$

En pratique

Une centaine d'itérations suffisent pour obtenir une approximation utilisable et ce calcul peut être réalisé hors ligne, par exemple, une fois par mois, pour mettre à jour le vecteur des scores.

En pratique, on se ramène à la matrice creuse H (possédant de nombreux zéros) :

$$\begin{array}{rcl} p^{(k)} & = & p^{(k-1)} G \\ & = & p^{(k-1)} (\alpha\, S + (1-\alpha)\, \dfrac{J}{n}) \\ & = & \alpha\, p^{(k-1)} S + (1-\alpha)\, \dfrac{\widetilde{e}}{n} \\ & = & \alpha\, p^{(k-1)} (H + a\, \dfrac{\widetilde{e}}{n}) + (1-\alpha)\, \dfrac{\widetilde{e}}{n} \\ & = & \alpha\, p^{(k-1)} H + (\alpha\, p^{(k-1)}\ a + (1-\alpha)) \dfrac{\widetilde{e}}{n} \end{array}$$

où

$$a = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}$$

est tel que $a_i = 1$ si $d^+(i) = 0$ et $a_i = 0$ si $d^+(i) > 0$.