

Graphes, matrices et probabilités: comment Google trouve la page web que vous cherchez

Isar Stubbe (LMPA, Université du Littoral)

Mini-cours à l'EILCO en février-mars 2023

Table de matières.

1. Introduction au logiciel	2
2. Multigraphe dirigé et matrice d'adjacence.....	4
3. Graphe pondéré, matrice stochastique	7
4. Equilibre: vecteur stationnaire	12
5. Matrice de Google	19

Description: Un graphe est un modèle d'un réseau, composé de sommets reliés par des arêtes (voire flèches). On peut par exemple penser à un graphe indiquant les vols directs entre les villes, ou un graphe indiquant les connections entre serveurs, ou encore à un graphe indiquant les hyperliens entre pages web. Pour étudier un graphe, on peut lui associer des matrices: la matrice d'adjacence ou la matrice stochastique. L'intérêt est clair: on dispose ainsi de toute la puissance du calcul matriciel pour formaliser – et résoudre! – des problèmes rencontrés dans la théorie des graphes... et donc aussi dans des situations “réelles” qui se modélisent par un graphe. Dans ce mini-cours, nous allons illustrer cette démarche, en prenant pour principal exemple le réseau formé par les pages web et les hyperliens entre elles—l'internet, donc. Et la principale question que l'on se posera: comment peut-on déterminer la page web “la plus intéressante” parmi un certain nombre de pages reliées entre elles? On découvrira notamment la réponse que Google donne à cette question, pour faire le classement des pages web lorsque vous faites une recherche sur internet. Et on indiquera comment un théorème mathématique – vieux de 100 ans! – est la clé pour trouver cette solution.

Prérequis: Bases du calcul matriciel, notions de probabilité.

Méthodologie: Cours en classe, exercices sur ordinateur (avec Octave/Matlab).

Autres ressources: <http://www-lmpa.univ-littoral.fr/~stubbe>

Evaluation: Questionnaire lors du dernier cours.

1. Introduction au logiciel

Exercice 1. Pour le calcul matriciel, on pourra utiliser Matlab (logiciel payant) ou Octave (logiciel libre, voir www.octave.org). On peut utiliser la Command Window comme une calculatrice capable de faire du calcul matriciel, par exemple:

```
>> 2+2 % on peut calculer
>> A=[1,2,3,4;5,6,7,8;9,10,11,12] % construire une matrice
>> A(2,3) % élément à la position (2,3) dans A
>> A(1,:) % première ligne de A
>> A(:,2) % deuxième colonne de A
>> B=[8,7;6,5;4,3;2,1] % une autre matrice
>> B(1:2,1:2) % construction d'une sous-matrice
>> [A(:,1:2),B(1:3,1)] % assemblage de matrices
>> [A(1:2,1:2);B] % un autre assemblage
>> C=A*B % le produit matriciel
>> C=A*B; % cacher le résultat avec le ";"
>> D=[1,2,3;4,5,6;7,8,9] % une matrice carrée
>> D^3 % puissance 3 de D, équivaut D*D*D
>> E=[1,0,1;0,1,0;1,0,1] % encore une matrice
>> D+E % somme de matrices
>> det(D) % déterminant de D
>> help det % explications concernant la fonction "det"
>> [m,n]=size(A) % le genre d'une matrice (lignes, colonnes)
>> F=ones([m,n]) % matrice 1 partout m x n
>> G=zeros([m,n]) % matrice 0 partout m x n
>> I=eye(n) % matrice unité n x n
```

Au lieu de travailler dans la Command Window, il est bien plus utile de faire un script de toutes les commandes que l'on souhaite exécuter: ainsi on peut sauvegarder (dans un fichier avec extension ".m") et modifier le script au besoin. Cela se fait facilement dans l'Editor intégré, qui permet aussi de lancer le script avec le bouton Run. Sauvegarder le code suivant dans un fichier `exo1.m`, puis exécuter-le:

```
clear all % effacer toutes les variables precedentes
clc % effacer l'ecran du terminal
A=[ 2, 2, 5;
   -2, 4, 8;
    +3,-2,16]
B=A+A^7
det(B)
```

Ce déterminant est égal à _____ .

Exercice 2. Parmi les structures de contrôle et boucles il y a:

<code>if (condition)</code>	<code>for i=1:n</code>	<code>while (condition)</code>
<code>% instructions</code>	<code>% instructions</code>	<code>% instructions</code>
<code>else</code>	<code>endfor</code>	<code>endwhile</code>
<code>% instructions</code>		
<code>endif</code>		

Les conditions et opérateurs logiques usuelles sont:

(i==j) % i égal à j	P & Q % P et Q
(i<=j) % i plus petit que j	P Q % P ou Q
(i>=j) % i plus grand que j	~ P % non P

Par ailleurs, on peut changer le format d'affichage des nombres:

```
>> format bank      % afficher deux décimaux
>> pi
>> pi^100
>> format short     % afficher 4 décimaux et puissance de 10 (affichage par défaut)
>> pi^100
>> format long      % afficher 10 décimaux et puissance de 10
>> pi^100
```

Ecrire un script `exo2.m` pour calculer la somme $\sum_{k=0}^{100} \frac{3^k - 2k}{k + 2}$.

La somme vaut _____ .

Exercice 3. Le logiciel Octave permet la création de nouvelles fonctions (en outre des fonctions déjà définies comme `det()`, `rank()`, `sin()`, etc.). Pour cela, on crée (dans Editor) un fichier, toujours avec extension `.m`, dont le nom est égal au nom de la fonction que l'on souhaite créer. Une fonction peut avoir plusieurs entrées et plusieurs sorties, qui peuvent être des matrices; le codage est typiquement:

```
function [y1,y2,...]=ma_fonction(x1,x2,...)
% instructions
end
```

Donc, les lignes ci-dessus doivent être sauvegardées dans un fichier `ma_fonction.m`, pour qu'Octave reconnaisse cette fonction; et ce fichier doit être placé dans le répertoire de travail. Une fonction peut donc avoir plusieurs entrées et plusieurs sorties, qui peuvent être des nombres ou des matrices. Une fonction peut en appeler d'autres; il suffit que toutes les fonctions (tous les fichiers `.m`) se trouvent dans le répertoire de travail.

Ecrire une fonction `H=matricehilbert(n)` pour construire la *matrice de Hilbert d'ordre* $n \in \mathbb{N}_0$:

$$H_n = \begin{pmatrix} 1 & 1/2 & 1/3 & \dots & 1/n \\ 1/2 & 1/3 & 1/4 & \dots & \vdots \\ 1/3 & 1/4 & 1/5 & & \\ \vdots & & & \ddots & \\ 1/n & \dots & & & 1/(2n-1) \end{pmatrix}$$

Calculer déterminant et rang avec les fonctions intégrées `det()` et `rank()` dans les cas suivants:

dét(H_{10}) = _____ et rang(H_{20}) = _____

2. Multigraphe dirigé et matrice d'adjacence

Exercice 4. Un *multigraphe dirigé* est la donnée d'un nombre fini de sommets reliés par un nombre fini de flèches. Pour des raisons pratiques, on écarte le graphe avec 0 sommets. Si on numérote les sommets du graphe de 1 à n , alors la *matrice d'adjacence* $A = (a_{ij})_{ij} \in \mathbb{R}^{n \times n}$ du graphe est définie par¹

$$a_{ij} = \text{nombre de flèches du sommet } j \text{ au sommet } i.$$

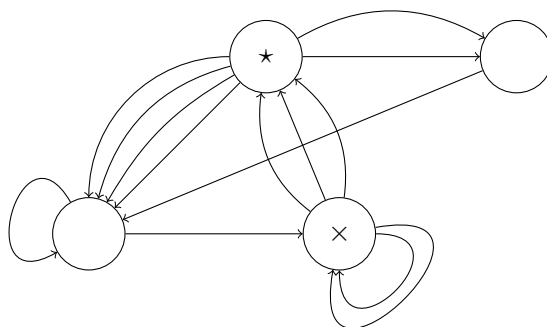
Il n'est pas difficile de montrer que l'élément en position (i, j) dans la matrice A^k (la puissance k -ième, pour $k \in \mathbb{N}$) donne le nombre de *chemins* de longueur k du sommet j au sommet i .

Faisons la démonstration pour A^2 . L'élément en position (i, j) dans $A^2 = A \cdot A$ est, par simple produit matriciel,

$$\sum_t a_{it}a_{tj}.$$

Dans cette formule, chaque terme $a_{it}a_{tj}$ est le produit du nombre de flèches $j \rightarrow t$ avec le nombre de flèches $t \rightarrow i$: c'est donc le nombre total de chemins $j \rightarrow t \rightarrow i$. Lorsqu'on fait la somme sur tous les sommets t (on fixe donc le sommet de départ j et le sommet d'arrivée i , mais on laisse varier le sommet intermédiaire t), on obtient en effet le nombre de chemins de longueur 2 de j à i . (Pour A^k on peut procéder par induction sur k .)

Soit maintenant le multigraphe dirigé



La matrice d'adjacence est

$$A = \begin{pmatrix} & & & \\ & & & \\ & & & \\ & & & \end{pmatrix}.$$

Faire un script `exo4.m` pour calculer:

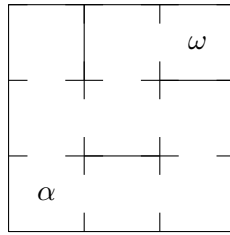
Le nombre de chemins de longueur 7 du sommet '×' au sommet '★' est _____ .

Le nombre de chemins de longueur au plus 15 du sommet '×' au sommet '★' est _____ .

Le nombre de chemins de longueur entre 27 et 113 du sommet '×' au sommet '★' est _____ .

¹Attention! Il y a *deux conventions* pour définir la matrice d'adjacence d'un graphe dirigé. Dans ce cours, nous avons adopté la *convention usuelle en sciences mathématiques*: a_{ij} correspond au nombre de flèches de j à i . En *sciences économiques et sociales*, où les graphes dirigés jouent aussi un rôle important, on note souvent a_{ij} pour le nombre de flèches de i à j ("dans le sens opposé", donc). Au final, la matrice d'adjacence "des mathématiciens" est la transposée de la matrice "des économistes".

Exercice 5. Dans le labyrinthe suivant, partant de la case α , on se déplace toujours à une case voisine, et une fois qu'on est à la case ω , on y reste. (Cela veut dire que toutes les portes entre les cases sont bi-directionnelles, sauf la porte vers la case ω , qui est "en sens unique".)



Le graphe représentant cette situation est:

La matrice d'adjacence de ce graphe est $A = \left(\begin{array}{c} \\ \\ \\ \end{array} \right)$.

En faisant au plus 23 pas, le nombre de chemins de α à ω est: _____ .

Imaginer une application réelle pour ces “labyrintes”:

Exercice 6. Un multigraphe dirigé est dit *irréductible* (aussi *fortement connexe*) si, pour toute paire de sommets, il existe au moins un chemin de l’un à l’autre. (“On peut aller de partout à partout.”)

- (a) Dessiner quelques exemples de graphes irréductibles et de graphes non-irréductibles.
- (b) Montrer que, s’il existe un chemin entre deux sommets dans un graphe à n sommets, alors il existe toujours un chemin de longueur au plus $n - 1$.
- (c) Soit A la matrice d’adjacence d’un graphe donné à n sommets. Montrer que ce graphe est irréductible si et seulement si tout élément de $A^0 + A^1 + \dots + A^{n-1}$ est strictement positif. De manière générale, on dit qu’une matrice $A \in \mathbb{R}^{n \times n}$ à éléments positifs est *irréductible* si tout élément de $A^0 + A^1 + \dots + A^{n-1}$ est *strictement* positif.

Exercice 7. Un multigraphe dirigé est dit *primitif* s’il existe un nombre $k > 0$ tel que, pour toute paire de sommets, il existe un chemin de longueur k de l’un à l’autre. (“On peut aller de partout à partout, et toujours avec un chemin de même longueur.”)

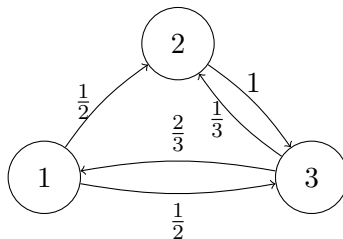
- (a) Dessiner quelques exemples de graphes primitifs et de graphes non-primitifs.
- (b) Comment peut-on vérifier la primitivité d’un graphe par un calcul avec sa matrice d’adjacence?
- (c) Formuler une définition mathématique pour la *primitivité* d’une matrice $A \in \mathbb{R}^{n \times n}$ à éléments positifs. Montrer qu’une matrice primitive est toujours irréductible, mais que l’implication réciproque est fausse.

3. Graphe pondéré, matrice stochastique

Exercice 8. On appelle une colonne $X \in \mathbb{R}^{n \times 1}$ un *vecteur stochastique* si tous ses éléments sont positifs et leur somme vaut 1. Une *matrice stochastique* (aussi appelée *matrice de Markov*) est une matrice carrée dont chaque colonne est un vecteur stochastique.

A toute matrice stochastique $M = (m_{ij})_{ij} \in \mathbb{R}^{n \times n}$ on peut associer un graphe pondéré (et *vice versa*): ce graphe a n sommets, et de tout sommet j à tout sommet i une flèche “pondérée” par le poids m_{ij} . Un tel graphe visualise un *système probabiliste discret* (aussi appelé *chaîne de Markov*): chaque sommet est un état du système, chaque flèche indique la probabilité de transition d’un état à un autre.

On considère maintenant le jeu suivant: on choisit une de trois positions initiales, et ensuite le hasard décide quel mouvement on fait, selon les probabilités données dans le graphe suivant:



La matrice stochastique associée au graphe est $M = \begin{pmatrix} & & \\ & & \\ & & \end{pmatrix}$.

Calculer:

$$M \cdot \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} \\ \\ \end{pmatrix}, \quad M^2 \cdot \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} \\ \\ \end{pmatrix}, \quad M^3 \cdot \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} \\ \\ \end{pmatrix}.$$

Que représentent ces résultats?

Calculer:

$$M^{100} \cdot \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} \\ \\ \end{pmatrix}, \quad M^{100} \cdot \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} \\ \\ \end{pmatrix}, \quad M^{100} \cdot \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} \\ \\ \end{pmatrix}.$$

Constat?

Calculer aussi:

$$M \cdot \begin{pmatrix} \frac{1}{3} \\ \frac{1}{3} \\ \frac{1}{3} \end{pmatrix} = \begin{pmatrix} \\ \\ \end{pmatrix} \quad M^3 \cdot \begin{pmatrix} \frac{1}{3} \\ \frac{1}{3} \\ \frac{1}{3} \end{pmatrix} = \begin{pmatrix} \\ \\ \end{pmatrix} \quad M^5 \cdot \begin{pmatrix} \frac{1}{3} \\ \frac{1}{3} \\ \frac{1}{3} \end{pmatrix} = \begin{pmatrix} \\ \\ \end{pmatrix}.$$

Constat?

Exercice 9. On a deux boîtes, appelées A et B . La boîte A contient 2 jetons avec le label a et 1 jeton avec le label b , alors que la boîte B contient 1 jeton avec le label a et 1 jeton avec le label b . On pioche au hasard un jeton dans la boîte A : si c'est un jeton a alors on reste sur la boîte A ; si c'est un jeton b alors on passe à la boîte B . On remet le jeton dans sa boîte, puis on pioche un nouveau jeton (dans la boîte A ou la boîte B , dépendant de la situation précédente), et on répète la procédure.

Le graphe probabiliste associé à ce procédé est:

La matrice stochastique associée au graphe est $M = \begin{pmatrix} & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \end{pmatrix}$.

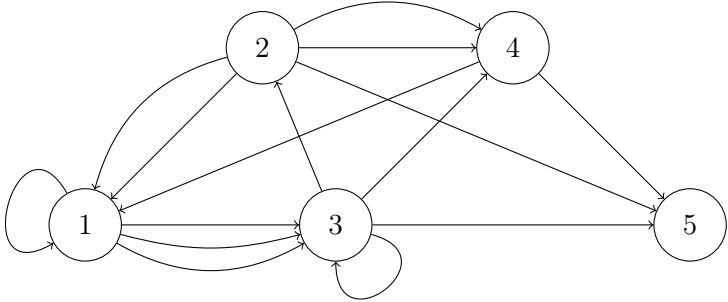
Si à chaque tirage on note le label du jeton, alors après 5 tours de jeu on obtient une séquence de longueur 5 de a 's et de b 's, comme par exemple $aabab$. La probabilité que la dernière lettre est un a est _____ .

Et si on fait 20 tours de jeu, la probabilité que la dernière lettre de la séquence est un a est _____ .

Et si on ne commence pas par piocher dans la boîte A mais plutôt dans la boîte B , alors après 20 tours de jeu la probabilité que la dernière lettre de la séquence est un a est _____ .

Exercice 10. Soit la matrice $A = \begin{pmatrix} \frac{1}{2} & y & \frac{1}{2} \\ x & \frac{3}{8} & 0 \\ \frac{1}{4} & \frac{1}{4} & z \end{pmatrix}$.

Il s'agit d'une matrice stochastique si on pose
 $x = \text{_____}$ $y = \text{_____}$ $z = \text{_____}$
 et elle correspond alors au graphe probabiliste suivant:



Exercice 11. Soit le multigraphe dirigé suivant (dont les sommets sont déjà numérotés):

La matrice d'adjacence associée au graphe est $A = \left(\begin{array}{c} \\ \\ \\ \\ \end{array} \right)$.

On observe qu'aucune flèche ne part du sommet 5. En termes de la matrice A cela se traduit ainsi:

Imaginons maintenant que l'on se trouve au sommet 2 du graphe. Au total 5 flèches partent de ce sommet: deux flèches vers le sommet 1, deux flèches vers le sommet 4, et une flèche vers le sommet 5. Sous l'hypothèse naturelle d'équiprobabilité des flèches, on peut calculer des *probabilités de transition* par une simple fraction: la probabilité de passer du sommet 2 au sommet 1 est de $\frac{2}{5}$, la probabilité de passer du sommet 2 au sommet 4 est aussi de $\frac{2}{5}$, et la probabilité de passer du sommet 2 au sommet 5 est de $\frac{1}{5}$.

Compléter le graphe pondéré reprenant les 5 sommets et les probabilités de transition entre eux:

A ce graphe probabiliste s'associe maintenant tout naturellement une matrice stochastique—ou presque! Car le sommet 5 n'a pas de flèches sortantes... Pour éviter une “probabilité nulle” pour quitter le sommet 5, on fixe la probabilité de transition du sommet 5 à chaque sommet à $\frac{1}{5}$.

La matrice stochastique associée ainsi construite est $M = \left(\begin{array}{c} \\ \\ \\ \\ \end{array} \right)$.

Mais le passage de la matrice d'adjacence A à la matrice stochastique M peut être fait directement...

Expliquer comment:

Finalement, créer une fonction `M=stoch(A)` qui transforme une matrice A à éléments positifs en une matrice stochastique suivant la procédure expliquée ci-dessus (en tenant compte d'éventuelles colonnes nulles!); on pourra utiliser la fonction `sum()` pour calculer la somme d'une colonne.

Exercice 12. Montrer les assertions suivantes.

- (a) Si M est une matrice stochastique et S est une colonne stochastique, alors aussi la colonne $M \cdot S$ est stochastique (lorsque le produit matriciel est bien défini).
- (b) Si M et N sont deux matrices stochastiques de même taille, alors pour tout $0 \leq \alpha \leq 1$ aussi $K = \alpha \cdot M + (1 - \alpha) \cdot N$ est stochastique. On dit que K est une *combinaison convexe* de M et N .
- (c) Pour tout $n > 0$, la matrice $n \times n$ dont chaque élément vaut $\frac{1}{n}$ est stochastique. On va la noter U_n dans la suite.
- (d) Soit M une matrice stochastique carrée, de taille $n \times n$. Pour tout $0 < \alpha < 1$, la matrice $\alpha \cdot M + (1 - \alpha) \cdot U_n$ est stochastique et à éléments strictement positifs; il suit donc que cette combinaison convexe est une matrice primitive.

4. Equilibre: vecteur stationnaire

Exercice 13. Faire un script `exo13.m` avec les commandes suivantes:

```
A=randi([1,99],10,10) % une matrice aléatoire 10x10 d'entiers entre 1 et 99
M=stoch(A)           % on la transforme en matrice stochastique
E=eig(M)             % calcul des valeurs propres de M
max(E)               % calcul de la plus grande valeur propre de M
```

et exécuter quelques fois ce script (pour renouveler à chaque fois la matrice aléatoire A).

Constat?

Exercice 14. On dit qu'une colonne stochastique $S \in \mathbb{R}^{n \times 1}$ est un *vecteur stationnaire* d'une matrice stochastique $M \in \mathbb{R}^{n \times n}$ si $MS = S$; autrement dit, il s'agit d'un *vecteur propre stochastique pour la valeur propre $\lambda = 1$* . Le *Théorème de Perron-Frobenius*² assure que toute matrice stochastique irréductible $M \in \mathbb{R}^{n \times n}$ a un unique vecteur stationnaire $S = (s_1, \dots, s_n)$. Ce vecteur S est donc la *seule* solution au système linéaire

$$\begin{cases} M \cdot \begin{pmatrix} s_1 \\ \vdots \\ s_n \end{pmatrix} = \begin{pmatrix} s_1 \\ \vdots \\ s_n \end{pmatrix} \\ s_1 + \dots + s_n = 1 \end{cases}$$

qui s'écrit de manière équivalente comme

$$\begin{pmatrix} M - I_n \\ \hline 1 \quad \dots \quad 1 \end{pmatrix} \begin{pmatrix} s_1 \\ \vdots \\ s_n \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}$$

Coder une fonction `S=vectstat(M)` pour effectuer ce calcul—pour cela, poser

$$A = \begin{pmatrix} M - I_n \\ \hline 1 \quad \dots \quad 1 \end{pmatrix} \quad \text{et} \quad B = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}$$

puis utiliser la commande `S=linsolve(A,B)` pour résoudre le système linéaire $AS = B$. Reprendre ensuite les Exercices 8 et 9: calculer les vecteurs stationnaires des matrices stochastiques respectives, et comparer les résultats avec les calculs déjà faits.

²D'après les mathématiciens allemands Oskar Perron (1880-1975) et Ferdinand Georg Frobenius (1849-1917). Pour l'énoncé complet du théorème général et sa démonstration, consulter une bonne référence (par exemple le livre *Théorie des matrices* de Félix R. Gantmacher). Ici nous n'avons besoin que du cas particulier des matrices stochastiques, où l'énoncé se simplifie considérablement.

Constat?

C'est encore le *Théorème de Perron-Frobenius* qui explique ce résultat: pour une matrice stochastique primitive M , on peut approcher son vecteur stationnaire S par

$$S \approx M^k \cdot U \quad \text{avec } U \text{ un vecteur stochastique quelconque et } k \text{ "assez grand".}$$

Ce calcul s'appelle la *Méthode de la Puissance Itérée* ('*Power Method*'). Pour la plus grande efficacité on prend souvent la distribution uniforme pour U .

Exercice 15. Pour leurs vols européens, une compagnie aérienne propose deux formules: un billet moins cher mais une seule pièce de bagage à main est autorisée, puis un billet plus cher mais avec la possibilité de prendre jusque 35kg de bagage. Peu après le lancement de ces deux formules, on a constaté que 40% des voyageurs ayant pris la formule la moins chère en sont contents et reprennent un tel billet pour leur prochain voyage; mais 60% change donc de formule. D'autre part, parmi ceux qui ont pris la formule avec plus de bagage, 80% en sont contents alors que 20% changent de formule.

Dessiner le graphe pondéré mettant en situation ces données:

Donner la matrice stochastique $M = \begin{pmatrix} & \end{pmatrix}$.

Calculer le vecteur stationnaire $S = \text{vectstat}(M) = \begin{pmatrix} & \end{pmatrix}$.

Comment cette information peut-elle être utile pour la compagnie aérienne?

Exercice 16. Pour visualiser comment le calcul du vecteur stationnaire peut se faire par la Méthode de la Puissance Itérée, il convient de faire une représentation graphique de chaque étape du calcul. L'idée est donc, pour une matrice stochastique (primitive) M et une distribution initiale de probabilités U_0 données, de calculer et afficher chaque élément de $U_k = M^k \cdot U_0$ en fonction de $k \geq 0$. Pour cela, on peut utiliser la capacité de Octave de faire un *plot*. Coder la fonction suivante, qui pour une matrice stochastique M , un vecteur stochastique U , et une puissance k , affiche la convergence vers $U_k = M^k \cdot U$:

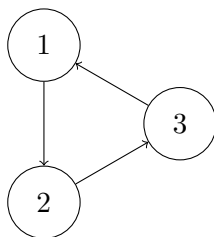
```

function visualiser_vectstat(M,U,k)
% on note le genre de M
[m,n]=size(M);
% on calcule les distributions finales pour chaque t inférieur à k:
S=zeros(m,k+1);
for t=0:k
    S(:,t+1)=M^t*U;
endfor
% on affiche les colonnes de S:
x=[0:1:k];
plot(x,S,'LineWidth',2);
endfunction

```

Appliquer cette fonction à diverses valeurs de M , U et k (p.e. les données des exercices précédents), pour observer la convergence du calcul de $M^k \cdot U$.

Exercice 17. Soit le graphe suivant:



On accorde à chaque flèche le poids 1.

La matrice stochastique et son vecteur stationnaire sont

$$M = \begin{pmatrix} & & \\ & & \\ & & \end{pmatrix} \quad \text{et} \quad S = \text{vectstat}(M) = \begin{pmatrix} \\ \\ \end{pmatrix}.$$

Si on tente d'approximer le vecteur stationnaire par la Méthode des Puissances Itérées, on observe en particulier que

$$M \cdot \begin{pmatrix} p_1 \\ p_2 \\ p_3 \end{pmatrix} = \begin{pmatrix} \\ \\ \end{pmatrix}, \quad M^2 \cdot \begin{pmatrix} p_1 \\ p_2 \\ p_3 \end{pmatrix} = \begin{pmatrix} \\ \\ \end{pmatrix}, \quad M^3 \cdot \begin{pmatrix} p_1 \\ p_2 \\ p_3 \end{pmatrix} = \begin{pmatrix} \\ \\ \end{pmatrix}, \quad \text{etc.}$$

et par conséquent...

Explication: la matrice stochastique M est irréductible et donc on peut bien calculer un unique vecteur stationnaire; mais M n'est pas primitive et donc la Méthode des Puissances Itérées n'est pas garantie de donner une bonne approximation du vecteur stationnaire!

Exercice 18. *Quelle est l'équipe la plus forte?* Aux JO de Rio en 2016, les équipes de volleyball féminin en Groupe A étaient (en ordre alphabétique) l'Argentine, le Brésil, le Caméroun, la Corée du Sud, le Japon et la Russie. Chaque équipe a rencontré chaque autre équipe une seule fois, et les résultats (fictifs) étaient les suivants:

JA	1 - 3	CO	BR	3 - 1	AR	RU	3 - 2	JA
BR	3 - 2	CA	RU	3 - 0	CA	BR	3 - 0	CO
RU	3 - 0	AR	CO	3 - 0	AR	CO	3 - 1	CA
JA	3 - 1	CA	BR	3 - 0	JA	JA	3 - 1	AR
RU	3 - 1	CO	AR	3 - 2	CA	BR	3 - 2	RU

Donc, par exemple, le Japon a gagné 1 set contre la Corée du Sud, alors que la Corée du Sud à gagné 3 sets contre le Japon; c'est donc ce dernier qui a gagné le match. On considère désormais le multigraphe dont les sommets sont les 6 pays participants, numérotés de 1 à 6 suivant leur ordre alphabétique, et où il y a une flèche du sommet x au sommet y **pour chaque set** que x a gagné contre y .

La matrice de multi-adjacence de ce graphe est: $A =$ $\left(\begin{array}{cccccc} & & & & & \\ & & & & & \\ & & & & & \\ & & & & & \\ & & & & & \\ & & & & & \end{array} \right)$.

Calculer la matrice stochastique: $M = \text{stoch}(A) =$ $\left(\begin{array}{cccccc} & & & & & \\ & & & & & \\ & & & & & \\ & & & & & \\ & & & & & \\ & & & & & \end{array} \right)$.

Calculer le vecteur stationnaire: $S = \text{vectstat}(M) =$

$$\begin{pmatrix} \\ \\ \\ \\ \\ \end{pmatrix}.$$

Le classement déterminé à l'aide du vecteur stationnaire du graphe (du "plus fort" au "plus faible") est:

1. _____
2. _____
3. _____
4. _____
5. _____
6. _____

Exercice 19. Voici les résultats des matches de foot dans le groupe D au premier tour du Mondial 2022:

Danmark	0-0	Tunésie
France	4-1	Australie
Tunésie	0-1	Australie
France	2-1	Danmark
Australie	1-0	Danmark
Tunésie	1-0	France

Avec 3 points pour un match gagné, 1 point pour un match nul, et 0 points pour un match perdu, cela donne le classement suivant:

1. _____
2. _____
3. _____
4. _____

Dessiner le multi-graphe dont les sommets sont les pays, numérotés par ordre alphabétique, et on ajoute une flèche de x à y pour chaque but que x a marqué contre y :

Donner la matrice de multi-adjacence du graphe: $A = \begin{pmatrix} & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \end{pmatrix}$.

Calculer la matrice stochastique associée: $M = \text{stoch}(A) = \begin{pmatrix} & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \end{pmatrix}$.

Calculer le vecteur stationnaire $S = \text{vectstat}(M) = \begin{pmatrix} \\ \\ \\ \\ \end{pmatrix}$.

Le classement déterminé à l'aide du vecteur stationnaire du graphe (du "plus fort" au "plus faible") est:

1. _____
2. _____
3. _____
4. _____

Exercice 20. *Qui est le meilleur de la classe?* Voici un tableau avec les résultats des examens (sur 20):

	Math	Phys	Chim	Info	Econ
Anne	14	12	15	11	9
Blaise	9	16	11	14	11
Christelle	17	19	15	16	11
Diego	13	9	12	12	15

Si on calcule les moyennes des notes, on obtient le classement suivant:

1. _____
2. _____
3. _____
4. _____

Si on fait un multigraphe dont les sommets sont les étudiants, et dans lequel on ajoute une flèche de x à y pour chaque examen que x a fait mieux que y , alors selon le vecteur stationnaire de la matrice stochastique de ce multigraphe, on obtient le classement suivant:

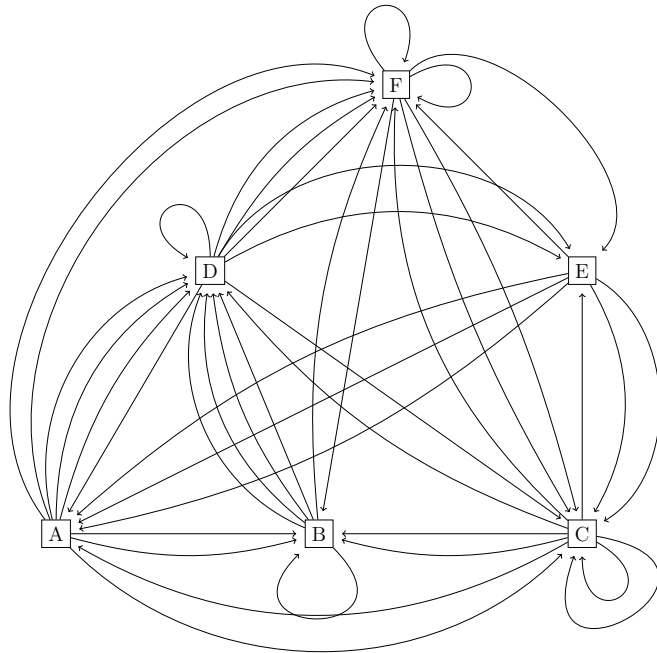
1. _____
2. _____
3. _____
4. _____

On peut aussi faire un multigraphe dont les sommets sont toujours les étudiants, mais on ajoute une flèche de x à y pour chaque point que x a plus que y . Selon le vecteur stationnaire de la matrice stochastique de ce deuxième multigraphe, on obtient le classement suivant:

1. _____
2. _____
3. _____
4. _____

5. Matrice de Google

Exercice 21. On considère le multigraphe suivant:



Numéroter ses sommets par ordre alphabétique et donner la matrice de multi-adjacence A , la matrice stochastique M , puis le vecteur stationnaire S de M :

$A =$

$\left(\begin{array}{c} \\ \\ \\ \\ \\ \end{array} \right)$

$M = \text{stoch}(A) =$

$\left(\begin{array}{c} \\ \\ \\ \\ \\ \end{array} \right)$

$S = \text{vectstat}(M) =$

$\left(\begin{array}{c} \\ \\ \\ \\ \\ \end{array} \right)$

Désormais on interprète ce multigraphe comme un réseau de pages web liées entre elles. On s’imagine une personne qui visite une première page web (choisie au hasard), puis suit un des hyperliens de cette page vers une autre page, etc. Si on répète ce procédé *ad infinitum*, alors la page la plus souvent visitée (donc “la page la plus importante”) est donnée par le score le plus élevé dans le vecteur stationnaire du graphe. Il y a, pourtant, un double problème avec ce modèle:

- (i) un graphe n’admet pas nécessairement un (unique) vecteur stationnaire (notamment s’il n’est pas irréductible!),
- (ii) un surfeur sur internet ne suit pas *ad infinitum* les hyperliens qui lui sont présentés.

Une légère modification de la matrice stochastique du graphe résout ces deux problèmes—et c’est ça l’invention de Larry Page, cofondateur de Google en 1998. On introduit pour cela la *matrice de Google* du graphe:

$$G = \alpha M + (1 - \alpha) \begin{pmatrix} 1/n & \dots & 1/n \\ \vdots & & \vdots \\ 1/n & \dots & 1/n \end{pmatrix}$$

où M est la matrice stochastique du graphe, n est le nombre total de pages (ici $n = 6$), et $\alpha \in [0, 1]$ est un paramètre. D’un côté, ce paramètre α exprime la “fidélité” du surfeur: il y a une probabilité α que le visiteur suit les hyperliens proposés, et une probabilité $1 - \alpha$ qu’il choisira au hasard une nouvelle page de départ. De l’autre côté, cette nouvelle matrice stochastique G est *primitive* pour tout $\alpha \neq 1$, et on peut donc calculer son (unique!) vecteur stationnaire efficacement par la Méthode des Puissances Itérées.

Coder une fonction `G=google(M,alpha)` pour effectuer ce calcul.

Pour $\alpha = 0.85$, calculer:

$$G = \text{google}(M, 0.85) = \begin{pmatrix} & & & & & \\ & & & & & \\ & & & & & \\ & & & & & \\ & & & & & \\ & & & & & \end{pmatrix} \quad \text{et} \quad S = \text{vectstat}(G) = \begin{pmatrix} \\ \\ \\ \\ \\ \end{pmatrix}$$

Selon ce modèle, quelle est donc le classement des pages web de “la plus importante” à “la moins importante”?

C’est exactement avec un tel procédé que Google classe les pages web lors d’une recherche sur internet; cet algorithme est désormais connu comme le ‘PageRank’ algorithme. On estime que la matrice que Google utilise, est de taille $3 \cdot 10^{10} \times 3 \cdot 10^{10}$; pour le calcul de son vecteur stationnaire, Google utilise la Méthode de la Puissance Itérée, et on estime que Google fait entre 50 et 100 itérations. On pense que la valeur de α est autour de 0.85 (mais c’est un secret industriel). Le calcul prend plusieurs jours.

Exercice 22. Deux candidats à la Présidentielle, Bernard et Cathérine, s’affrontent lors d’un débat. Au lendemain du débat, une enquête montre que, parmi les gens qui comptaient auparavant voter pour Bernard, 85% confirme son choix alors que 7% va finalement voter pour Cathérine; le reste décide de s’abstenir. Parmi les gens qui comptaient voter pour Cathérine, 23% change de candidat et 19% s’abstiendra, alors que le reste confirme son choix pour Cathérine. Finalement, de ceux qui voulaient s’abstenir, 40% va finalement quand-même voter pour Bernard et 17% pour Cathérine, et le reste ne votera donc pas.

Le graphe représentant cette situation est:

La matrice stochastique de ce graphe, et son vecteur stationnaire, sont donc

$$M = \begin{pmatrix} & \\ & \end{pmatrix} \quad \text{et} \quad S = \begin{pmatrix} \\ \end{pmatrix} .$$

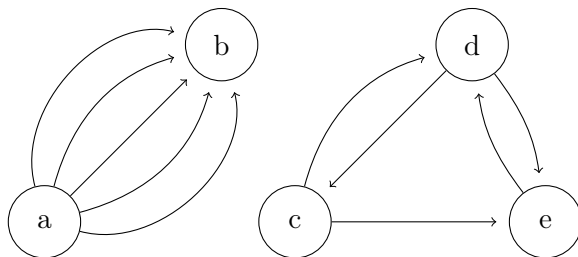
Mais l'institut de sondage estime que seul 92% des électeurs interrogés tiendra parole. Pour tenir compte de cette volatilité, l'institut décide d'utiliser une matrice de google.

La matrice utilisée sera donc $G = \text{google}(M, \alpha)$ avec $\alpha = \underline{\hspace{2cm}}$.
 Ainsi on trouve la matrice et son vecteur stationnaire suivant:

$$G = \begin{pmatrix} & \\ & \end{pmatrix} \quad \text{et} \quad S' = \begin{pmatrix} \\ \end{pmatrix} .$$

Formuler à l'aide de ces calculs une prévision du résultat de l'élection:

Exercice 23. *Quel est l'impact d'une publication scientifique? On interprète le multigraphe*

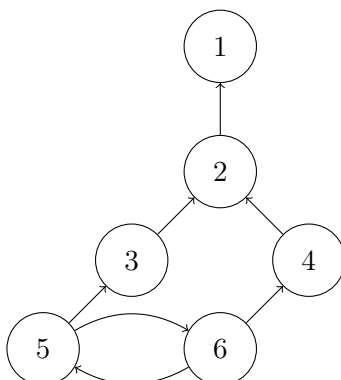


comme une représentation de cinq publications scientifiques et les références qu'elles font les unes aux autres: chaque flèche d'un sommet x à un sommet y représente donc une citation de y dans la publication x . Quelle est la publication "la plus importante"?

Première approche: "la publication la plus importante est celle qui est la plus souvent citée". C'est donc la publication _____ .

Deuxième approche: "la publication la plus importante est celle qui a le meilleur score dans le vecteur stationnaire de une matrice stochastique adéquate associée au graphe". Faire toutes les démarches nécessaires (en observant notamment que ce graphe n'est pas irréductible!) pour conclure que c'est donc la publication _____ .

Exercice 24. Soit le graphe



Selon le vecteur stationnaire de la matrice de Google du graphe, prenant $\alpha = 0.6$, le classement des sommets de ce graphe est le suivant: _____ .
 Selon le vecteur stationnaire de la matrice de Google du graphe, prenant $\alpha = 0.7$, le classement des sommets de ce graphe est le suivant: _____ .
 Morale?

Exercice 25. Montrer que, pour toute matrice stochastique M et tout $0 < \alpha < 1$, la matrice de Google $G = \alpha M + (1 - \alpha)U_n$ est stochastique et primitive; cela implique que l'on peut calculer son vecteur stationnaire efficacement par la Méthode des Puissances Itérées. (On rappelle que U_n est la notation pour la matrice $n \times n$ dont chaque élément vaut $\frac{1}{n}$.)