

Ranking et recommandation – TD

M1 AMIS Informatique

Version du 10 février 2025

Franck Quessette – franck.quessette@uvsq.fr

Table des matières

1	TD 1 – Représentation matricielle du graphe du web	2
2	TD 2 – Le vrai Pagerank	4
3	TD 3 – Manipulations de matrices	5
4	TD 4 – Recommandation	7
5	TD 5 – SVD	9

1 TD 1 – Représentation matricielle du graphe du web

Exercice 1.1 Graphe du web

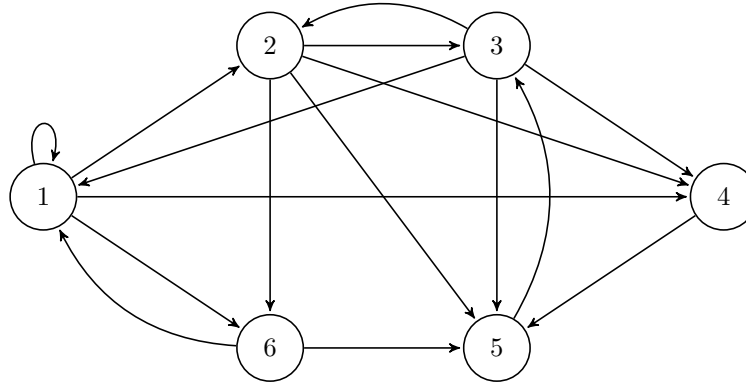


FIGURE 1 – Graphe 1 du WEB

1. Donnez pour le graphe du web de la figure 1 une représentation sous forme de matrice pleine stochastique en utilisant la méthode de Pagerank (inspirez vous d'une matrice d'adjacence). Écrire ce graphe dans un fichier texte pour être lu par votre futur programme de recherche de note de pertinence.
2. Donnez la complexité en taille de votre structure si le graphe est de taille N .
3. Proposez une représentation creuse de la structure codée dans un fichier pour n'écrire que les arcs qui font partie du graphe. La taille de la structure doit être le nombre d'arcs multiplié par une constante.
4. Proposez une représentation creuse de la structure en mémoire avec la même complexité en taille et dont la complexité pour l'opération "produit à gauche d'un vecteur par une matrice creuse" est le nombre d'arcs multiplié par une constante.
5. L'algorithme des puissances consiste à itérer $\pi^{(k+1)} = \pi^{(k)}P$ jusqu'à convergence. L'initialisation est $\pi^{(0)} = 1/N \mathbf{e}$ où N est la taille de la matrice. Le test de fin est $\|\pi^{(k+1)} - \pi^{(k)}\|_1 \leq \varepsilon$. \mathbf{e} est un vecteur ligne rempli de 1 et ε est fixé à 10^{-6} . Codez l'algorithme des puissances. Faites le test de votre programme avec la matrice pleine construite en 1.
6. Donnez la complexité en temps de votre algorithme si le graphe est de taille N .
7. Vérifiez vos résultats avec SCILAB. Le vecteur solution peut être trouvé en élevant la matrice P à une puissance grande (100 devrait suffire).
8. Codez l'algorithme des puissances avec la matrice creuse construite en 2 et 3.
9. Donnez la complexité en temps de votre algorithme pour matrice creuse si le graphe a N sommets et M arcs.

Rappel sur les normes, pour un vecteur $x = (x_1, x_2, \dots, x_n)$, on a :

$$\begin{aligned}
 \|x\|_1 &= \sum_{i=1}^n |x_i| \\
 \|x\|_2 &= \left(\sum_{i=1}^n x_i^2 \right)^{\frac{1}{2}} \\
 &\dots \\
 \|x\|_\infty &= \sup_{i=1, \dots, n} |x_i|
 \end{aligned}$$

Exercice 1.2 Graphe du web 2

1. Écrire ce nouveau graphe dans un fichier texte en format matrice pleine pour être lu par votre futur programme de recherche de note de pertinence.

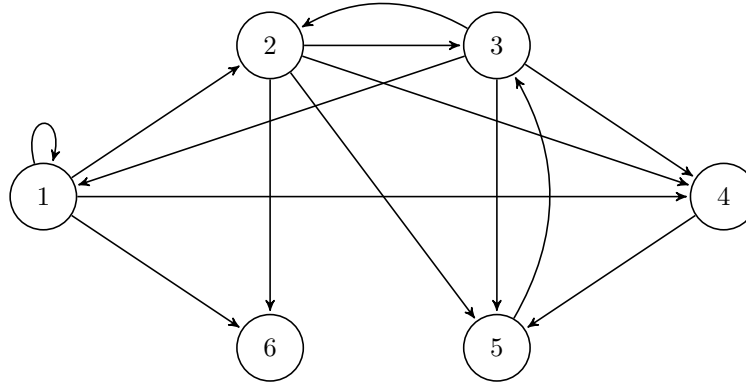


FIGURE 2 – Graphe 2 du WEB

2. Vérifier la convergence de votre algorithme.
3. Vérifier vos résultats avec SCILAB.
4. Y-a-t il un problème ?
5. Implémentez la solution proposée par Pagerank pour résoudre ce problème ?

Exercice 1.3 *Valeurs propres*

Soit P la matrice :

$$P = \begin{pmatrix} 0.6 & 0.4 \\ 0.8 & 0.2 \end{pmatrix}$$

1. Vérifiez que 1 est valeur propre à gauche et à droite.
2. Calculez l'autre valeur propre.
3. Calculez les vecteurs propres à gauche et à droite.
4. Vérifiez vos calculs avec Scilab.

2 TD 2 – Le vrai Pagerank

Site web de matrices creuses : <https://sparse.tamu.edu>.

Rappels

Soit

- P la matrice stochastique
- N la taille de la matrice P
- α fixé
- e un vecteur ligne de taille N ne contenant que des 1
- f est un vecteur ligne de taille N tel que $f[i] = 1$ si la ligne i de P ne contient que des zéros et sinon $f[i] = 0$
- f^t est la transposée de f .

On itère sur :

$$\begin{cases} x^{(0)} = (1/N)e \\ x^{(k+1)} = \alpha x^{(k)} P + [(1-\alpha)(1/N) + \alpha(1/N)(x^{(k)} f^t)]e \end{cases}$$

Exercice 2.1 *Pagerank*

1. Implémenter la version Google (le surfer aléatoire) avec la complexité minimale vue en cours (en fonction du nombre de sommets et d'arcs du graphe du web). Cela signifie en particulier que vous utilisez une structure de matrice creuse permettant une multiplication à gauche avec la complexité minimale. On prend pour le moment $\alpha = 0.85$.
2. Tester votre programme sur les petites matrices du TD1
3. Télécharger la matrice `wb-cs-stanford.mtx` depuis le site de tamu <https://sparse.tamu.edu/Gleich/wb-cs-stanford> au format matrix market ou depuis e-campus, section DATA.
4. Tester votre programme sur cette matrice `wb-cs-stanford.mtx` et vérifier la convergence pour un test d'arrêt avec une précision de 10^{-6} .
5. Faire une courbe avec en abscisse la précision et en ordonnée le nombre d'itérations. Vous prendrez comme précisions de les puissances de 10 de 10^{-3} à 10^{-9} .
6. Garder la précision à 10^{-6} mais changer α . Faire une courbe du nombre d'itérations pour la convergence avec α égal à 0.5, 0.7, 0.85, 0.9, 0.99, 0.999
7. Télécharger toutes les autres matrices (il y en a 6) depuis la section DATA.
8. Tester la convergence sur toutes les autres matrice

Exercice 2.2 *Plus gros*

testez votre code avec :

- <https://sparse.tamu.edu/Gleich/wikipedia-20051105>
- <https://sparse.tamu.edu/Gleich/wb-edu>

3 TD 3 – Manipulations de matrices

Exercice 3.1 Matrice I

- e est un vecteur ligne ne contenant que des 1. Sa taille est donnée par le contexte.
- Id est la matrice identité. Sa taille est donnée par le contexte.
- Donnez la définition d'une matrice stochastique.
- Donnez la définition d'une matrice irréductible.
- Donnez la définition d'une matrice primitive.
- Donnez la définition d'une valeur propre à droite, à gauche.
- Prouver que pour toute matrice stochastique, 1 est valeur propre à gauche.
- La matrice Id est elle une matrice stochastique? Est elle irréductible? Est elle primitive?
- Quels sont les vecteurs propres associées à la valeur propre 1 pour la matrice Id .
Soit M une matrice stochastique quelconque et π tel que $\pi M = \pi$
- Montrez que $Me^t = e^t$
- en déduire que 1 est valeur propre à droite.
- Montrez que M^2 est stochastique et que $\pi M^2 = \pi$
- Montrez que $\pi(1/2M + 1/2Id) = \pi$
- Est ce que la propriété suivante est vraie : si M est irréductible alors M^2 est irréductible (si c'est faux donnez un contre exemple, si c'est vrai donner une preuve).
- Est ce que la propriété suivante est vraie : si M est primitive alors M^2 est primitive (si c'est faux donnez un contre exemple, si c'est vrai donner une preuve).

Exercice 3.2 Graphe du web I

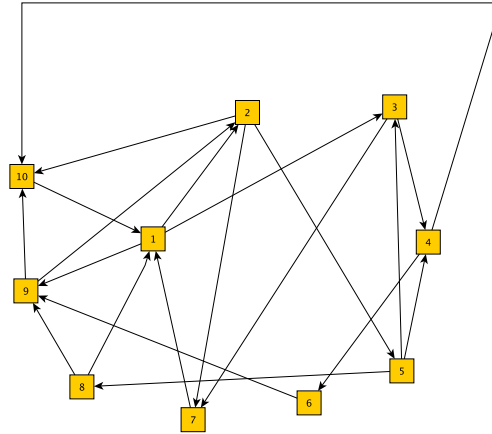
Supposez que le graphe du web soit donné par la matrice suivante :

$$P = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 & 0 & 0 & 0 \\ 1/2 & 1/2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1/2 & 0 & 1/2 \\ 0 & 0 & 0 & 1/2 & 0 & 0 & 1/2 \\ 0 & 0 & 0 & 1/2 & 1/2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1/2 & 1/2 & 0 \end{bmatrix}$$

- Supposons $\pi_0 = (1, 0, 0, 0, 0, 0, 0)$. Avec scilab calculez $\pi_1 = \pi_0 P$. Puis calculez la 10ieme itération π_{10} .
- Supposons $\pi_0 = (1/2, 0, 0, 0, 0, 0, 1/2)$. Avec scilab calculez $\pi_1 = \pi_0 P$. Puis calculez la 10ieme itération π_{10} .
- Est ce que les deux valeurs sont égales? Pourquoi?
- Est ce que la matrice est irréductible?

Exercice 3.3 Graphe du web II

Supposez que le graphe du web ci-dessous :



- Donnez la matrice stochastique associée à ce graphe du web.
- Montrez que le graphe est fortement connexe
- Calculez les pertinences pour ce graphe du web sans faire la modification de Google
- Calculez les pertinences pour ce graphe du web en faisant la modification de Google. Comparez les.

Exercice 3.4 *Graphe du web III*

Supposez que le graphe du web soit un circuit de longueur N , que sera la pertinence de chaque noeud ?

Exercice 3.5 *Matrices II*

- Que vaut $(e^t e)(e^t e)$?
- Soit A une matrice non singulière, que vaut $(eA)(A^{-1}e^t)$

4 TD 4 – Recommandation

Exercice 4.1 *Calcul de Similarité*

L'utilisateur $U1$ a la liste d'achats suivante ($O1, O4, O5, O6, O7, O9, O11, O15$). L'utilisateur $U2$ a la liste d'achats suivante ($O2, O3, O4, O5, O6, O11, O12, O13, O15$).

1. Calculer la similarité entre $U1$ et $U2$
2. $U2$ achète un objet $O1$, que devient la similarité ?
3. Est-il possible que la similarité diminue ?
4. Un utilisateur $U3$ arrive dans le système et achète un objet $O8$, calculer la similarité entre $U3$ et $U2$.
5. Puis il acquiert $O2$ et ensuite $O1$, calculer la similarité entre $U3$ et $U2$ à chaque étape.

Exercice 4.2 *Similarité*

Montrez qu'avec la définition de l'indice de similarité entre clients (vu en cours), pour tout client U , on a $S_{U,U} = 1$ dès que U a fait un achat ou donné une note.

Exercice 4.3 *Recommandation à base de graphe*

On a les archives suivantes pour les achats d'un site marchand (voir ci-dessous).

1. Tracer le graphe quadriparti issu de Joe lié aux achats.
2. Quelle recommandation faire à Joe si on a un algo de recommandations reposant sur un achat unanime des voisins ?
3. Quelle recommandation faire à Joe si on a un algo de recommandations reposant sur un achat de type 2 parmi 3 (c'est à dire que si il y a 3 voisins, il est suffisant que 2 achètent pour faire une recommandation).

Liste des achats par ordre chronologique dans le journal : (les objets portent un numéro)

JOE 1
U4 1
U4 2
U4 3
U4 4
JOE 4
U1 2
U1 1
U1 3
U2 4
U2 12
U4 12
U4 13
JOE 2
JOE 3
U2 13
U5 3
U5 4
U5 14
U5 15
U6 1
U6 2
U6 3
U3 4
U3 3
U3 13

U1 4
 U1 12
 U2 2
 U2 3
 U3 14
 U3 15
 U4 14
 U6 15
 U1 13
 U1 14
 U4 15
 U5 1
 U5 2
 U2 14
 U2 15
 U3 1
 U3 2
 U6 4
 U6 14
 U1 15
 U2 1

Exercice 4.4 *Recommandation*

On étudie un système où les utilisateurs notent les hôtels et qui propose des recommandations. Voici la table des notes (entre 1 et 5). Un point signifie que l'hôtel n'a pas été noté.

Utilisateur	Similarité	Ibis	Mercure	H10	Holliday Inn
Alice	0.9	.	3	5	1
Bob	0.2	.	3	1	1
Charlie	0.7	2	.	.	5
David	0.9	.	3	5	2
Estelle	0.4	1	5	5	4
Franck	0.8	2	.	.	1

La similarité avec l'utilisateur cible est déjà calculée et elle est en colonne 2 de la table.

1. Calculer la note prédite pour chacun des hôtels pour l'utilisateur cible et indiquer l'hôtel proposé (donner tous les détails du calcul).
2. Le premier hôtel dans la liste (Ibis) veut améliorer son classement. Il crée un utilisateur virtuel "Gus" et réussit à avoir une note de similarité égale à 0.8. Dans un premier temps, "Gus" donne la note maximale (5) à Ibis, et ne fait que cela. Que deviennent les notes globales pour les hôtels et que devient la recommandation ?
3. Maintenant, "Gus" donne de mauvaises notes (1) à tous les autres hôtels pour baisser leur note globale. Que deviennent les notes globales pour les hôtels et que devient la recommandation ?

5 TD 5 – SVD

Exercice 5.1 SVD avec SCILAB

Syntaxe : $[U, S, V] = \text{svd}(X)$

Faire la décomposition en valeur singulière de $\begin{bmatrix} 0 & 3 \\ -2 & 1 \\ 0 & 0 \end{bmatrix}$. puis vérifier.

Exercice 5.2 SVD et approximation de rang faible

1. Construire une SVD de la matrice 5×10 qui se trouve dans le fichier SVD2.
2. Proposez une approximation de rang 2 de cette matrice
3. Quelle est l'erreur (en norme infinie) de cette approximation ?

Exercice 5.3 SVD théorique

Supposez qu'on a une décomposition SVD de $A = U\Sigma V^t$. Montrez que cela implique que :

$$A(i, j) = \sum_k U(i, k) \Sigma(k, k) V(j, k)$$

Supposez qu'on a une décomposition SVD de $A = U\Sigma V^t$.

1. Montrez que $AV = U\Sigma$
2. Montrez que Σ^2 contient les valeurs propres à gauche de $A^t A$
3. Montrez que V contient les vecteurs propres à gauche de $A^t A$
4. Montrez que U contient les vecteurs propres à droite de AA^t
5. Montrez que si σ est la racine carrée d'un élément diagonal de Σ alors il existe deux vecteurs u et v vérifiant $\sigma u = Mv$ et $\sigma v = M^t u$.
6. Montrez que si A est une matrice carrée, alors le déterminant de A est le produit des valeurs singulières de A .
7. Montrez que si A a une SVD égale à $U\Sigma V^t$, alors $U^t AV$ est une matrice diagonale. Que vaut elle ?