

Activités sur PageRank

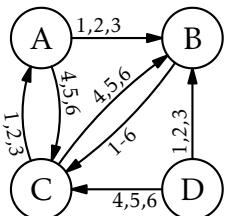
Nom et prénom :

Première approche

EXERCICE 1 : On considère le graphe ci-contre. Chaque sommet correspond à une page. Un flèche allant du sommet A au sommet B signifie que la page A contient un lien pointant vers la page B.

La règle du jeu est la suivante :

- On choisit un sommet de départ.
- On jette le dé et on suit l'arc correspondant à la valeur. Et on recommence.



Avoir "1,2,3" sur un arc veut dire qu'on la prend si on fait 1, 2 ou 3. Par exemple, sur le sommet A, si on fait "3", on va en B. De même "1-6" veut dire qu'on prend l'arc quelque soit la valeur du dé.

On a fait les lancers de dés à l'avance et on a mis les résultats dans le tableau ci-contre.

Dé		2	1	6	5	2	6	1	3	2	2
Sommet	D	B	C								

Puis on a choisi D comme sommet de départ et on a fait les premiers déplacements en suivant les résultats du dé.

- 1) Compléter le tableau ci-dessus pour déterminer les différents sommets visités.
- 2) Lancer le dé 20 fois et compléter la ligne "Dé" du tableau ci-dessous, sauf la case grisée.

Dé											
Sommet											

3) Choisir un sommet de départ et compléter la ligne "Sommet" en tenant compte des résultats du dé.

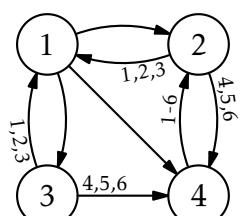
4) Compter le nombre de visites pour chaque sommet. On ne compte pas le sommet de départ.

Sommet	A	B	C	D
Visites				
Classement				

5) Classer les sommets du plus au moins visité.

EXERCICE 2 : On recommence avec le graphe ci-contre.

- 1) Rajouter des valeurs sur les arcs partant du sommet 1 pour qu'on ait autant de chance de prendre chacune.
- 2) Lancer 50 fois le dé et remplir la première ligne des deux tableaux ci-dessous.
- 3) Choisir un sommet de départ et effectuer les déplacements correspondant aux dés.



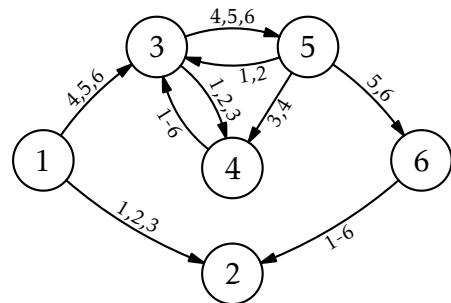
Dé											
Sommet											

- 4) Compléter le tableau ci-contre en mettant les fréquences.
- 5) Établir un classement de ces pages.

Sommet	1	2	3	4
Visites				
Fréquence				
Classement				

EXERCICE 3 : On considère le graphe ci-contre.

- 1) Quel est le problème pour faire une marche aléatoire sur ce graphe?
- 2) Modifier le graphe pour régler ce problème. Attention, il ne faut pas favoriser une page plus que les autres.
- 3) Réaliser 50 lancers du dé pour remplir le tableau ci-dessous, puis déterminer les déplacements.



Dé																				
Sommet																				

- 4) Compléter le tableau ci-contre en mettant les fréquences.
- 5) Établir un classement de ces pages.

Sommet	1	2	3	4	5	6
Visites						
Fréquence						
Classement						

Utilisation des matrices

Il est possible de calculer directement le classement en utilisant des matrices. Une matrice est un tableau. Chaque ligne et chaque colonne correspond à un sommet. Une valeur à l'intersection des lignes i et des colonnes j correspond à la probabilité de passer du sommet i au sommet j . La matrice ci-contre correspond au graphe de l'exercice 1.

EXERCICE 4 : Compléter les matrices associées aux exercices 2 et 3 (avec la version modifiée du graphe). Vous pouvez mettre des fractions.

$$\begin{array}{c|cccc} \curvearrowleft & A & B & C & D \\ \hline A & 0 & 0,5 & 0,5 & 0 \\ B & 0 & 0 & 1 & 0 \\ C & 0,5 & 0,5 & 0 & 0 \\ D & 0 & 0,5 & 0,5 & 0 \end{array}$$

$$\begin{array}{c|cccc} \curvearrowleft & 1 & 2 & 3 & 4 \\ \hline 1 & \dots & \dots & \dots & \dots \\ 2 & \dots & \dots & \dots & \dots \\ 3 & \dots & \dots & \dots & \dots \\ 4 & \dots & \dots & \dots & \dots \\ \hline 5 & \dots & \dots & \dots & \dots \\ 6 & \dots & \dots & \dots & \dots \end{array}$$

Ces matrices permettent de calculer la probabilité d'arriver sur une page donnée en partant d'une page choisie au hasard. Par exemple, pour calculer la probabilité d'arriver à chaque sommet après 20 déplacements, on fait :

$$\begin{bmatrix} 0,25 & 0,25 & 0,25 & 0,25 \end{bmatrix} \times \begin{bmatrix} 0 & 0,5 & 0,5 & 0 \\ 0 & 0 & 1 & 0 \\ 0,5 & 0,5 & 0 & 0 \\ 0 & 0,5 & 0,5 & 0 \end{bmatrix}^{20} \approx [0,22222 \ 0,33333 \ 0,44445 \ 0]$$

Cela veut dire qu'on a environ 22% de chances d'arriver sur la page 1, 33% d'arriver sur la page 2, 45% pour la page 3 et enfin 0% pour la page 4. On peut ainsi établir le classement des pages : 3, 2, 1 puis 4.

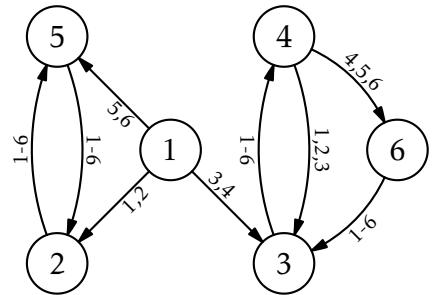
EXERCICE 5 : Utiliser la calculatrice pour établir les probabilités d'arriver sur les différentes pages de l'exercice 2. Comparer les résultats avec ceux de l'exercice 2.

Sommet	1	2	3	4
Fréquence				
Classement				

L'approche de Google

EXERCICE 6: Que va-t-il se passer si on essaie de faire un classement des pages d'après le graphe suivant?

Pour résoudre ce problème, et celui du puit, Larry Page et Sergey Brin ont eu l'idée suivante : à chaque étape avant de déterminer le prochain lien à suivre, il faut d'abord faire un premier tirage qui permet de savoir si on va suivre un des liens ou si on va sauter directement vers n'importe quelle page. Ils ont fixé une probabilité de 85% de suivre les liens et de 15% de sauter vers n'importe quelle page.



EXERCICE 7 : Pour appliquer cette nouvelle méthode, voici la nouvelle marche à suivre :

- Choisir un sommet de départ.
 - Jeter un dé :

- Avec un 6 : rejeter un dé et se rendre au sommet correspondant.
 - Sinon relancer le dé et suivre le lien correspondant.

- 1) Réaliser 50 lancers de dé pour remplir la première ligne des deux tableaux ci-dessous.
 - 2) Choisir un sommet de départ et réaliser les déplacements correspondants aux lancers du dé. Les lancers au dessus d'une case grisée servent à savoir si le dé suivant indique sur quel sommet aller ou s'il faut suivre les arcs.

3) Compléter le tableau ci-contre et établir un classement.

Sommet	1	2	3	4	5	6
Visites						
Classement						

Utilisation de Python

EXERCICE 8 : Nous allons étudier le graphe de l'exercice 1. Il est déjà défini à l'aide des instructions `ajouter_page(1)`... ainsi que `ajouter_lien(1, 2)`...

Pour simplifier, les noms des sommets sont des nombres : 1 pour A, 2 pour B, 3 pour C et 4 pour D.

- 1) Exécuter le script si ce n'est pas déjà fait et taper la commande `afficher_graphe()` dans le terminal. Vérifier que la description correspond bien au graphe de l'exercice 1.

2) Taper `marche_simple(1, 100)` pour simuler une marche aléatoire en 100 étapes en partant du sommet 1. Compléter le tableau.

Sommet	1	2	3	4
Score				
Classement				

3) Relancer plusieurs fois cette commande (vous pouvez utiliser la flèche du haut pour remonter dans l'historique des commandes). Que remarquez vous sur le classement obtenu?

4) Faire des marches aléatoires en partant des autres sommets. Que remarquez-vous sur le classement obtenu?

Sommet	1	2	3	4
Score				
Classement				

- 5) Faire des marches aléatoires en 10000 étapes en partant du sommet de votre choix. Compléter le tableau ci-contre avec un de vos résultats.

Sommet	1	2	3	4
Score				
Classement				

- 6) Taper `afficher_matrice()` pour voir la matrice associée à ce graphe et vérifier que c'est la même que celui de la feuille d'exercices.
- 7) Taper `pagerank_matrice(100)` pour faire le calcul avec la matrice. Comparer le résultat avec ceux des marches aléatoires.

Sommet	1	2	3	4
Score				
Classement				

EXERCICE 9 : Nous considérons maintenant le graphe de l'exercice 2.

- 1) Modifier les lignes `ajouter_lien(..., ...)` pour indiquer les liens entre les différentes pages du graphe. Indiquer les couples de valeurs utilisés comme paramètres :
- (1, 2), (1, 3),
.....
- 2) Afin de simuler plusieurs marches aléatoires, nous allons utiliser la commande `marches_multiples(100, 1000)` pour faire 1000 marches aléatoires de 100 étapes chacunes. Compléter le tableau ci-contre.
- 3) Vérifier la matrice trouvée sur la feuille avec celle obtenue avec `afficher_matrice()`.
- 4) Remplir le tableau avec les résultats obtenus avec `pagerank_matrice(100)`.

Sommet	1	2	3	4
Score				
Classement				

EXERCICE 10 : Nous considérons maintenant le graphe de l'exercice 3, sans rajouter la modification.

- 1) Remplir le tableau avec les résultats obtenus avec `marches_multiples(100, 1000)`.

Sommet	1	2	3	4	5	6
Score						
Classement						

2) Rajouter dans le fichier les liens allant de 2 vers les 6 sommets, y compris lui-même.

- 3) Exécuter le script, refaire `marches_multiples(100, 1000)` avec ce nouveau graphe et remplir le tableau ci-contre.

Sommet	1	2	3	4	5	6
Score						
Classement						

4) Comparer les résultats trouvés dans les deux questions précédentes.

5) Comparer `pagerank_matrice(100)` avec les résultats précédents.

EXERCICE 11 : On considère maintenant le graphe de l'exercice 6.

- 1) Modifier le fichier pour obtenir le graphe souhaité. Indiquer les couples de valeurs utilisés dans les liens :
.....
- 2) Comparer les résultats obtenus avec `marches_multiples_sauts(100, 1000)` et `pagerank_matrice_sauts(1000)`.
- 3) D'après `pagerank_matrice_sauts(1000)`, quelle est la page la plus visitée?