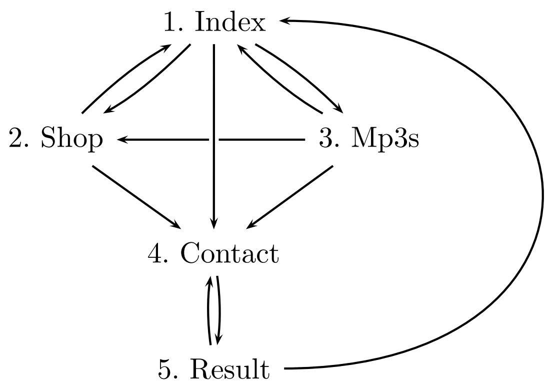
Pertinence d'une page Web
(TS spécialité)

Exercice 1

Un groupe de musique a mis en place un site internet selon le plan ci-dessous, où les flèches représentent des liens hypertexte.



- La page **Index** est la page d'accueil du site. Elle renvoie vers les pages Shop, Mp3s et Contact.
- La page **Shop** offre des produits en vente directe. Elle renvoie vers les pages Index et Contact.
- La page **Mp3s** propose des titres en écoute. Elle renvoie vers les pages Shop, Index et Contact.
- La page **Contact** présente un formulaire permettant d'envoyer un message au webmaster du site. Elle renvoie automatiquement sur la page Result.
- La page **Result** est une page d'accusé d'envoi du message de la page Contact. À partir de cette page, on peut revenir vers la page Contact ou aller à la page Index.

Le concepteur du site souhaite que les moteurs de recherche renvoient en priorité la page d'accueil. Pour cela, il étudie la pertinence de ces 5 pages selon différents modèles.

Partie A – Comptage naïf

Dans ce premier type de comptage, la pertinence de chaque page est égale au nombre de liens qui renvoient vers cette page.

1. Calculer avec cette méthode la pertinence de chaque page.
2. Indiquer l'ordre de pertinence des pages selon ce modèle. Le souhait du concepteur du site est-il respecté ?

Partie B – Comptage pondéré

Dans ce deuxième type de comptage, on considère qu'un surfeur suit une marche aléatoire dans le graphe constitué de ces 5 pages et de leurs liens hypertexte. Le surfeur visitant une page P visite ensuite une des pages qui y sont liées avec une probabilité égale à $\frac{1}{m}$ où m est le nombre de liens issus de la page P. Chaque lien est ainsi pondéré par une probabilité. La pertinence d'une page est la somme des probabilités affectées aux liens pointant vers cette page.

1. Calculer avec cette méthode la pertinence de chaque page.

2. Indiquer l'ordre de pertinence des pages selon ce modèle. Le souhait du concepteur du site est-il respecté ?

Partie C – Comptage récursif

Les comptages naïfs et pondérés ne sont pas satisfaisants pour plusieurs raisons. En particulier, ils sont facilement manipulés en créant des liens artificiels, à partir de pages dont l'unique objectif est d'augmenter la pertinence des pages vers lesquelles elles renvoient.

Pour éviter ce problème, on peut "distribuer" la pertinence de chaque page aux pages qui y sont liées. Pour cela on va considérer à nouveau que le surfeur suit une marche aléatoire sur le graphe, et s'intéresser à la loi de probabilité de la position du surfeur sur le graphe après un temps infini.

La position initiale du surfeur est notée par une matrice à 1 ligne et 5 colonnes U_0 dont tous les termes sont nuls hormis celui correspondant à la page où se trouve le surfeur, égal à 1. À chaque étape de la marche aléatoire sur le graphe correspond une matrice ligne U_p qui constitue la loi de probabilité de la position du surfeur lors de cette étape. On admet que cette suite de matrices a comme limite une matrice ligne W.

En notant $\mu_{p,1}$, $\mu_{p,2}$, $\mu_{p,3}$, $\mu_{p,4}$ et $\mu_{p,5}$ les termes de la matrice U_p, et de manière analogue les termes de la matrice U_{p+1}, on peut établir la relation de récurrence entre ces deux matrices. Le terme $\mu_{p,1}$ est la probabilité qu'à l'étape p, le surfeur soit sur la page 1, à partir de laquelle on peut visiter les pages 2, 3 et 4. Cette probabilité contribuera donc équitablement, au rang suivant, aux probabilités $\mu_{p+1,2}$, $\mu_{p+1,3}$, $\mu_{p+1,4}$. On trouvera donc dans les formules de ces trois probabilités le terme $\frac{1}{3}\mu_{p,1}$. À l'inverse, les pages 2, 3 et 5 renvoient à la page 1, et ces pages émettent respectivement 2, 3 et 2 liens. Par conséquent,

$$\mu_{p+1,1} = \frac{1}{2}\mu_{p,2} + \frac{1}{3}\mu_{p,3} + \frac{1}{2}\mu_{p,5}.$$

1. De manière analogue, donner les formules pour $\mu_{p+1,2}$, $\mu_{p+1,3}$, $\mu_{p+1,4}$ et $\mu_{p+1,5}$.

2. Écrire la matrice carrée d'ordre 5 A telle que pour tout entier naturel p, $U_{p+1} = U_p \times A$.

3. De quelle nature est la suite de matrices $(U_p)_{p\in\mathbf{N}}$? En déduire la formule explicite de la matrice U_p pour tout entier naturel p.

4. À l'aide d'un logiciel de calcul formel, on observe que la suite $(A^p)_{p\in\mathbf{N}}$ des puissances de la matrice A admet comme limite la matrice carrée

$$L = \begin{pmatrix} \frac{9}{40} & \frac{1}{10} & \frac{3}{40} & \frac{3}{10} & \frac{3}{10} \\ \frac{9}{40} & \frac{1}{10} & \frac{3}{40} & \frac{3}{10} & \frac{3}{10} \\ \frac{9}{40} & \frac{1}{10} & \frac{3}{40} & \frac{3}{10} & \frac{3}{10} \\ \frac{9}{40} & \frac{1}{10} & \frac{3}{40} & \frac{3}{10} & \frac{3}{10} \\ \frac{9}{40} & \frac{1}{10} & \frac{3}{40} & \frac{3}{10} & \frac{3}{10} \end{pmatrix}$$

En déduire, quelle que soit la position initiale U_0, la matrice limite de la suite (U_p).

5. Indiquer l'ordre de pertinence des pages selon ce modèle. Le souhait du concepteur du site est-il respecté ?

Partie D – Comptage récursif avec saut direct

La méthode de calcul de la pertinence d'une page Web développée par Google prend en considération la possibilité que le surfeur atteigne directement une page, par exemple en tapant directement une adresse, ou bien en utilisant un marque-page. On évite ainsi les "trous noirs" que pourraient constituer des pages n'émettant aucun lien.

Le *coefficient d'échappement* évalue la probabilité que le surfeur effectue ainsi un saut direct. On prend souvent comme valeur $c = 0,15$. Cette probabilité est répartie équitablement entre toutes les pages du graphe, et son complémentaire est appliqué à la marche aléatoire. Avec les notations de la partie précédente, on a donc

$$U_{p+1} = \frac{0,15}{5}C + 0,85U_pA = U_p\left(\frac{0,15}{5}J + 0,85A\right)$$

où C est la matrice 1×5 dont tous les termes sont égaux à 1 et J est la matrice carrée d'ordre 5 dont tous les termes sont égaux à 1.

1. Expliquer la formule $U_{p+1} = \frac{0,15}{5}C + 0,85U_pA$.

2. En vérifiant que $U_p \times \frac{0,15}{5}J = \frac{0,15}{5}C$, justifier la formule $U_{p+1} = U_p\left(\frac{0,15}{5}J + 0,85A\right)$

3. Calculer la matrice $B = \frac{0,15}{5}J + 0,85A$.

4. Préciser la nature et les éléments caractéristiques de la suite de matrices $(U_p)_{p\in\mathbf{N}}$.

5. À l'aide d'un logiciel de calcul formel, on observe que la suite $(B^p)_{p\in\mathbf{N}}$ s'approche de la matrice carrée approximativement égale à

$$L' = \begin{pmatrix} 0,2242 & 0,1200 & 0,0935 & 0,2877 & 0,2746 \\ 0,2242 & 0,1200 & 0,0935 & 0,2877 & 0,2746 \\ 0,2242 & 0,1200 & 0,0935 & 0,2877 & 0,2746 \\ 0,2242 & 0,1200 & 0,0935 & 0,2877 & 0,2746 \\ 0,2242 & 0,1200 & 0,0935 & 0,2877 & 0,2746 \end{pmatrix}$$

En déduire, quelle que soit la position initiale U_0, la matrice limite de la suite (U_p).

6. Indiquer l'ordre de pertinence des pages selon ce modèle. Le souhait du concepteur du site est-il respecté ?

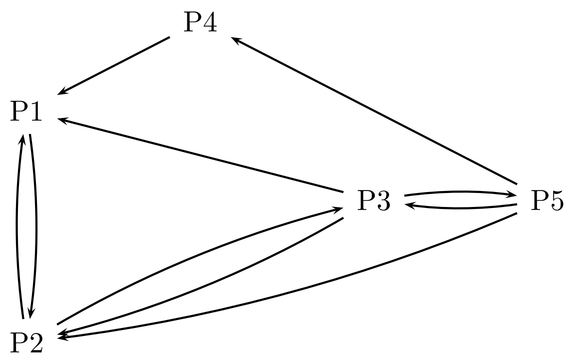
Partie E – Une modification du graphe

1. La suppression d'un unique lien permet de modifier l'ordre de pertinence des pages pour le comptage naïf et de se rapprocher du souhait du webmaster . Quel est ce lien ?

2. Étudier la pertinence des pages de ce nouveau graphe selon les 3 autres modèles. On utilisera un logiciel de calcul formel ou la calculatrice pour approcher les limites des suites de matrices. Le souhait du concepteur du site est-il respecté ?

Exercice 2

On s'intéresse au graphe orienté ci-dessous, représentant un ensemble de 5 pages internet et de leurs liens hypertexte.



Partie A – Comptage naïf et comptage pondéré

1. Classer les pages par ordre de pertinence selon le comptage naïf.
2. Classer les pages par ordre de pertinence selon le comptage pondéré.

Partie B – Comptage récursif

On considère un surfeur suivant une marche aléatoire sur ce graphe. Pour tout entier naturel p, on note U_p la matrice ligne donnant la loi de probabilité de la position du surfeur à l'étape p.

1. Déterminer la matrice A telle que $U_{p+1} = U_p \times A$.
2. Préciser la nature et les éléments caractéristiques de la suite de matrices lignes $(U_p)_{p\in\mathbf{N}}$.
3. À l'aide d'un logiciel de calcul formel ou d'une calculatrice, déterminer la limite de la suite des puissances de A.
4. En déduire, en justifiant, le classement des pages par ordre de pertinence selon le comptage récursif.

Partie C – Comptage récursif avec saut direct

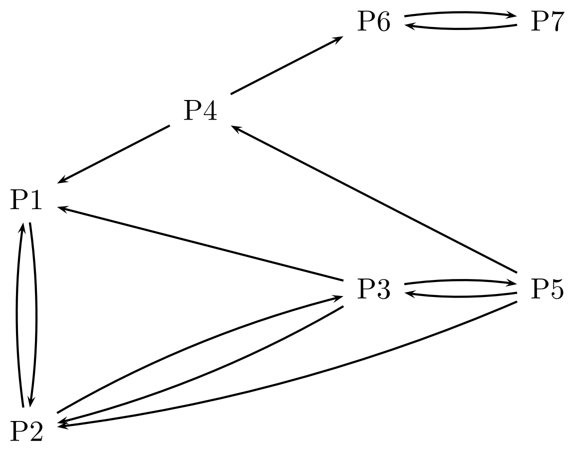
Dans cette partie, on considère que le surfeur utilise un saut direct vers une des pages du graphe avec une probabilité égale à $0,15$. On conserve les mêmes notations que dans la partie précédente.

1. Déterminer la matrice B telle que $U_{p+1} = U_p \times B$.
2. Préciser la nature et les éléments caractéristiques de la suite de matrices lignes $(U_p)_{p\in\mathbf{N}}$.
3. À l'aide d'un logiciel de calcul formel ou d'une calculatrice, déterminer la limite de la suite des puissances de B.
4. En déduire, en justifiant, le classement des pages par ordre de pertinence selon le comptage récursif avec saut direct.
5. Comparer les classements obtenus avec les différents modes de comptage.

Partie D – Une modification malheureuse

On ajoute maintenant au graphe précédent deux pages, P6 et P7, ainsi que les liens P4→P6, P6→P7 et P7→P6.

1. Étudier la pertinence des pages de ce nouveau graphe avec le comptage récursif. Qu'observe-t-on ?
2. Étudier la pertinence des pages de ce nouveau graphe avec le comptage récursif avec saut direct. Observe-t-on le même phénomène qu'avec le mode de comptage précédent ?


P6
P7
P4
P1
P3
P5
P2