



SCIENCES NUMÉRIQUES ET TECHNOLOGIE

Thème : le Web

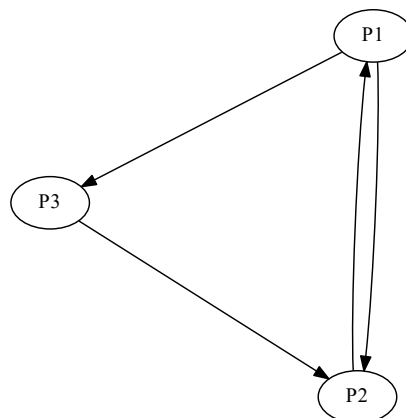
Activité pour estimer la popularité¹ d'une page en utilisant
le modèle du « surfeur aléatoire »

Fiche élève



Situation n°1

Dans l'activité qui suit, on propose un modèle extrêmement réduit du Web avec ... trois pages. Les liens hypertextes entre ces pages sont représentés à l'aide d'un schéma (graphe orienté) ; voici la première situation étudiée :



1. en exagérant un peu, on qualifiera cette popularité de « PageRank » de la page

? Question 1

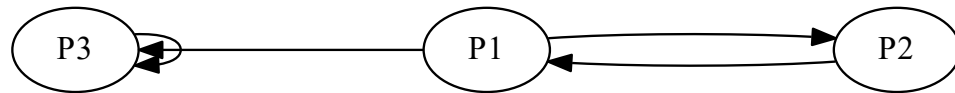
À l'aide d'un dé à six faces si besoin :

1. positionnez-vous sur l'une des trois pages au hasard ;
2. suivez le lien hypertexte ; au besoin (si plusieurs choix sont possibles), utilisez le dé pour décider pour vous !
3. comptabilisez le nombre de passages sur chaque page après un grand nombre d'étapes et donnez la fréquence de passage de chaque page : cette valeur mesure en quelque sorte la **popularité** de la page ;
4. une fois ce travail fait, modélisez la situation par un algorithme que vous allez programmer et exécuter.



Situation n°2

On vous propose une deuxième situation modélisée par le graphe suivant :



? Question 2

Si on procède de la même manière que précédemment, que va-t-il se passer ?
Vous pouvez adapter le programme précédent pour appuyer votre conjecture.

Dans cette situation, la page 3 est qualifiée de « puits » ; pour pallier ce problème, on ajoute au modèle précédent une autre composante : **lorsque l'on se retrouve sur la page n°3, on va pointer aléatoirement et de manière équiprobable vers une des trois pages.**

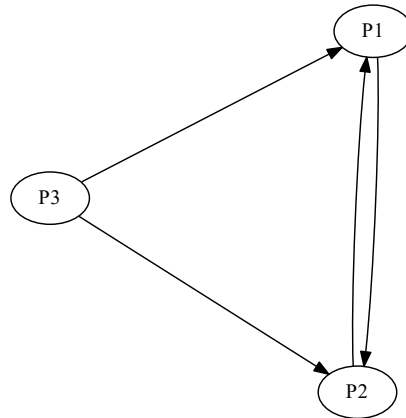
? Question 3

Adaptez votre programme avec cette nouvelle indication et donnez une estimation de la popularité de chaque page.



Situation n°3

On vous propose une troisième situation modélisée par le graphe suivant :



? Question 2

Si on procède de la même manière que précédemment, que va-t-il se passer ?

Vous pouvez adapter le programme précédent et vérifier.

Dans ce cas, à moins d'être tombé sur la page n°3 au départ, on n'y reviendra jamais ! C'est pourquoi on ajoute au modèle de départ : **on introduit une probabilité - en général de l'ordre de 0,15 - d'aller sur une page au hasard, quel que soit le lien entre les pages**

? Question 4

Adaptez votre programme avec cette nouvelle indication et donnez une estimation de la popularité de chaque page.

💡 Pour aller plus loin

On peut modifier la fonction `rep()` et mettre la fonction qui permet de passer d'une page à la suivante (les fonctions `pagesuivante`) comme **paramètre** ; ainsi, il suffira de changer de fonction pour s'adapter à la situation ; cela donne :

```
ÉDITEUR : PAGERANK
LIGNE DU SCRIPT 0095
def rep2(f, n):
    c1,c2,c3,pa = 0,0,0,page1()
    for i in range(n):
        if pa == 1:
            c1+=1
        elif pa == 2:
            c2+=1
        elif pa == 3:
            c3+=1
        pa = f(pa)
    return c1/n , c2/n , c3/n
Fns... a A # Outils Exéc Script
```

`rep` prend une fonction en paramètre

```
PYTHON SHELL
>>> rep2(pagesuiv3,100)
(0.5, 0.49, 0.01)
>>> rep2(pagesuiv32,100)
(0.46999999999999999, 0.48, 0.05)
>>> rep2(pagesuiv3,1000)
(0.5, 0.5, 0.0)
>>> rep2(pagesuiv32,1000)
(0.474, 0.46999999999999999, 0.056000000000000001)
>>> |
Fns... a A # Outils Éditer Script
```

résultats