Seconde – SNT – Le Web
Séance n°5 : Algorithme PageRank

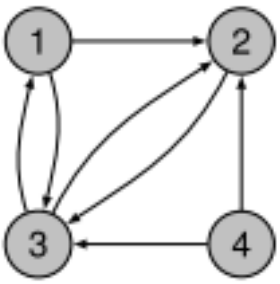
https://fr.wikipedia.org/wiki/PageRank

Le PageRank est l'algorithme d'analyse des liens concourant au système de classement des pages Web utilisé par le moteur de recherche Google. Il mesure quantitativement la popularité d'une page web.
Le PageRank n'est qu'un indicateur parmi d'autres dans l'algorithme qui permet de classer les pages du Web dans les résultats de recherche de Google.
Ce système a été inventé par Larry Page, cofondateur de Google.
Le principe est d'attribuer à chaque page un score proportionnel au nombre de fois que passerait par cette page un utilisateur parcourant le graphe du Web en cliquant aléatoirement, sur un des liens apparaissant sur chaque page.

Le « surfeur aléatoire »

Après avoir fait la liste des sites traitant du sujet qui l'intéresse, le surfeur aléatoire en choisit un au hasard un. Puis, il choisit toujours au hasard de suivre l'un des liens hypertextes du site sur lequel il se trouve vers les autres sites qu'il a listés. Il répète cette opération sans s'arrêter en comptant pour chacun des sites combien de fois il l'a visité. Chaque site est alors affiché avec son score c'est-à-dire son nombre de visites.

Exercice 1.



▶1. Prenons l'exemple de quatre sites qui traitent du sujet recherché : la flèche allant de 1 vers 2 signifie que la page 1 a un lien hypertexte vers la page 2. L'absence de flèche de 2 vers 4 signifie qu'il n'y a pas de lien hypertexte de la page 2 vers la page 4.

1. Choisissez un site parmi les 4.
2. Simulez un déplacement aléatoire de notre surfeur.
3. Proposez un classement de ces quatre pages.
4. Comparez votre classement avec les autres membres de votre groupe.

▶2. Taper l'algorithme ci-dessous, que permet-il ?

```
from random import *
def suivant1(site):
    if site==1:
        site=choice([2,3])
    elif site==2:
        site=3
    elif site==3:
        site=choice([1,2])
    else:
        site=choice([2,3])
    return site
site=int(input("Quelle est le premier site ? "))
site=suivant1(site)
print(site)
```

►3. Modifier la fin de l'algorithme.

```
site=int(input("Quelle est le premier site ? "))
L=[]
L.append(site)
for i in range(10-1):
    site=suivant1(site)
    L.append(site)
for i in range(1,5):
    score=L.count(i)
    print(i, score)
```

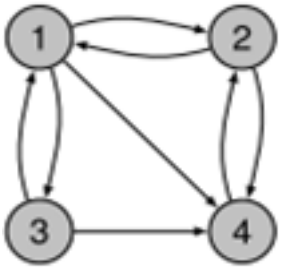
Ce qu'il faut comprendre :

- `L=[]` permet de créer une liste qui contiendra les sites visités au fur et à mesure,
- `L.append(site)` permet de rajouter le numéro du site visité,
- `L.count(i)` permet de compter le nombre de passage sur chaque site.

►4. Modifiez le nombre total de sites visités, exécutez, éventuellement plusieurs fois, le programme en choisissant la première page puis compléter les scores ci-dessous :

		Score			
		Site 1	**Site 2**	**Site 3**	**Site 4**
Nombre total de sites visités	**100**				
	1000				
	10000				
PageRank estimé en %					

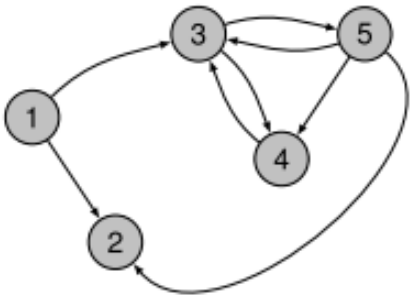
Exercice 2.



Modifier le programme précédent pour estimer le « PageRank » des sites représentés par le graphe ci-contre.

		Score			
		Site 1	**Site 2**	**Site 3**	**Site 4**
Nombre total de sites visités	**100**				
	1000				
	10000				
PageRank estimé en %					

Exercice 3.

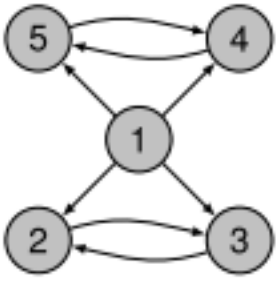


►1. Considérer le graphe ci-contre. A votre avis, que risque t'il d'arriver à notre surfeur aléatoire ?
Comment ce problème se traduirait au niveau de notre fonction ?
►2. Pour pallier à ce genre de problème, l'algorithme a évolué et le surfeur aléatoire a laissé place au « surfeur raisonnable »
Parmi les améliorations de cet algorithme, il y a la possibilité de se télétransporter. On pourra supposer que si on est bloqué, alors on pourra au coup suivant aller n'importe où.
Modifier le programme précédent pour essayer d'estimer le « PageRank » des sites représentés par le graphe ci-contre avec notre version simplifiée du surfeur raisonnable.

		Score				
		Site 1	**Site 2**	**Site 3**	**Site 4**	**Site 5**
Nombre total de sites visités	**100**					
	1000					
	10000					
PageRank estimé en %						

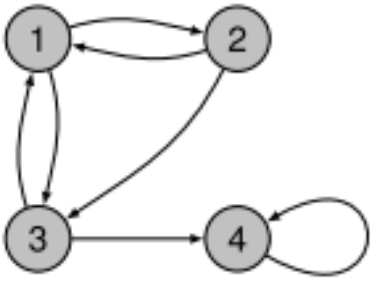
Exercice n°4 :

Essayez d'estimer le « PageRank » des sites représentés par le graphe ci-dessous avec le surfeur aléatoire.
Quel problème est soulevé ?
Imaginez une solution pour pallier ces problèmes ?



Exercice n°5 :

Essayez d'estimer le « PageRank » des sites représentés par le graphe ci-dessous.
Quel problème est soulevé ?
Imaginez une solution pour pallier ces problèmes ?



Pour approfondir : https://www.netinshape.fr/pagerank/