Rapport 3I005
Chaînes de Markov : PageRank

MENG Fanshuo :3403051
XIANG Yangfan :3300401

I. Préliminaire

Le code est en Python3, comportant 6 fichiers.
datastructure : les classes de structures de données avec une fonction pour le méthode de la limite de puissance de la matrice de transition.
internaute : la classe internaute pour le méthode “processus particulier“.
simulationpi : la classe pour le méthode de $\pi_n = \pi_o \times P^n$.
tools :les fonctions pour la génération du nanoWeb 1, 2, 3 et le générateur ergodique
main1 : comparaison des différentes méthodes sur les nanoWeb 1, 2 et 3
main2 : comparaison des différentes méthodes sur les nanoWeb générer par le générateur ergodique de tailles diffèrent.

1) $\pi_{t+1} = \pi_t \times P$

Soit π_i la distribution de probabilité de se trouver dans chaque nœud a l’instant i et P la matrice de transition.
Le produit entre $\pi_i \times P = \pi_{i+1}$ nous donne la distribution de probabilité a l’instant suivant, i.e chaque valeur de π_{i+1} est la somme des probabilités des chemins possible pour rejoindre un nœud sachant la distribution π_i dans l’état précédant.

2) $P^2, P^3, P^n, lim_{n\to+\infty} P^n$

Soit P la matrice de transition, P^2, P^3, P^n représente la transition en 2, 3 et n étapes, i.e ${P^m}_{i,j} = P(X_m = j|X_0 = i)$ la probabilité de passer de la position i à la position j en m étapes. Si on a convergence (existence de régime permanent) tous les lignes de $lim_{n\to+\infty} P^n$ sont égaux et correspond la probabilité de distribution stationnaire.

II. Analyse des 3 nanoWebs

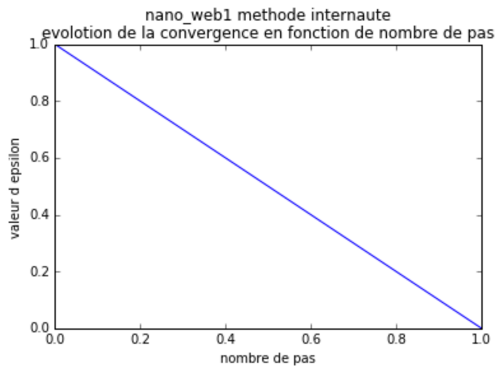
1) nanoWeb1

a) Caractéristiques

Le nanoWeb1 contient un nœud 9 qui est absorbant, donc il n’est pas irréductible et n’est pas apériodique. On en déduit qu’il n’existe pas de probabilité stationnaire pour les nœuds du nonoWeb1 (pas de convergence)

b) Méthode de processus particulier

Cas1 : un point absorbant

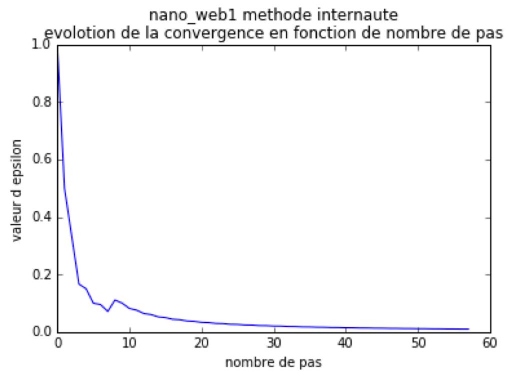


Avec comme point de départ le nœud 3, il suffit d'une seule étape pour atteindre une convergence car le nœud 3 ne peut se diriger que vers le nœud 9 qui est un nœud absorbant.

Nœud	0	1	2	3	4	5	6	7	8	9
Proba	0	0	0	0	0	0	0	0	0	1

Tableau de la distribution de probabilité avec comme point de départ nœud 3

Cas2 : un point avec entre et sortie

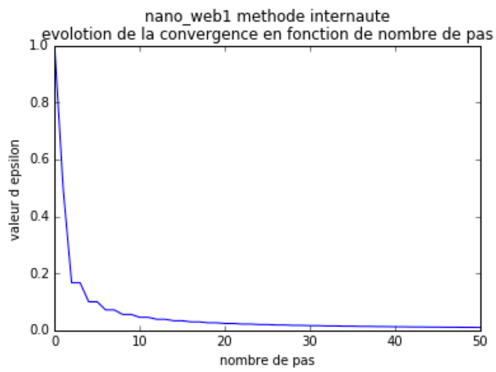


Avec le point de départ le nœud 0, on atteint une convergence au bout de 56 itérations.

Nœud	0	1	2	3	4	5	6	7	8	9
Proba	0	0	0	0	0.017	0.05	0.06	0.43	0.43	1

Tableau de la distribution de probabilité avec comme point de départ nœud 0

Cas3 : une boucle



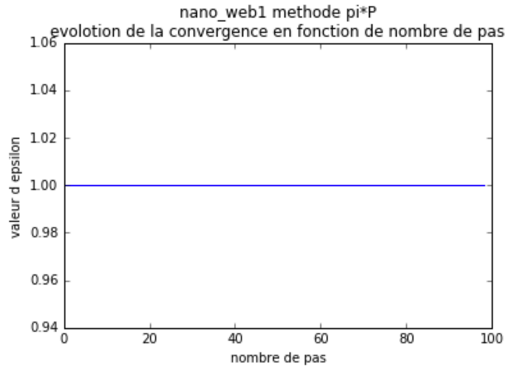
Avec comme point de départ le nœud 7, on atteint une convergence au bout de 50 itérations.

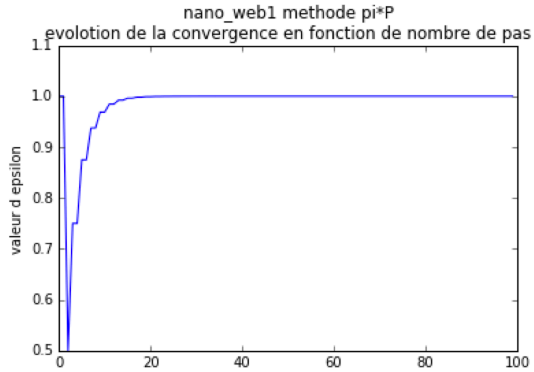
Nœud	0	1	2	3	4	5	6	7	8	9
Proba	0	0	0	0	0	0	0	0.49	0.51	0

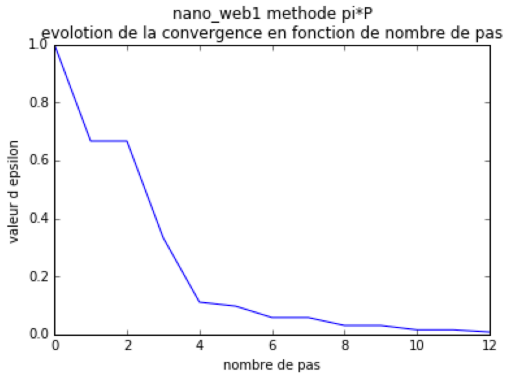
Tableau de la distribution de probabilité avec comme point de départ nœud 7

Même si on n'est pas dans un cas convergent, la méthode d'internaute fini toujours par se converger. Cette méthode n'est pas fiable dans le cas de non convergence.

c) Méthode de $\pi_n = \pi_o \times P^n$

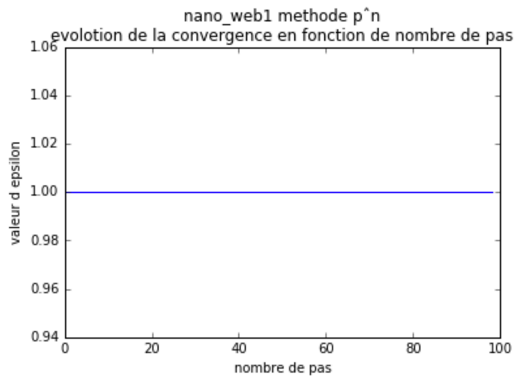






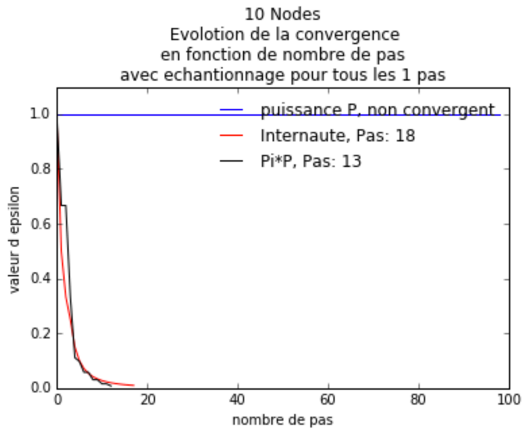
On observe que dans des cas, on a la convergence et dans des autres cas non.
La convergence ou non est dû à la position de départ, par exemple si on part de 0 on atteint une convergence (figure du bas)
Si on part de 5 (figure de droit) on observe une baise de seuil rapide et puis un redressement et se stagne au 1, cela est sans doute du du passage de la boucle formée par 5 et 6 au boucle formée par 7 et 8.
Si on part de 8 (figure gauche) on se trouve dans la boucle formée par 7 et 8, le seuil reste toujours à 1 puis qu'on passe de l'un à l'autre de chaque itération, pas de convergence.

d) Méthode de $lim_{n \to +\infty} P^n$



On a une convergence si le nanoWeb1 est ergodique, comme il n'est pas, les produits successifs entre P ne convergeront jamais donc ε reste constant à 1. Le calcule s'arrête au bout d'un moment (100 itérations) pour éviter la boucle infinie.

e) Conclustion



Seul la méthode de $lim_{n \to +\infty} P^n$ est fiable dans le cas de non ergodique.

2) nanoWeb2
a) Caractéristique

Le nanoWeb2 est irréductible car de l'import quel nœud on peut accéder a n'importe quel nœud. En plus on observe une cycle de période 3 entre les nœuds 2 3 4 et une cycle de période 4 entre 2 1 0 9 et pgcd(3, 4) = 1 donc le nanoWeb2 est apériodique.
Le nanoWeb2 est donc ergodique, on a dont la convergence pour les 3 méthode.

b) Méthode de processus particulier

Nœud	0	1	2	3	4	5	6	7	8	9
Proba	0.037	0.151	0.227	0.189	0.126	0.126	0.126	0.012	0.012	0.037

Tableau de la distribution de probabilité avec comme point de départ nœud 3

c) Méthode de $\pi_n = \pi_o \times P^n$

Nœud	0	1	2	3	4	5	6	7	8	9
Proba	0.055	0.113	0.217	0.197	0.193	0.084	0.028	0.026	0.028	0.054

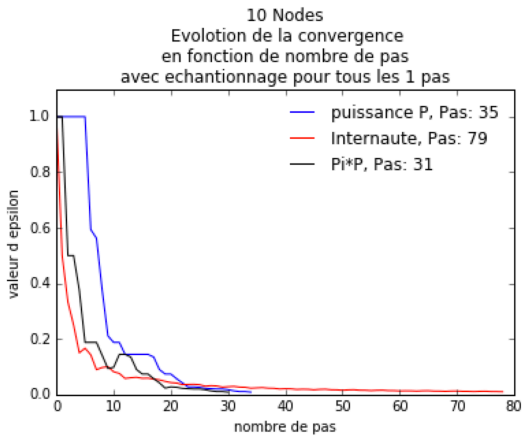
Tableau de la distribution de probabilité avec comme point de départ nœud 3

d) Méthode de $lim_{n \to +\infty} P^n$

Nœud	0	1	2	3	4	5	6	7	8	9
Proba	0.056	0.109	0.223	0.195	0.192	0.084	0.027	0.028	0.027	0.054

Tableau de la distribution de probabilité

e) Conclusion

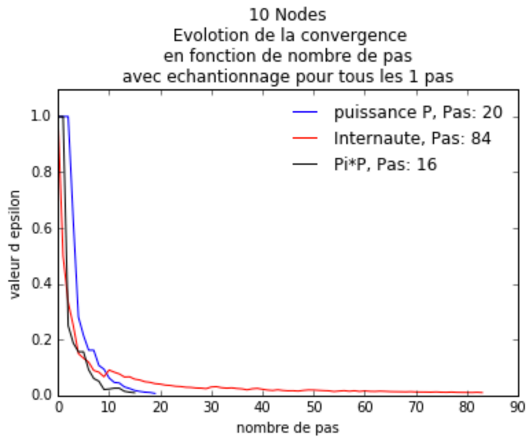


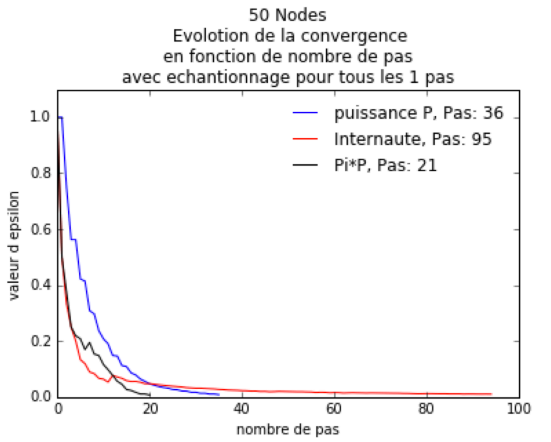
En terme d'efficacité, la méthode d'internaute est la plus mauvaise avec plus de 79 itérations pour atteindre le seuil de convergence, la méthode de $\pi_n = \pi_o \times P^n$ a une léger avance sur $lim_{n\to+\infty} P^n$. Nivaux résultat, la méthode d'internaute est encore une fois la plus mauvais avec des points qu'ont une grande variation de probabilité, les deux autres s'égalise.
Sachant que la méthode $lim_{n\to+\infty} P^n$ détecte en cas de non ergodique, il est préférable de l'utiliser.

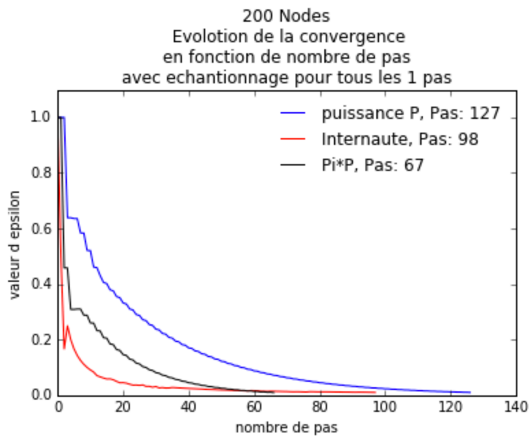
3) nanoWeb3

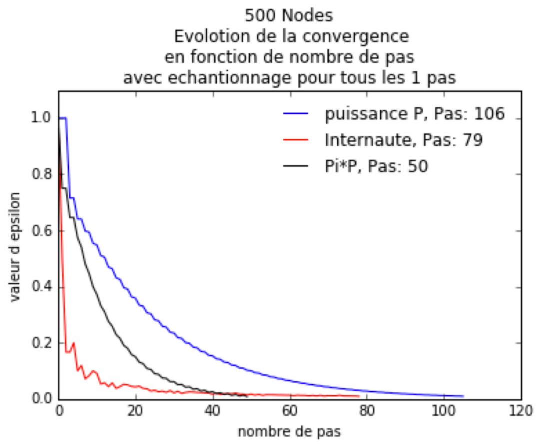
L'analyse pour le nanoWeb3 est la même pour le nanoWeb2 car il est non ergodique. Cf. 1).

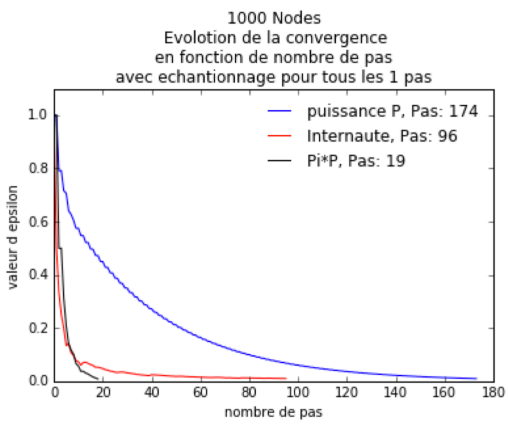
III. Générateur ergodique et test sur des grosses structures

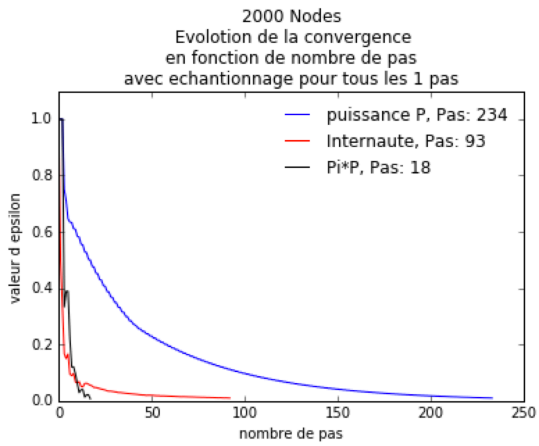












Dans le cas ergodique, en testant pour des tailles variables, la méthode $\pi_n = \pi_o \times P^n$ est le plus efficace. Surtout lorsque la taille dépasse les milles, le nombre d'itération nécessaire pour cette méthode reste constant alors que l'un est 10 fois plus l'autre 5 fois plus.

Conclusion : si on est dans le cas ergodique il faut mieux utiliser la méthode de $\pi_n = \pi_o \times P^n$ pour déterminer la distribution stationnaire sinon si on n'a pas cette certitude, la méthode $lim_{n\to+\infty} P^n$ est celui qu'il faut appliquer. La méthode d'internaute est donc le pire choix possible, il faut absolument l'éviter.