LINFO1114 - Projet - Groupe 19

Henne Alexis
Meessen Vincent
Arnaud Alexis

Décembre 2025

1 Introduction

Ce projet vise à implémenter et tester trois méthodes de calcul du score PageRank des noeuds d'un graphe orienté.

Les trois méthodes sont :

- La Power method
- La résolution via un système d'équations linéaires
- La simulation d'une marche aléatoire sur le graphe

Ces méthodes seront détaillées dans le rappel théorique qui suit.

2 Rappels théoriques

Le PageRank est une mesure d'importance des noeuds d'un graphe orienté, interprétée comme la distribution stationnaire d'un marcheur aléatoire se déplaçant de noeud en noeud. Pour un graphe à n noeuds, on définit :

- P : matrice de transition du graphe (chaque ligne distribue uniformément la probabilité vers les voisins sortants).
- $\alpha \in [0, 1]$: paramètre de téléportation.
- $v \in R^n$: vecteur de personnalisation tel que $\sum_i v_i = 1$.
- $r \in R^n$: vecteur des scores PageRank recherchés.

L'équation générale s'écrit :

$$r = \alpha P^\top + (1 - \alpha)v$$

Cette équation exprime bien que le score du noeud dépend à la fois des contributions des noeuds qui pointent vers lui (premier terme) ainsi que de la probabilité de téléportation (deuxième terme).

2.1 Calcul du PageRank par système d'équations linéaires

L'équation générale du PageRank peut s'écrire sous la forme d'un problème linéaire:

$$(I - \alpha P^\top)r = (1 - \alpha)v$$

où I est la matrice identité.
Il faut donc résoudre :

$$r = (I - \alpha)P^\top)^{-1}(1 - \alpha)v$$

2.2 Calcul du PageRank par la Power Method

La Power Method exploite le fait que le PageRank correspond au vecteur propre de la matrice de transition.

$$G^\top = \alpha\, P^\top + (1 - \alpha)\, e\, v^\top$$

On cherche donc un vecteur r tel que :

$$r = G^\top r$$

La Power Method consiste donc à itérer :

$$r^{(k+1)} = r^{(k)} G^\top$$

jusqu'à convergence, c'est-à-dire tant que la norme

$$\|pageRankLinear - r^{(k+1)}\| < \varepsilon$$

pour un ε choisi.

Cette méthode converge toujours car la matrice G est stochastique, irréductible et apériodique grâce à la téléportation.

2.3 Impact du paramètre α

Le paramètre α contrôle le taux de téléportation
Plus α est grand, plus le PageRank reflète la structure du graphe mais converge lentement.
Plus α est petit, plus la téléportation domine, ce qui rend les scores plus uniformes et la convergence plus rapide.
La valeur standard est $\alpha = 0.85$.

2.4 Impact du vecteur de personnalisation v

Le vecteur v définit vers quels noeuds le marcheur est susceptible de se téléporter.

Un vecteur uniforme donne le PageRank classique, tandis qu'un vecteur biaisé oriente les scores vers les noeuds privilégiés, permettant une personnalisation des scores.

3 Présentation numérique du système d'équations linéaires avec exemple

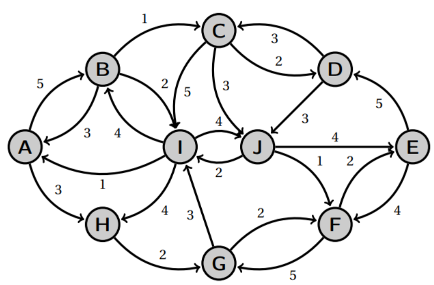

Figure 1: Graphe exemple

3.1 Matrice de transition P

Les poids sortants sont normalisés pour chaque noeud. La matrice $P^\top$ (obtenue par transposition de la matrice de transition définie par lignes) s'écrit :

$$P^\top = \begin{pmatrix} 0 & \frac{3}{6} & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{13} & 0 \\ \frac{5}{8} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{4}{13} & 0 \\ 0 & \frac{1}{6} & 0 & \frac{3}{6} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{2}{10} & 0 & \frac{5}{9} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{2}{7} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{4}{9} & 0 & \frac{2}{5} & 0 & 0 & \frac{1}{3} \\ 0 & 0 & 0 & 0 & 0 & \frac{5}{7} & 0 & 1 & 0 & 0 \\ \frac{3}{8} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{4}{13} & 0 \\ 0 & \frac{2}{6} & \frac{5}{10} & 0 & 0 & 0 & \frac{3}{5} & 0 & 0 & \frac{2}{3} \\ 0 & 0 & \frac{3}{10} & \frac{3}{6} & 0 & 0 & 0 & 0 & \frac{4}{13} & 0 \end{pmatrix}.$$

3.2 Vecteur de personnalisation

Le vecteur de personnalisation est :

$$v = \begin{pmatrix} 0.0953 \\ 0.1858 \\ 0.1068 \\ 0.0452 \\ 0.0089 \\ 0.1469 \\ 0.0951 \\ 0.1138 \\ 0.0616 \\ 0.1406 \end{pmatrix}.$$

3.3 Système linéaire final

Le système d'équations linéaires permettant de calculer le PageRank est donc :

$$(I - \alpha P^\top) \begin{pmatrix} r_A \\ r_B \\ r_C \\ r_D \\ r_E \\ r_F \\ r_G \\ r_H \\ r_I \\ r_J \end{pmatrix} = (1 - \alpha) \begin{pmatrix} 0.0953 \\ 0.1858 \\ 0.1068 \\ 0.0452 \\ 0.0089 \\ 0.1469 \\ 0.0951 \\ 0.1138 \\ 0.0616 \\ 0.1406 \end{pmatrix}$$

4 Evolution de l'erreur dûe à la marche aléatoire

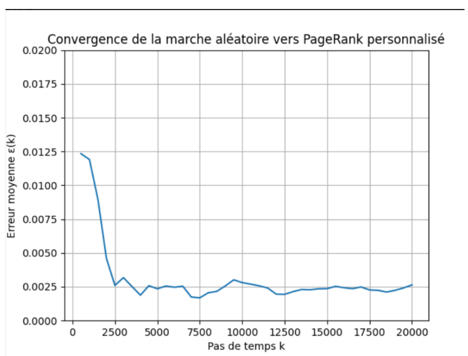

Figure 2: Convergence de la marche aléatoire

Le graphique montre la convergence progressive de l'estimation du PageRank par marche aléatoire vers le score PageRank exact.

L'erreur moyenne est calculée par :

```
mean_error : np.abs(page_rank_linear - pageRankRandomWalk(visits)).mean()
```

Cette formule calcule l'erreur absolue moyenne entre le PageRank exact et l'estimation par marche aléatoire où $\varepsilon(k) = 0$ représente une convergence parfaite , et une valeur très élevée signifie une grande divergence.

Ce que montre le graphique :

- Au début $(k = 500)$: $\varepsilon(k) \approx 0.02$
  L'erreur moyenne est déjà très faible, montrant la convergence rapide vers la valeur PageRank correcte.

- Evolution de $k = 500$ à $k = 2000$: $\varepsilon(k)$ converge vers 0.

  L'erreur diminue jusqu'à ≈ 0.002, soit un gain de facteur 10 en précision par rapport à l'itération 500.
  Le bruit (fluctuations) est normal et attendu pour une marche aléatoire, de par la nature aléatoire de l'algorithme.

5 Méthodes utilisées

5.1 Fonctionnement commun

Les deux implémentations de l'algorithme PageRank (Power Method et Marche Aléatoire) commencent par calculer la matrice de probabilité de transition P. Cette matrice est obtenue en normalisant chaque colonne de la matrice d'adjacence A afin que leur somme soit égale à 1. Pour les nœuds pendants (c'est-à-dire sans liens sortants), une distribution uniforme est appliquée.
Ensuite, les deux méthodes calculent la matrice Google transposée $G^\top$ à partir de la matrice de transition transposée $P^\top$, du vecteur identité e, de la transposée du vecteur de personnalisation $v^\top$ et du coefficient d'amortissement α, selon la formule suivante :

$$G^\top = \alpha\, P^\top + (1 - \alpha)\, e\, v^\top$$

Les étapes suivantes diffèrent selon l'algorithme utilisé.

5.2 Fonctionnement de l'implémentation PageRank avec la Power Method

Le vecteur π, contenant les futurs scores PageRank de chaque nœud, est initialisé proportionnellement au degré entrant de chaque nœud, puis normalisé afin que la somme de ses composantes soit égale à 1.
Le calcul du PageRank final s'effectue ensuite par itérations successives. À chaque itération, le vecteur π est multiplié par la matrice Google :

$$\pi^{(k+1)} = \pi^{(k)} G^\top$$

Cette opération correspond à une étape de propagation des scores PageRank à travers le graphe.
La boucle itérative se termine lorsque l'une des conditions suivantes est vérifiée :

- le nombre maximal d'itérations est atteint (150 itérations) ;
- la convergence est atteinte, c'est-à-dire lorsque la variation entre deux itérations successives devient inférieure au seuil de tolérance fixé à 10^{-9}.

5.3 Fonctionnement de l'implémentation PageRank avec la Marche Aléatoire

Un vecteur `visites` est initialisé avec autant de composantes que de nœuds dans le graphe, toutes initialement égales à zéro. Ce vecteur permet de comptabiliser le nombre de passages sur chaque nœud.
Un nœud initial est ensuite sélectionné pour démarrer la marche aléatoire (par défaut, le nœud 0).
Une boucle itérative est exécutée pour un nombre d'étapes fixé à l'avance (par défaut 20 000 itérations). À chaque itération :

1. un nœud suivant est sélectionné aléatoirement selon les probabilités de transition contenues dans la ligne correspondante de la matrice Google ;
2. le compteur associé au nœud visité dans le vecteur `visites` est incrémenté de 1.

À l'issue de la marche aléatoire, le score PageRank est obtenu en normalisant le vecteur `visites` par la somme totale de ses composantes :

$$\text{PageRank} = \frac{\texttt{visites}}{\sum \texttt{visites}}$$

Cette normalisation transforme les comptages de visites en une distribution de probabilité, où chaque valeur représente la proportion du temps passé sur chaque nœud au cours de la marche aléatoire.

6 Scores obtenus

Voici les scores PageRank obtenus après application de nos méthodes :

$$\text{pageRankLinear} = \begin{pmatrix} 0.06639078 \\ 0.10120283 \\ 0.05378640 \\ 0.06205773 \\ 0.08741389 \\ 0.12046142 \\ 0.15810790 \\ 0.07906491 \\ 0.16350395 \\ 0.10801018 \end{pmatrix} \quad \text{pageRankPower} = \begin{pmatrix} 0.06639078 \\ 0.10120283 \\ 0.05378640 \\ 0.06205773 \\ 0.08741389 \\ 0.12046142 \\ 0.15810790 \\ 0.07906491 \\ 0.16350395 \\ 0.10801018 \end{pmatrix}$$

$$\text{pageRankRandomWalk} = \begin{pmatrix} 0.06564672 \\ 0.09719514 \\ 0.05014749 \\ 0.06024699 \\ 0.08494575 \\ 0.12254387 \\ 0.16474176 \\ 0.08279586 \\ 0.16424179 \\ 0.10749463 \end{pmatrix}$$

7 Importance de chacun des noeuds

Après application de nos méthodes, voici un graphe insistant sur l'importance que prennentchaque noeuds de notre graphique.

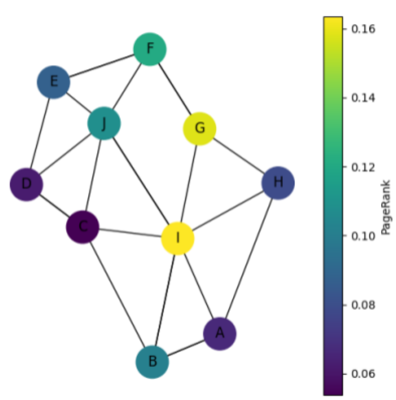

Figure 3: Importance des noeuds