

5. Matrice de Google appliquée à la gestion des affaires

Difficulté : plus difficile

L'organisation d'une vraie entreprise de consulting a été modélisée à l'aide d'un graphe orienté constitué de noeuds représentant les entités de l'entreprise, reliés par des liens représentant les relations de cause à effet. Ce graphe peut être vu comme une chaîne de Markov. La matrice de Google a été appliquée au graphe organisationnel de l'entreprise pour étudier l'importance de ses différentes entités. Les résultats de cette études sont parus dans l'article suivant :

"Google matrix of business process management" par M.W. Abel et D.L. Shepelyansky (fichier **article business management.pdf** dans le répertoire **students_etudiants**, les parties barrées ne sont pas indispensables pour ce projet sauf si vous le terminez en avance).

Les données sont accessibles à l'adresse suivante :

<http://www.quantware.ups-tlse.fr/QWLIB/cheirankbusiness/>

Elles ont été recopiées dans les fichiers **business_noeuds.txt** et **business_liens.txt** dans le répertoire **students_etudiants**.

Vérifiez de manière aussi complète que possible les résultats présentés dans l'article. Testez ensuite vos propres idées pour caractériser cette matrice de Google et cette entreprise. Vous pouvez, par exemple, étudier la convergence du vecteur population vers la population à l'équilibre. Donnez vos conclusions sur ce que nous apprend cette étude sur l'entreprise de consulting.

6. Simulation du moteur de recherche de Google

Difficulté : plus difficile

Ecrivez votre propre moteur de recherche Google. Il devra trouver, dans un ensemble de pages fictives que vous aurez créée, les pages contenant les mots demandés par l'utilisateur puis présenter le résultat en classant les pages par ordre décroissant de PageRank.

- Commencez par créer des pages fictives qui peuvent, par exemple, être des fichiers .txt contenant du texte et des pointeurs vers d'autres pages fictives. Ces pointeurs peuvent être simplement le nom du fichier .txt correspondant à la page pointée, précédés du mot "pointeursvers:"
- Attribuer un PageRank à chacune des pages grâce à la technique de la matrice Google
- Ecrivez un moteur de recherche qui ira chercher les pages contenant les mots demandés (commencer par un mot). La liste des pages trouvées devra être présentée à l'utilisateur par ordre décroissant de PageRank.

Vous débuterez peut-être ce projet avant que le cours sur la matrice Google n'ai eu lieu. La matrice Google contient les probabilités de transition des "surfers" d'une page internet à une autre. Vous pouvez débuter le projet en créant une matrice bidon (ne contenant pas d'état absorbant) et en écrivant un programme qui calcule l'évolution temporelle du vecteur population et la visualise. Après le cours sur Google, vous remplacerez votre matrice par une matrice Google.

7. Le PageRank de Google

Difficulté : moyenne

Le but de ce projet est d'analyser et de caractériser le plus complètement possible l'algorithme PageRank de classement des pages internet utilisé par Google. Vous pouvez utiliser toutes les informations disponibles sur internet. Il convient au minimum d'écrire un code permettant de tester cet algorithme et d'observer l'influence des différents paramètres. Ce code devra permettre de traiter un réseau contenant un très grand nombre de pages, chacune comportant un petit nombre (éventuellement 0) de liens. Vous rendrez un rapport aussi complet et original que possible.

Vous débuterez probablement ce projet avant que le cours sur la matrice Google n'ai eu lieu. La matrice Google contient les probabilités de transition des "surfers" d'une page internet à une autre. Vous pouvez débuter le projet en créant une matrice bidon (ne contenant pas d'état absorbant) et en écrivant un programme qui calcule l'évolution temporelle du vecteur population et la visualise. Après le cours sur Google, vous remplacerez votre matrice par une matrice Google.