

Fiche pédagogique

Activité : Pagerank avec les trains.

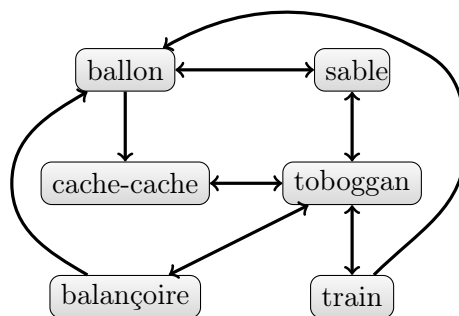
Objectifs pédagogiques : Découvrir comment fonctionnent certains algorithmes de recommandation, notamment PageRank.

Notions abordées : Algorithmes, programmes, algorithmes de recommandation, marche aléatoire.

Matériel nécessaire : Réseau de trains Intelino (+ application sur l'éthique des données).

Niveau : A partir du cycle 3.

Déroulement : On dispose d'un réseau de trains Intelino modélisé par le graphe ci-dessous.



Les sommets représentent des bâtiments sur lesquels sont dessinés les jeux d'enfants dont le nom apparaît sur les sommets. De chaque sommet/bâtiment, un train peut aller vers les sommets/bâtiments vers lesquels une flèche pointe. Par exemple depuis "ballon", le train peut aller vers "cache-cache" ou "sable". Le choix de la direction prise se fait par l'intermédiaire d'aiguillages où le train choisit de manière aléatoire (et équiprobable) s'il part à droite ou à gauche. Ainsi en partant de "ballon", le train a une chance sur deux d'arriver en "cache-cache" et une chance sur deux d'arriver en "sable". De la même manière, en partant de "toboggan", le train a une chance sur quatre d'arriver à chacun des sommets/bâtiments "sable", "cache-cache", "balançoire", "train" grâce à des aiguillages successifs.

On commence par montrer le réseau au public et par expliquer comment les trains fonctionnent (sur batterie, avec un capteur sous le train qui détecte la couleur des différentes cellules sur les rails, etc....) Typiquement, on montre que le comportement du train n'est pas le même suivant la suite de cellules sur laquelle il passe. Entre autres :

- blanc et 1,2 ou 3 verts → aller à vitesse 1 (lente), 2 (moyenne), 3 (rapide).
- blanc, bleu marine → faire demi-tour.
- blanc, 1, 2 ou 3 rouges → faire une pause de 2, 5, ou 10 secondes.

Le train exécute alors un algorithme. Les instructions sont celles données par les suites de cellules, les rails permettent l'enchaînement des instructions, les aiguillages permettent d'avoir un enchaînement aléatoire.

Ensuite, on fait une expérimentation en lançant un train. Lors de cette expérimentation, certaines personnes sont chargées de compter le nombre de passages à chaque bâtiment/sommet. En pratique, la personne en charge de compter le nombre de passages à un bâtiment ajoute un bâton à côté de ce bâtiment à chaque fois que le train y passe.

On fait ensuite observer ce qui se passe. Certains sommets/bâtiments sont plus visités que d'autres. Demander s'ils ont une idée de pourquoi. En général certains voient que le sommet/bâtiment le plus visité ("toboggan") est celui qui est le plus connectée aux autres. Préciser que le nombre de liens ne suffit pas, ce qui compte c'est aussi la "popularité" des sommets/bâtiments qui pointent vers nous. (On peut faire le lien avec des faux profils twitter dans la suite).

On peut ensuite passer à la simulation scratch disponible sur le site suivant.

<https://scratch.mit.edu/projects/558471406/>

Rappeler au passage la différence entre un algorithme et un programme. Un algorithme est une suite d'instructions. Un programme est une suite d'instructions écrites dans un langage (de programmation) pour être comprise par la machine qui va l'exécuter. Le réseau de train et le programme Scratch sont deux programmes qui effectuent le même algorithme. L'un est codé avec les cellules intellino (+ français pour le comptage fait par des humains), l'autre en Scratch. L'avantage de faire un programme sur ordinateur plutôt que d'effectuer l'algorithme avec des trains ou à la main, est qu'on peut aller beaucoup plus vite.

Les simulations sont l'occasion de faire un petit calcul de fréquences. On observe également ce qui se passe. On voit que le train passe environ une fois sur trois par "toboggan". On peut même observer que la fréquence tend vers 1/3 quand le nombre de déplacements augmente.

Faire le lien avec les différents algorithmes de recommandation qu'ils peuvent connaître (la plupart connaissent Youtube, Netflix, Spotify, Twitter etc). Expliquer que l'algorithme Pagerank de Google utilise le même genre de procédés. Expliquer aussi que les algorithmes de recommandation sur les réseaux sociaux, en général, s'adaptent à chaque profil en fonction des données collectées, l'objectif étant en général de nous maintenir plus longtemps en activité sur le site.

On peut faire le lien avec les bulles de filtres, qui peuvent poser des problèmes personnels : si j'écoute beaucoup de jazz, Spotify ne va pas spontanément me faire découvrir des morceaux de rock. Mais aussi des problèmes sociétaux, typiquement sur Twitter : si je partage certaines idéologies politiques, je vais naturellement suivre des personnes qui partagent ces mêmes idées. Avec le temps, j'ai de moins en moins de chances que Twitter me confronte à une idéologie différente de la mienne. On observe ainsi des communautés qui grandissent indépendamment les unes des autres et qui se confrontent de moins en moins (les fameuses "bulles de filtres").

On peut également en profiter pour les mettre en garde sur toutes les informations personnelles qui sont collectées lors de nos pratiques en utilisant la page <https://view.genial.ly/6267b76f14f285001a8ca169/interactive-content-carte-donnees>