

## Activité : Popularité d'une page web

Lorsqu'une personne fait une requête sur un moteur de recherche, celui-ci affiche une liste de pages correspondantes, en présentant les pages les plus populaires en premier. Comment fait ce moteur de recherche pour déterminer les pages les plus populaires ?

Dans cette activité, nous allons présenter une version simplifiée de l'algorithme *PageRank*, inventé par Larry Page (fondateur de Google).

### Présentation de l'algorithme

La plupart des pages web présentent des liens vers d'autres pages web, appelés liens hypertextes. Le principe général est que si une page web est beaucoup référencée par les autres pages web, cette page est plus importante, et sera bien classée par les moteurs de recherche (ceci sera affiné plus tard).

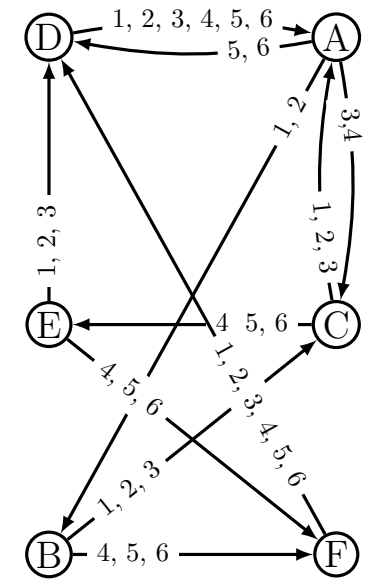
Pour mesurer cette popularité par un nombre, on applique l'algorithme suivant :

- on part d'une page web au hasard (on suppose que cela n'a pas d'importance, ce qui est partiellement vrai) ;
- sur chaque page visitée, on clique sur un lien au hasard sur cette page (pour visiter une nouvelle page) ;
- on répète cette étape un grand nombre de fois.

Durant l'exécution de l'algorithme, on n'oublie pas de compter le nombre de fois que chaque page a été visitée. Le score de chaque page est alors la fréquence du nombre de visites de cette page. Les pages avec le plus haut score seront visitées plus souvent que les autres.

### Premier exemple

Le graphe suivant représente un ensemble de page web, nommées A à F. Les arêtes (flèches) représentent des liens entre les pages. Par exemple, il y a une arête allant de la page B à la page F, ce qui signifie qu'il est possible, en visitant la page B, de cliquer sur un lien pour se retrouver sur la page F (ignorer les numéros pour le moment).



En vous mettant par deux, en vous munissant d'un dé à six face, parcourez aléatoirement ces pages web, en notant bien le nombre de visites. Pour cela, par binôme, pendant quelques minutes, en partant d'un sommet choisi au hasard :

- un élève lance un dé encore et encore ;
- l'autre élève suit le chemin parcouru, en se déplaçant sur le graphe en suivant le résultat du dé.

Au signal de l'enseignant, venez écrire les résultats trouvés sur l'ordinateur du professeur.

### Analyse

1. Reporter sur le graphe plus haut les scores *PageRank* calculés avec la classe.
2. Compter, pour chaque page, le nombre de liens qui proviennent d'autres pages.

On s'intéresse à l'affirmation : « Plus une page a de liens qui viennent vers elle, plus son score *PageRank* est élevé. »

3. En comparant les pages A, B, C, cette affirmation vous semble-t-elle correcte ?
4. En comparant les pages A et D, cette affirmation vous semble-t-elle correcte ? Comment expliquer cela ?

### Search Engine Optimization

Vous êtes l'auteur de la page F, et vous voulez augmenter votre score *PageRank*. En vous servant des réponses aux questions précédentes, imaginez un moyen pour augmenter artificiellement votre score (en créant de nouveaux liens depuis la page F, ou depuis de nouvelles pages que vous créez).