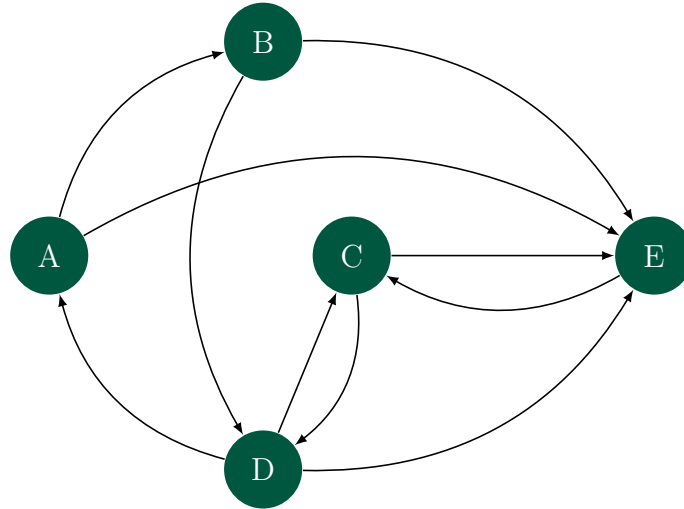


TP 5 – Algorithme de PageRank

En informatique, l'algorithme Page Rank du moteur de recherche de Google est par exemple basé sur l'utilisation des chaînes de Markov. Il utilise le principe d'un surfeur aléatoire se promenant de site en site en cliquant aléatoirement sur un des liens apparaissant sur les pages. Les pages que l'algorithme aura le plus visité seront alors les mieux classées dans le résultat de recherche. On considère le réseau simplifié suivant, constitué de 5 sites Internet.



Le surfeur aléatoire est placé initialement, à $t = 0$, en A. Chaque seconde, il choisit un site vers lequel il se dirige. Chaque lien possible est équiprobable. Cela signifie par exemple qu'à $t = 1$, le surfeur aléatoire sera en B avec une probabilité 0,5 ou en E avec une probabilité 0,5.

Écrire un algorithme en langage Python afin d'estimer la probabilité d'être sur chacun des sites au bout d'un temps très long. Vérifier ensuite que le choix du site de départ ne modifie pas les probabilités calculées.

Histoire

L'idée de PageRank a été présentée en 1998 par **Sergey Brin** et **Larry Page**, les fondateurs de Google, dans un article intitulé *The Anatomy of a Large-Scale Hypertextual Web Search Engine*. Depuis, de nombreuses améliorations ont été apportées. Par exemple, avec l'algorithme décrit dans ce TP, il est possible pour un utilisateur de booster artificiellement le classement de ses pages en se citant soi-même. Avec l'algorithme de PageRank actuel, ce type de liens n'est pas pris en compte.

