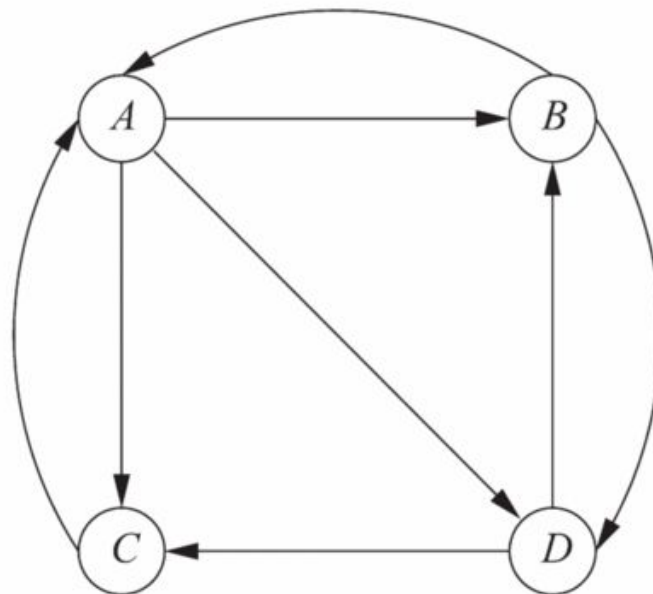


Лабораторная работа. PageRank.

Задание:

реализовать алгоритм поисковой оптимизации PageRank.

- 1) Предусмотреть возможность создания произвольного графа веб-страниц (минимум 4 вершины). Вершина - это веб-страница;
- 2) Связи между вершинами задаются также произвольным образом. Связь между вершинами (ребро графа) означает, что одна веб-страница ссылается на другую веб-страницу;
- 3) В соответствии со связями вершин определить матрицу переходов M , отражающую вероятности переходов со страницы на страницу. Например, зададим произвольный граф, состоящий из 4 вершин:



Тогда матрица M примет следующий вид:

$$M = \begin{bmatrix} 0 & 1/2 & 1 & 0 \\ 1/3 & 0 & 0 & 1/2 \\ 1/3 & 0 & 0 & 1/2 \\ 1/3 & 1/2 & 0 & 0 \end{bmatrix}$$

Т.е. из вершины A можно с равной вероятностью $= 1/3$ попасть в вершину B , C или D . Так как из вершины B выходят два ребра в вершины A и D , то с вероятностью $1/2$ можно попасть как в вершину A , так и в вершину D .

4) Также определим первоначальный вектор v_0 , отражающий вероятности, с которыми пользователь может оказаться на той или иной веб-странице в начале своего “блуждания”. Пользователь с равной долей вероятности может оказаться на любой веб-странице. В примере рассматриваются четыре страницы, следовательно, с вероятностью $\frac{1}{4}$ пользователь может оказаться на странице А, с вероятностью $\frac{1}{4}$ на странице В, с вероятностью $\frac{1}{4}$ на странице С и с вероятностью $\frac{1}{4}$ может оказаться на странице D.

$$\begin{bmatrix} 1/4 \\ 1/4 \\ 1/4 \\ 1/4 \end{bmatrix}$$

Таким образом, $v_0 =$

5) Для понимания того, как изменится вектор v_0 после первого шага пользователя, необходимо умножить матрицу переходов M на вектор v_0 :

Умножение вектора на матрицу

$$\text{Матрица: } \hat{A} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}.$$

$$\text{Векторы: } \vec{B} = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} \quad \text{и} \quad \vec{C} = \begin{pmatrix} c_1 \\ c_2 \\ c_3 \end{pmatrix}$$

$$\vec{C} = \hat{A} \vec{B}$$

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \cdot \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} = \begin{pmatrix} a_{11} \cdot b_1 + a_{12} \cdot b_2 + a_{13} \cdot b_3 \\ a_{21} \cdot b_1 + a_{22} \cdot b_2 + a_{23} \cdot b_3 \\ a_{31} \cdot b_1 + a_{32} \cdot b_2 + a_{33} \cdot b_3 \end{pmatrix} = \begin{pmatrix} c_1 \\ c_2 \\ c_3 \end{pmatrix}$$

Таким образом, мы получим вероятности попадания пользователя на ту или иную страницу. Страница, для которой это значение больше - важнее. Повторив умножение многократно, можно однозначно определить, какая страница важнее остальных (в примере - это страница А):

$$\begin{bmatrix} 1/4 \\ 1/4 \\ 1/4 \\ 1/4 \end{bmatrix}, \begin{bmatrix} 9/24 \\ 5/24 \\ 5/24 \\ 5/24 \end{bmatrix}, \begin{bmatrix} 15/48 \\ 11/48 \\ 11/48 \\ 11/48 \end{bmatrix}, \begin{bmatrix} 11/32 \\ 7/32 \\ 7/32 \\ 7/32 \end{bmatrix}, \dots, \begin{bmatrix} 3/9 \\ 2/9 \\ 2/9 \\ 2/9 \end{bmatrix}$$