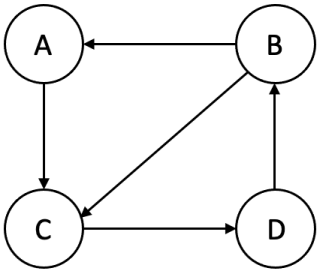
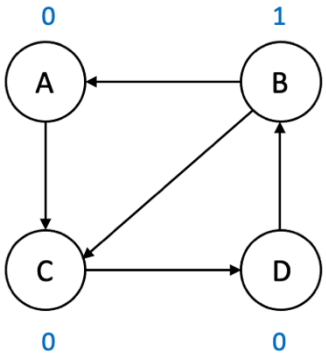
Exercise on PageRank (Second midterm 2022/23)

Let's apply the basic PageRank update rule for the first 2 iterations to the graph below, starting with page B having PageRank equal to 1 and all the other nodes (A, C and D) with PageRank equal to 0.



$PR_0(A) = 0$
$PR_0(B) = 1$
$PR_0(C) = 0$
$PR_0(D) = 0$

The PageRank values are shown in blue.



The basic PageRank update rule is:

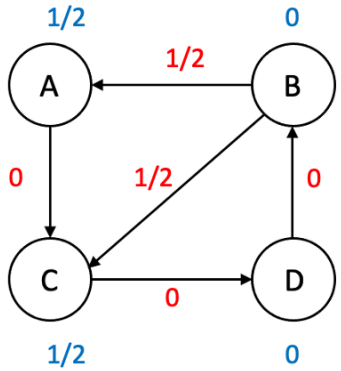
$$PR_{i+1}(u) = \sum_{v \in B_u} \frac{PR_i(v)}{N_v}$$

where:

- $PR_{i+1}(u)$ = PageRank of node u at step i+1
- B_u = set of nodes pointing to u
- v = a node pointing to u
- N_v = number of outgoing links from node v
- $PR_i(v)/N_v$ = fraction of the PageRank value from node v passed to node u

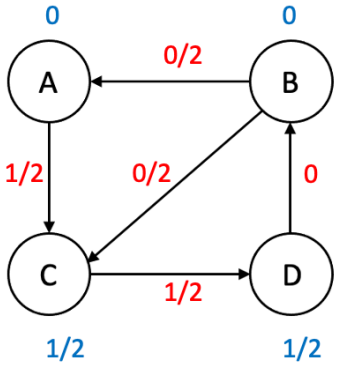
First iteration: i = 1
$PR_1(A) = PR_0(B)/2 = 1/2$
$PR_1(B) = PR_0(D) = 0$
$PR_1(C) = PR_0(A) + PR_0(B)/2 = 0 + 1/2 = 1/2$
$PR_1(D) = PR_0(C) = 0$

Let's show in red how the PageRank is reassigned according to the basic PageRank update rule, in blue the new values. Observe that the sum of the PageRank values is equal to 1.



Second iteration: i = 2
$PR_2(A) = PR_1(B)/2 = 0$
$PR_2(B) = PR_1(D) = 0$
$PR_2(C) = PR_1(A) + PR_1(B)/2 = 1/2 + 0/2 = 1/2$
$PR_2(D) = PR_1(C) = 1/2$

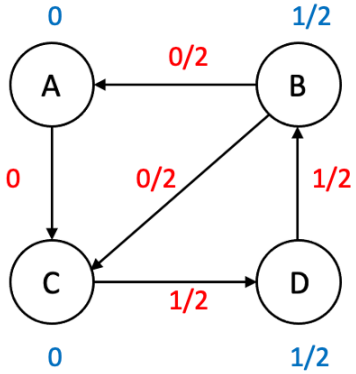
Let's show in red how the PageRank is reassigned according to the basic PageRank update rule, in blue the new values.



Let's consider also the third and fourth iterations.

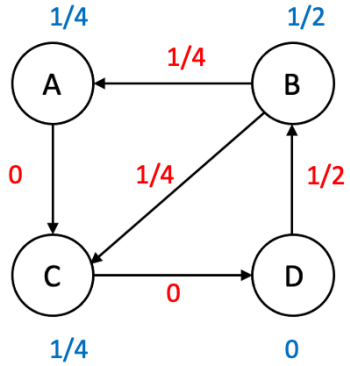
Third iteration: i = 3
$PR_3(A) = PR_2(B)/2 = 0$
$PR_3(B) = PR_2(D) = 1/2$
$PR_3(C) = PR_2(A) + PR_2(B)/2 = 0 + 0/2 = 0$
$PR_3(D) = PR_2(C) = 1/2$

Let's show in red how the PageRank is reassigned according to the basic PageRank update rule, in blue the new values.



Fourth iteration: i = 4
$PR_4(A) = PR_3(B)/2 = 1/4$
$PR_4(B) = PR_3(D) = 1/2$
$PR_4(C) = PR_3(A) + PR_3(B)/2 = 0 + 1/4 = 1/4$
$PR_4(D) = PR_3(C) = 0$

Let's show in red how the PageRank is reassigned according to the basic PageRank update rule, in blue the new values.



Continuing with the iterations, we reach equilibrium with the following PageRank distribution after 111 iterations (with a threshold difference between two solutions set to 10^{-6}):
$PR_k(A) = 1/7$
$PR_k(B) = 2/7$
$PR_k(C) = 2/7$
$PR_k(D) = 2/7$
We obtain the same solution (in 98 iterations) even when starting from an initial condition in which the PageRank is equally distributed among the nodes.
In summary, nodes B, C, and D share the highest rank equally, while node A has half their rank because it only receives half of the flow originating from node B.

Note: alternatively, in this example with no sink nodes, we can also solve directly the linear system using basic algebra.