# MINING MASSIVE DATASETS

## Assignment - 1

## Map Reduce and Page Rank

① $map(15) = [(3, 15), (5, 15)]$

$map(21) = [(3, 21), (7, 21)]$

$map(24) = [(2, 24), (3, 24)]$

$map(30) = [(2, 30), (3, 30), (5, 30)]$

$map(49) = [(7, 49)]$

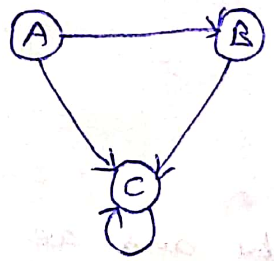
The reducer function is addition

i.e. $reduce(P, [i_1, i_2 \ldots i_k]) = (P, i_1 + i_2 \ldots i_k)$

The result is

$[(2, 54), (3, 90), (5, 45), (7, 70)]$

② Given




$\Rightarrow$

$r_A = 0$

$r_B = \frac{r_A}{2}$

$r_C = \frac{r_A}{2} + r_B + r_C$

$\beta = 0.7$

$r_A + r_B + r_C = 3$

$$M = \begin{array}{c} \\ a \\ b \\ c \end{array} \begin{array}{c} \begin{array}{ccc} a & b & c \end{array} \\ \begin{bmatrix} 0 & 0 & 0 \\ 1/2 & 0 & 0 \\ 1/2 & 1 & 1 \end{bmatrix} \end{array}$$

$$A = \beta . M + (1-\beta) \frac{1}{N} . e . e^T$$

$$A = 0.7 \begin{bmatrix} 0 & 0 & 0 \\ 1/2 & 0 & 0 \\ 1/2 & 1 & 1 \end{bmatrix} + 0.3 \begin{bmatrix} 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \end{bmatrix}$$

$$A = \begin{bmatrix} 0.1 & 0.1 & 0.1 \\ 0.45 & 0.1 & 0.1 \\ 0.45 & 0.8 & 0.8 \end{bmatrix}$$

Initially $r_0 = \begin{bmatrix} 1/3 \\ 1/3 \\ 1/3 \end{bmatrix}$

$$r_1 = \begin{bmatrix} 0.1 & 0.1 & 0.1 \\ 0.45 & 0.1 & 0.1 \\ 0.45 & 0.8 & 0.8 \end{bmatrix} \begin{bmatrix} 0.33 \\ 0.33 \\ 0.33 \end{bmatrix} = \begin{bmatrix} 0.09 \\ 0.21 \\ 0.68 \end{bmatrix}$$

$$r_2 = \quad '' \quad \begin{bmatrix} 0.09 \\ 0.21 \\ 0.68 \end{bmatrix} = \begin{bmatrix} 0.096 \\ 0.129 \\ 0.75 \end{bmatrix}$$

$$r_3 = \begin{bmatrix} 0.097 \\ 0.13 \\ 0.74 \end{bmatrix}$$

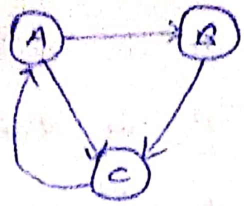
This is the resultant vector and we observe that the values are converged at $r_3$.

So, the page rank for node A $= 0.097 \times 3 = 0.3$

node B $= 0.13 \times 3 = 0.39$

node C $= 0.74 \times 3 = 2.22$

(3) Given




$$\Rightarrow \quad r_A = r_C$$
$$r_B = r_A/2$$
$$r_C = r_A/2 + r_B$$

$$M = \begin{bmatrix} 0 & 0 & 1 \\ 1/2 & 0 & 0 \\ 1/2 & 1 & 0 \end{bmatrix} \qquad \beta = 0.85$$

$$A = \beta \cdot M + (1-\beta)\frac{1}{n} \cdot e \cdot e^T$$

$$A = 0.85 \begin{bmatrix} 0 & 0 & 1 \\ 1/2 & 0 & 0 \\ 1/2 & 1 & 0 \end{bmatrix} + 0.15 \begin{bmatrix} 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \end{bmatrix}$$

$$A = \begin{bmatrix} 0.05 & 0.05 & 0.9 \\ 0.475 & 0.05 & 0.05 \\ 0.475 & 0.9 & 0.05 \end{bmatrix}$$

we multiply A with $r_0$ to get $r_1$

$$r_1 = \begin{bmatrix} 0.33 \\ 0.189 \\ 0.47 \end{bmatrix} \qquad r_2 = A \cdot r_1 = \begin{bmatrix} 0.44 \\ 0.189 \\ 0.35 \end{bmatrix}$$

$$r_3 = A \cdot r_2 = \begin{bmatrix} 0.35 \\ 0.23 \\ 0.39 \end{bmatrix}$$

$$r_4 = A \cdot r_3 = \begin{bmatrix} 0.38 \\ 0.2 \\ 0.39 \end{bmatrix}$$

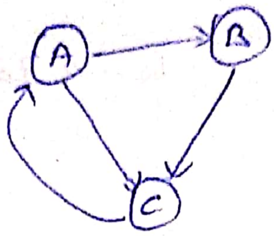
$$r_5 = A \cdot r_4 = \begin{bmatrix} 0.38 \\ 0.2 \\ 0.38 \end{bmatrix}$$

∴ Pagerank of webpage A is 0.38

B is 0.2

C is 0.38

(4) Given




$r_A = r_C$

$r_B = r_A/2$

$r_C = r_A/2 + r_B$

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

Given $r_A = r_B = r_C = 1$

So, $r_0 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$

$$r_1 = \begin{bmatrix} 1 \\ 0.5 \\ 1.5 \end{bmatrix} \quad r_2 = \begin{bmatrix} 1.5 \\ 0.5 \\ 1 \end{bmatrix} \quad r_3 = \begin{bmatrix} 1 \\ 0.75 \\ 1.25 \end{bmatrix} \quad r_4 = \begin{bmatrix} 1.25 \\ 0.5 \\ 1.25 \end{bmatrix} \quad r_5 = \begin{bmatrix} 1.25 \\ 0.6 \\ 1.125 \end{bmatrix}$$

∴ PageRank of webpage A is 1.25

B is 0.6

C is 1.125