

PAGE RANK ALGORITHM

Aman Shankar(as5171)

1. Page rank is like a voting algorithm. The rank of a page completely depends on the votes by other pages.
2. The page rank of each page depends on the page rank of the pages pointing to it. Hence to calculate the page rank of the current page we should calculate the page rank of all the previous pages through which we reach this page.
3. At the start of each page we will set the page rank to 1.
4. After every iteration, we use the following formulate to update the page:

$$\text{Rank}(p) = 0.15 + 0.85 * \text{Contribution}(p)$$
$$\text{Contributions}(p) = \text{Rank}(p) / \text{Neighbors}(p)$$

Where Rank(p) is the rank for page 'P', Contributions (p) is the contributions for the page 'P' and Neighbors(p) are all the neighbors of 'P'

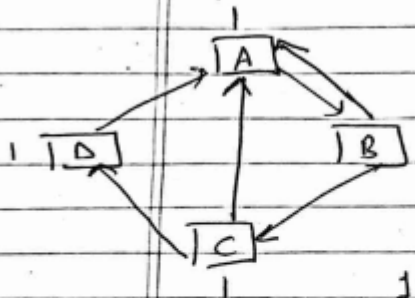
5. Explanation about the formula for Rank(p):

The factor of 0.15 comes for the reason there are no links to the page 'P'. The factor of 0.85 is added to give this page a higher influence than other pages. The vote for page 'P' is given by the $\text{Contributions}(p) = \text{Rank}(p) / \text{Neighbors}(p)$.

6. Code Explanation:

We initialize two RDDs using the spark context object. Two RDDs are the Links and Ranks. Links are page links and their corresponding page ranks are stored in the Ranks RDDs. The first step is to join the two RDDs and change the RDDs using the flatMap method using the following formula: $[(\text{dest}, \text{rank}/\text{len}(\text{links})) \text{ for dest in links}]$. Then we use the reduceByKey method to accumulate the same keys. And then we use MapValues method to change the value of RDDs to the given formulas $(\lambda x: 0.15 + 0.85 * x)$. The above steps are for one iteration. We run the algorithm for a set number of iterations.

PAGE RANK EXAMPLE



First Step Initialisation

All page ranks are 1

$$P(A) = 1 \quad P(B) = 1 \quad P(C) = 1$$

$$P(D) = 1$$

1st Iteration

$$P(A) = 0.15 + 0.85 \left(\frac{P(B)}{2} + P(C) + \frac{P(D)}{2} \right)$$

$$= 0.15 + 0.85 (0.5 + 1 + 0.5) = 1.85$$

$$P(B) = 0.15 + 0.85 (P(A)) = 1$$

$$P(C) = 0.15 + 0.85 \left(\frac{P(A)}{2} \right) = 0.575$$

$$P(D) = 0.15 + 0.85 \left(\frac{P(C)}{2} \right) = 0.575$$

So after 1st Iteration the ranks will be

$$P(A) = 1.85$$

$$P(B) = 1$$

$$P(C) = 0.575$$

$$P(D) = 0.575$$