



**UFES**

**UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO  
CENTRO UNIVERSITÁRIO DO NORTE DO ESTADO - CEUNES**

**MARIANA FERREIRA ROCHA**

**NATALIA ANDRADE CALMON**

**O ALGORITMO GOOGLE DE PAGERANK E SEU CÁLCULO ATRAVÉS DO  
MÉTODO DAS POTÊNCIAS**

**SÃO MATEUS**

**2018/1**

# INTRODUÇÃO

Neste documento será abordado um pouco sobre como Google classifica a pontuação de páginas (PageRank), além de uma aplicação desse método dado um simples exemplo, obtendo o resultado desses cálculos através do Método da Potências.

## 1. PageRank

O sucesso do Google se dá por grande parte por causa do seu algoritmo de classificação de páginas, o pagerank, que identifica a ordem das páginas de acordo com sua importância utilizando um autovetor de uma matriz de links ponderados.

### 1.1. Funcionamento

A atribuição de importância de uma página web é determinada por um número não negativo, que é relacionado aos links feitos à mesma. Para isso, toma-se como base um grafo, onde seus vértices representam as páginas e as arestas são as ligações entre elas.

A partir do grafo é feita a interpretação do mesmo criando-se uma matriz de Markov, onde cada entrada representa a ligação da página  $j$  – identificada como coluna da matriz – com a página  $i$  – identificada como linha da matriz –, considerando a probabilidade de acesso da página  $i$  à partir da página  $j$ . A soma de todos os valores de células de cada coluna não pode ultrapassar 1.

Com a matriz feita, é iniciado um cálculo para obter a porcentagem de tempo gasto para acesso de cada página, assumindo como inicial uma página aleatória. A matriz é elevada a um número consideravelmente grande e logo em seguida é multiplicada pela entrada aleatória, obtém assim a matriz com as porcentagens de tempo gasto de acesso cada página.

### 1.2. Aplicação no exemplo

Para aplicação de um exemplo, tomamos como base um grafo contendo 5 vértices (A, B, C, D e E) que representam as páginas e seus links que são expostos na figura 1.

Como o próximo passo é criar a matriz de Markov, vamos calcular as probabilidades para preenchê-la, seguindo a tabela 1.

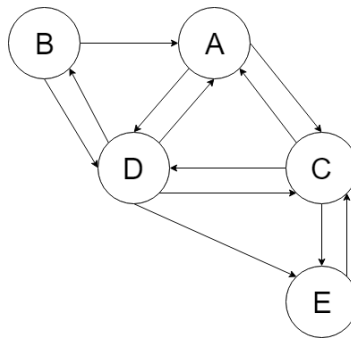


Figura 1 - Grafo

Tabela 1 - Tabela de probabilidades

Página (j)	Quantidade de Links (j->i)	Páginas Incidentes (i)	Probabilidade
A	2	D e C	$\frac{1}{2}$
B	2	A e D	$\frac{1}{2}$
C	3	A, D e E	$\frac{1}{3}$
D	4	A, B, C e D	$\frac{1}{4}$
E	1	C	1

Com isso temos a matriz de Markov e a matriz Y indicando a página inicial como:

$$A = \begin{bmatrix} 0 & \frac{1}{2} & \frac{1}{3} & \frac{1}{4} & 0 \\ 0 & 0 & 0 & \frac{1}{4} & 0 \\ \frac{1}{2} & 0 & 0 & \frac{1}{4} & 1 \\ \frac{1}{2} & \frac{1}{2} & \frac{1}{3} & 0 & 0 \\ 0 & 0 & \frac{1}{3} & \frac{1}{4} & 0 \end{bmatrix} \quad Y_0 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}$$

O algoritmo implementado no scilab para esse cálculo, pode ser visto na abaixo.

```
function pageRank()

    A=[0    1/2 1/3 1/4 0;
        0    0  0  1/4 0;
        1/2  0  0  1/4 1;
        1/2 1/2 1/3  0  0;
        0    0 1/3 1/4 0]

    erro=0.0001

    y0=[0;1;0;0;0]

    A=A^25

    metPotencia(A, erro, y0)

endfunction
```

## 2. Método das Potências

Como o pagerank é determinado pelo cálculo do autovetor, podemos fazer a obtenção pelo método das potências. O método das potências é um método iterativo para o cálculo do maior autovalor de uma matriz e seu autovetor associado.

### 2.1. Funcionamento

Dado uma matriz (A) com as probabilidades de acesso às páginas - como a utilizada pelo método descrito no item 1 deste documento - e uma matriz coluna (Y) identificando uma página inicial aleatória é feita uma multiplicação das duas, obtendo uma matriz Z e, em seguida, calcula-se a nova matriz Y. A nova matriz coluna Y é calculada pela divisão de 1 pelo maior valor entre os elementos de Y já existente, multiplicado pela matriz Z.

Logo depois, é feita a normalização de Y, onde divide o mesmo pelo somatório dos valores pertencente a ele.

Esse processo se repete até que o erro seja menor que 0,001.

### 2.2. Aplicação no exemplo

Para a aplicação desse método, continuamos os passos a partir no método anterior:

1ª iteração :

- Resultado do cálculo da matriz Z ( A\*Y ) :

$$Z_1 = \begin{bmatrix} 0.2 \\ 0.0599 \\ 0.3299 \\ 0.24 \\ 0.17 \end{bmatrix}$$

- Resultado do cálculo do  $\alpha = \max(Z)$ :

$$\alpha = 0.329978$$

- Resultado do cálculo do novo Y ( ( 1/MAX[Y] ) \* Z )

$$Y_1 = \begin{bmatrix} 0.6061 \\ 0.1818 \\ 1 \\ 0.7273 \\ 0.5152 \end{bmatrix}$$

- Resultado do cálculo da normalização de Y ( Y/ $\Sigma Y$  )

$$Y_1 = \begin{bmatrix} 0.2 \\ 0.0599 \\ 0.3299 \\ 0.24 \\ 0.17 \end{bmatrix}$$

2ª iteração :

- Resultado do cálculo da matriz Z ( A\*Y ) :

$$Z_2 = \begin{bmatrix} 0.2 \\ 0.06 \\ 0.33 \\ 0.24 \\ 0.17 \end{bmatrix}$$

- Resultado do cálculo do  $\alpha = \max(Z)$ :

$$\alpha = 0.33$$

- Resultado do cálculo do novo Y ( ( 1/MAX[Y] ) \* Z )

$$Y_2 = \begin{bmatrix} 0.6060 \\ 0.1818 \\ 1 \\ 0.7272 \\ 0.5151 \end{bmatrix}$$

- Resultado do cálculo da normalização de Y ( Y/ $\Sigma Y$ )

$$Y_2 = \begin{bmatrix} 0.2 \\ 0.006 \\ 0.33 \\ 0.24 \\ 0.17 \end{bmatrix}$$

Com isso, podemos analisar o resultado final, após a 2ª iteração com erro 0,000022, que é menor que 0,0001, e concluir a ordem de importância das páginas como  $C \rightarrow D \rightarrow A \rightarrow E \rightarrow B$ .

O algoritmo implementado no scilab para esse cálculo, pode ser visto abaixo.

```
function metPotencia(A, e, y0)

    interacoes = 1 //Número de Interações
    erro=1
    alphaAntigo=0;

    while (erro>e && interacoes <50) //adicionar uma quantidade max de interações
        //Calculo do novo Z
        Z=A*y0

        //Calculo alpha
        alpha=max(Z)

        //Calculo do novo Y
        y=(1/alpha)*Z

        //Calculo erro
        erro = abs(alpha - alphaAntigo)

        //Normalização
        somaY = sum(y)
        y=y/somaY;

        //Atualiza y0 e alpha
        y0=y
        alphaAntigo = alpha

        interacoes=interacoes+1
    end

    disp(alpha, "autovalor")
    disp(y, "autovetor")

endfunction
```

## CONCLUSÃO

Apesar de bem simples, o método das potências pode ser usado em diversas aplicações, como exemplo o cálculo do pagerank que foi descrito e apresentado em exemplo.

## REFERÊNCIAS

Kurt Brian e Tanya Leise, *The Linear Algebra Behind Google*,  
<https://www.rose-hulman.edu/~bryan/googleFinalVersionFixed.pdf>