

PAGERANK 公式

$$PR(a)_{i+1} = \sum_{i=0}^n \frac{\overbrace{PR(T_i)_i}^{\text{其他指向节点 } a \text{ 的节点的 PR}}}{\underbrace{L(T_i)}_{\text{节点 } T_i \text{ 的出链数} / T_i \text{ 的出度}}}$$

代表“循环次数”

例如:

A → B	PR(A)	PR(B)	PR(C)	PR(D)
↑ ↘	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$
D ← C	$\frac{3}{8}$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

$$PR(A)_1 = \frac{PR(C)_0}{L(C)} + \frac{PR(D)_0}{L(D)} = \frac{\frac{1}{4}}{2} + \frac{\frac{1}{4}}{1} = \frac{3}{8}$$

$i = \infty \rightarrow$ 平稳分布 (stationary distribution)

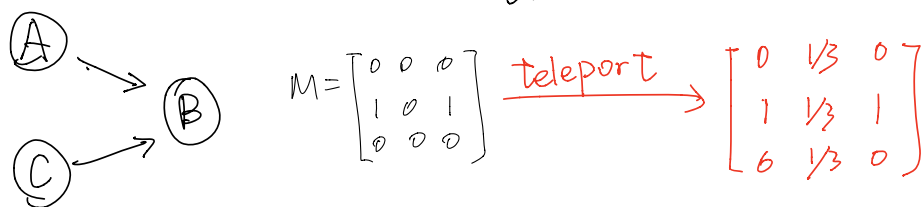
马尔可夫矩阵表示:

$$M = \begin{bmatrix} A & B & C & D \\ 0 & 0 & 1/2 & 1 \\ 1/2 & 0 & 0 & 0 \\ 1/2 & 1 & 0 & 0 \\ 0 & 0 & 1/2 & 0 \end{bmatrix}$$

B只能
走到C

$$r = \begin{bmatrix} \frac{1}{4} \\ \frac{1}{4} \\ \frac{1}{4} \\ \frac{1}{4} \end{bmatrix}, \text{ 所以 } t=1 \text{ 时 } r' = Mr = \begin{bmatrix} 3/8 \\ 1/8 \\ 3/8 \\ 1/8 \end{bmatrix}$$

Dead Ends问题: $\lim_{t \rightarrow \infty} PR(A) = \lim_{t \rightarrow \infty} PR(B) = \lim_{t \rightarrow \infty} PR(C) = 0$



解决方法: Teleport (修正转移概率矩阵 M)

$$M += a^T \left(\frac{e}{n} \right)$$

其中 $a = [a_0, \dots, a_{n-1}]$. $a_i = \begin{cases} 1 & \text{有一列全为0} \\ 0 & \text{其它} \end{cases}$

e 是 $[1, 1, \dots, 1]^T$

n 为矩阵的行数/列数

对于 $M = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$ 得到 $a = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix}$

$$a^T \left(\frac{e}{n} \right) = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1/3 \\ 1/3 \\ 1/3 \end{bmatrix} = \begin{bmatrix} 0 & 1/3 & 0 \\ 0 & 1/3 & 0 \\ 0 & 1/3 & 0 \end{bmatrix}$$

Spider Traps问题: 自环

$$M = \beta M + (1-\beta) \frac{ee^T}{n} \quad \text{默认 } \beta \in (0.8, 0.9)$$

跟随出链到
开网页的概率

随机跳到其
它网页的概率