

# PageRank

## 定义

- 基本定义
  - 非周期图
  - 强连通

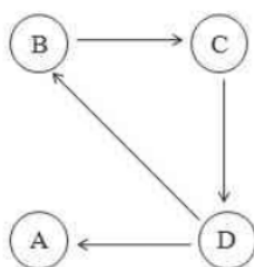
$$MR = R$$

排序泄露



## 排序泄露(Rank Leak)

- 一个独立的网页没有outlink
- 得到不合理的Rank值，并且影响收敛速度



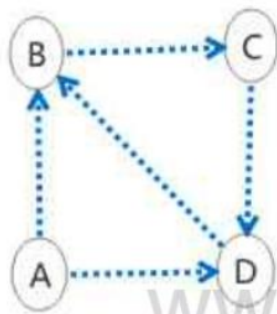
	PR(A)	PR(B)	PR(C)	PR(D)
初始	0.25	0.25	0.25	0.25
一次迭代	0.125	0.125	0.25	0.25
二次迭代	0.125	0.125	0.125	0.25
三次迭代	0.125	0.125	0.125	0.125
...	...	...	...	...
n次迭代	0	0	0	0

排序沉入



## 排序沉入(Rank Sink)

- B、C、D形成loop，没有outlink，不向外分发Rank
- 也会得到不合理的Rank值，并且影响收敛速度



	PR(A)	PR(B)	PR(C)	PR(D)
初始	0.25	0.25	0.25	0.25
一次迭代	0	0.375	0.25	0.375
二次迭代	0	0.375	0.375	0.25
三次迭代	0	0.25	0.375	0.375
四次迭代	0	0.375	0.25	0.375
五次迭代	0	...	...	...

- 一般定义

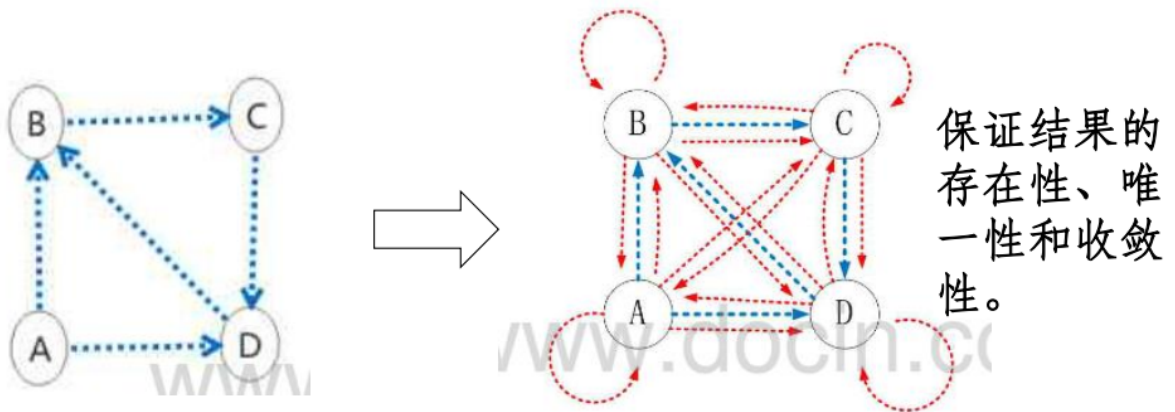
$$R = dMR + \frac{1-d}{n} \mathbf{1}$$



# 改进-RWR

## ➤ Random Walk with Restart

- 随机游走过程中重新开始浏览一个新网页



任意两个顶点之间都有直接通路

在每个顶点处以概率 $\alpha$ 按原来蓝色方向转移，以概率 $1 - \alpha$ 按红色方向转移

35

## 计算

- 迭代算法
- 幂法
- 代数算法