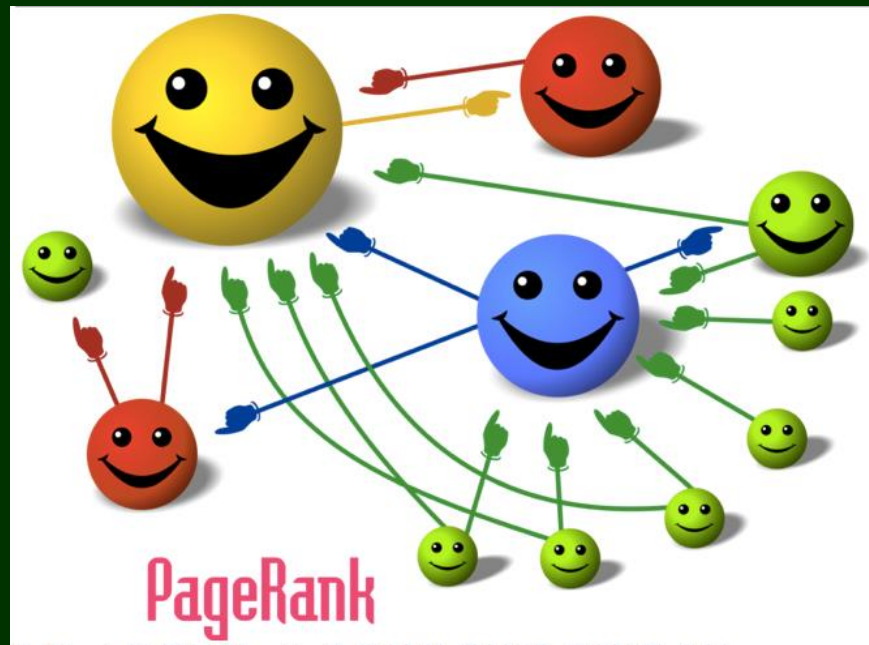


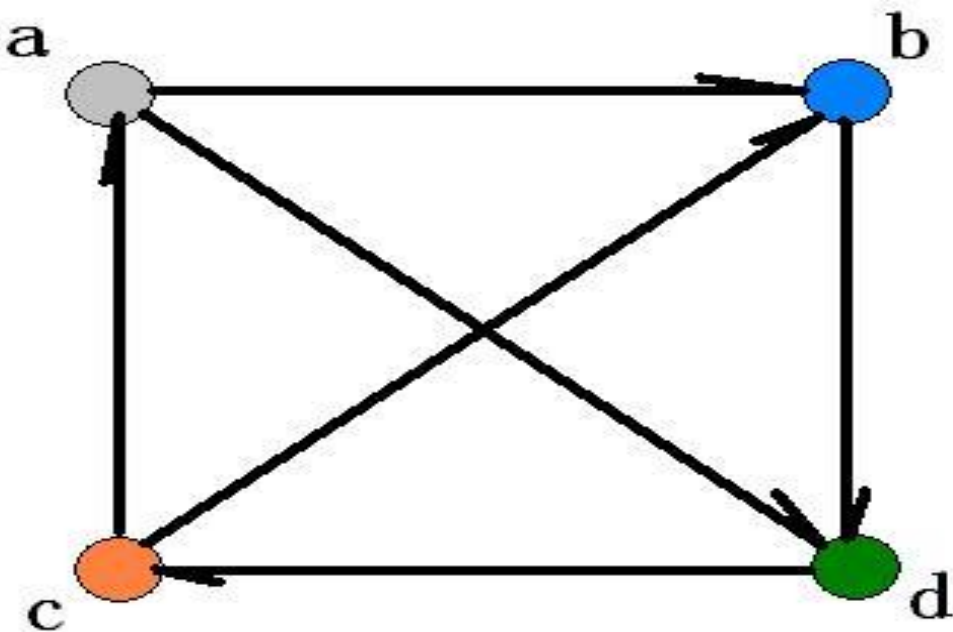
# 网页排名



资料来源: <http://en.wikipedia.org/wiki/File:PageRank-hi-res.png>

# PageRank: 节点重要性的一种测度

基本要领：每一个节点将自己的值均分给出向邻居



每个节点将从邻居收到的值加起来

$$a = c / 2$$

$$b = a / 2 + c / 2$$

$$c = d$$

$$d = a / 2 + b$$

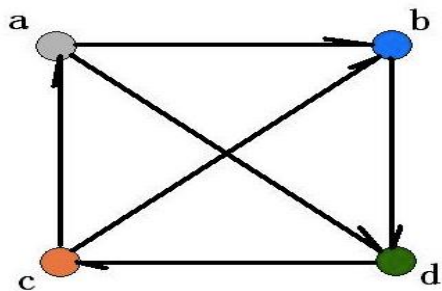
多次迭代！

搜索引擎形成查询结果网页排序的重要参数

# 上图的算例

经过约70次迭代，最后收敛到：  
 $A=0.615$ ,  $B=0.923$ ,  $C=D=1.231$

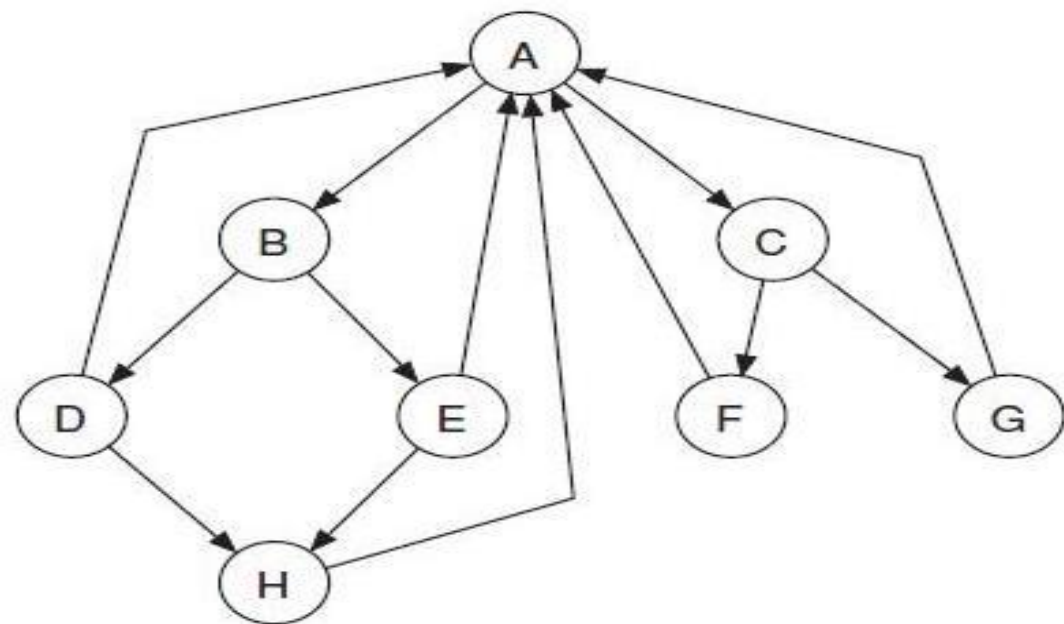
E14					
	A	B	C	D	E
1	$a=c/2$	$b=a/2+c/2$	$c=d$	$d=a/2+b$	传递关系
2	a	b	c	d	
3	1	1	1	1	初值
4	0.5	1	1	1.5	按传递关系计算
5	0.5	0.75	1.5	1.25	
6	0.75	1	1.25	1	
7	0.625	1	1	1.375	
8	0.5	0.8125	1.375	1.3125	
9	0.6875	0.9375	1.3125	1.0625	
10	0.65625	1	1.0625	1.28125	
11	0.53125	0.859375	1.28125	1.328125	
12	0.640625	0.90625	1.328125	1.125	
13	0.6640625	0.984375	1.125	1.2265625	



# PageRank基本算法描述

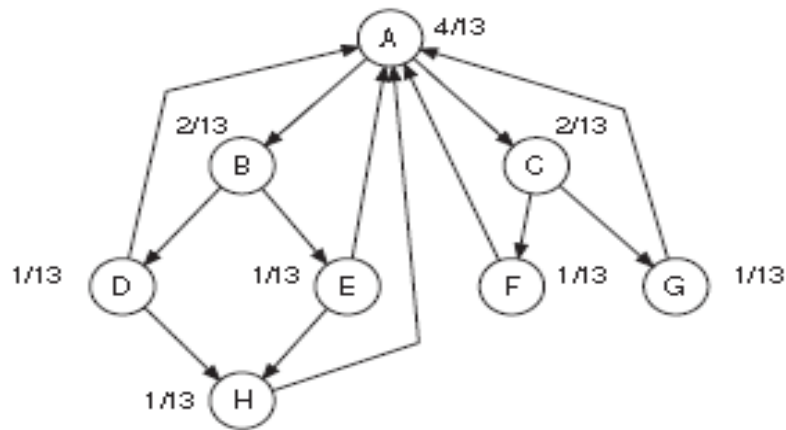
- 输入：一个有 $n$ 个节点的网络（有向图），设所有节点的PageRank初始值为 $1/n$ 。
- 选择操作的步骤数 $k$
- 按照下列规则，同时对每个节点进行操作，做 $k$ 次：
  - 每个节点将自己当前的PageRank值通过出向链接均分传递给所指向的节点
    - 若没有出向链接，则认为传递给自己（或者说保留给自己）
  - 每个节点以从入向链接获得的（包括可能自传的）所有值之和更新它的PageRank

# 一个计算网页排名的实例



Step	A	B	C	D	E	F	G	H
1	$\frac{1}{2}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{8}$
2	$\frac{3}{16}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{32}$	$\frac{1}{32}$	$\frac{1}{32}$	$\frac{1}{32}$	$\frac{1}{16}$

- 每个节点的初值都是  $\frac{1}{8}$
- 最后收敛结果见下图



# 小结

- 在一个由“引用”或者“推荐”关系构成的信息网络中，每个节点的重要性可以认为取决于有多少人推荐，以及推荐人的重要性
- 这种重要性可通过“PageRank算法”得到量化
- PageRank算法的基本精神是基于信息网络的结构，让每个节点不断把自己的重要性分给出向邻居，同时用从入向邻居收到的重要性之和来更新自己