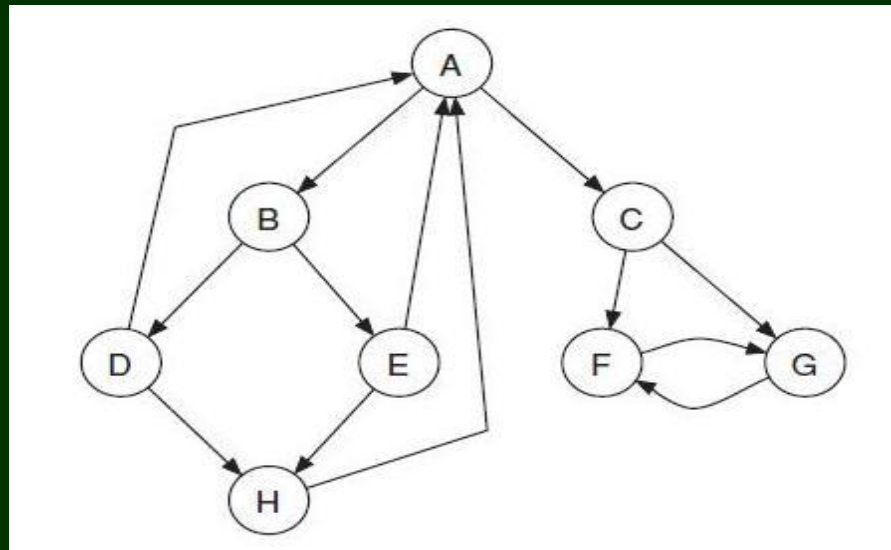
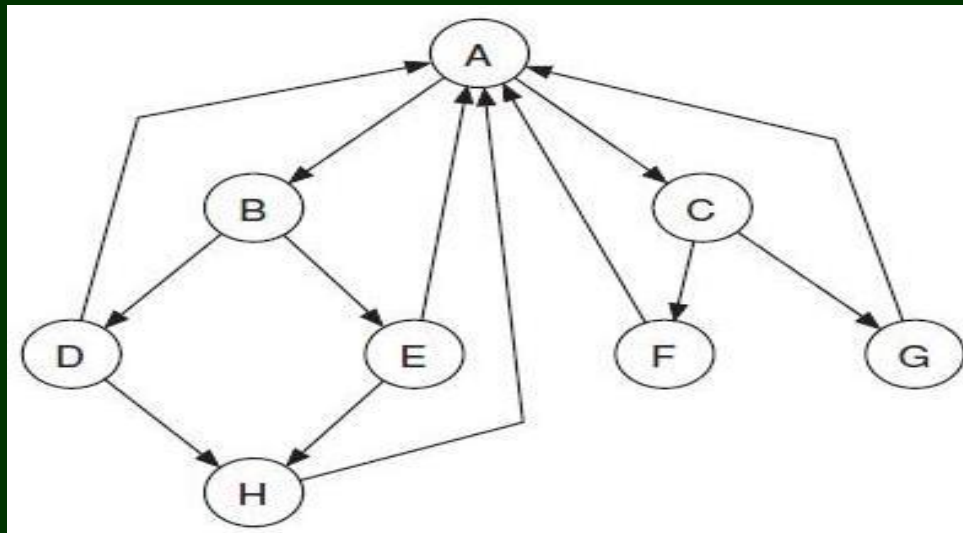


同比缩减与等量补偿

PageRank基本算法在某些结构上的“病态”



- F和G两个节点显得很“自私”：不断吸收别人的价值，但不向外分享

PageRank值很快集中到F和G

pagerank.xlsx								
在工作表中搜索								
开始 布局 表格 图表 SmartArt 公式 数据 审阅								
J82								
	A	B	C	D	E	F	G	H
62	$a=d/2+e/2+h$	$b=a/2$	$c=a/2$	$d=b/2$	$e=b/2$	$f=c/2+g$	$g=c/2+f$	$h=d/2+e/2$
63	a	b	c	d	e	f	g	h
64	0.125	0.125	0.125	0.125	0.125	0.125	0.125	0.125
65	0.25	0.0625	0.0625	0.0625	0.0625	0.1875	0.1875	0.125
66	0.1875	0.125	0.125	0.03125	0.03125	0.21875	0.21875	0.0625
67	0.09375	0.09375	0.09375	0.0625	0.0625	0.28125	0.28125	0.03125
68	0.09375	0.046875	0.046875	0.046875	0.046875	0.328125	0.328125	0.0625
69	0.109375	0.046875	0.046875	0.0234375	0.0234375	0.3515625	0.3515625	0.046875
70	0.0703125	0.0546875	0.0546875	0.0234375	0.0234375	0.375	0.375	0.0234375
71	0.046875	0.03515625	0.03515625	0.02734375	0.02734375	0.40234375	0.40234375	0.0234375
72	0.05078125	0.0234375	0.0234375	0.01757813	0.017578125	0.41992188	0.41992188	0.02734375
73	0.04492188	0.02539063	0.02539063	0.01171875	0.01171875	0.43164063	0.43164063	0.01757813
74	0.02929688	0.02246094	0.02246094	0.01269531	0.012695313	0.44433594	0.44433594	0.01171875
75	0.02441406	0.01464844	0.01464844	0.01123047	0.011230469	0.45556641	0.45556641	0.01269531
76	0.02392578	0.01220703	0.01220703	0.00732422	0.007324219	0.46289063	0.46289063	0.01123047
77	0.01855469	0.01196289	0.01196289	0.00610352	0.006103516	0.46899414	0.46899414	0.00732422
78	0.01342773	0.00927734	0.00927734	0.00598145	0.005981445	0.47497559	0.47497559	0.00610352
79	0.01208496	0.00671387	0.00671387	0.00463867	0.004638672	0.47961426	0.47961426	0.00598145
80	0.01062012	0.00604248	0.00604248	0.00335693	0.003356934	0.48297119	0.48297119	0.00463867
81	0.00799561	0.00531006	0.00531006	0.00302124	0.00302124	0.48599243	0.48599243	0.00335693
82	0.00637817	0.0039978	0.0039978	0.00265503	0.002655029	0.48864746	0.48864746	0.00302124

PageRank的同比缩减与统一补偿规则

- 同比缩减
 - 在每次运行基本PageRank更新规则后，将每一节点的PageRank值都乘以一个小于1的比例因子 s ， $0 < s < 1$ ，经验值在0.8-0.9之间。
- 统一补偿
 - 在每一节点的PageRank值上统一加上 $(1-s)/n$ 。

这样，既维持了“ $\Sigma \text{PageRank} = 1$ ”的性质，也防止了PR值过度集中到个别节点。

pagerank.xlsx									
在工作表中搜索									
开始 布局 表格 图表 SmartArt 公式 数据 审阅									
G86									
	A	B	C	D	E	F	G	H	
62	a=(d/2+e/2+10*0.8+0.2/8			取s=0.8		其他类推			
63	a	b	c	d	e	f	g	h	
64	0.125	0.125	0.125	0.125	0.125	0.125	0.125	0.125	
65	0.225	0.075	0.075	0.075	0.075	0.175	0.175	0.125	
66	0.185	0.115	0.115	0.055	0.055	0.195	0.195	0.085	
67	0.137	0.099	0.099	0.071	0.071	0.227	0.227	0.069	
68	0.137	0.0798	0.0798	0.0646	0.0646	0.2462	0.2462	0.0818	
69	0.14212	0.0798	0.0798	0.05692	0.05692	0.25388	0.25388	0.07668	
70	0.13188	0.081848	0.081848	0.05692	0.05692	0.260024	0.260024	0.070536	
71	0.1269648	0.077752	0.077752	0.0577392	0.0577392	0.2657584	0.2657584	0.070536	
72	0.12762016	0.07578592	0.07578592	0.0561008	0.0561008	0.26870752	0.26870752	0.07119136	
73	0.12683373	0.07604806	0.07604806	0.05531437	0.055314368	0.27028038	0.27028038	0.06988064	
74	0.12515601	0.07573349	0.07573349	0.05541923	0.055419226	0.27164353	0.27164353	0.06925149	
75	0.12473658	0.0750624	0.0750624	0.0552934	0.055293396	0.27260822	0.27260822	0.06933538	
76	0.12470302	0.07489463	0.07489463	0.05502496	0.055024961	0.27311154	0.27311154	0.06923472	
77	0.12440774	0.07488121	0.07488121	0.05495785	0.054957852	0.27344708	0.27344708	0.06901997	
78	0.12418226	0.0747631	0.0747631	0.05495248	0.054952483	0.27371015	0.27371015	0.06896628	
79	0.12413501	0.0746729	0.0746729	0.05490524	0.054905239	0.27387336	0.27387336	0.06896199	
80	0.12409378	0.074654	0.074654	0.05486916	0.054869161	0.27396785	0.27396785	0.06892419	
81	0.12403468	0.07463751	0.07463751	0.0548616	0.054861602	0.27403588	0.27403588	0.06889533	
82	0.12400554	0.07461387	0.07461387	0.054855	0.054855005	0.27408371	0.27408371	0.06888928	
工作表1									
普通视图 就绪 求和=0									

随机游走：PageRank的另一种等价理解

- 想象一个人从一篇随机选择的网页开始，然后随机选择其中的链接浏览到下一篇网页，并不断如此进行，称为“随机游走”
- 考虑任意一篇网页X，问：经过k步随机游走到达X的概率是多少？
- 可以证明：到达X的概率等于运行PageRank基本算法k步得到的值
- 随机游走概念稍加修改也可以和同比缩减统一补偿的PageRank等价

小结

- 信息一旦刻画成一种网络，其中的边经常自然地隐含着一种“推荐”或者“引用”关系，人们可以利用这种关系对信息的作用进行评估
 - 影响力，重要性，权威性，新颖性，...
- 先进的评估方法不仅考虑局部结构，而且会考虑全局结构带来的影响（节点特征性质在网络中的传播）
- HITS算法，PageRank算法
- 当理想遇到现实——重要现实情况的处理
 - 数据范围问题，退化网络结构问题