UNIVERSITY OF Southampton

Page Rank

COMP3220 Web Infrastructure

Dr Heather Packer – hp3@ecs.soton.ac.uk

UNIVERSITY OF Southampton

PageRank

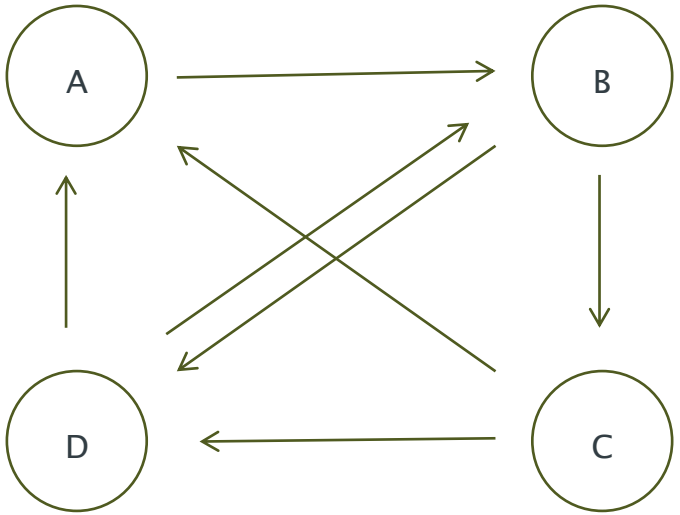
- Algorithm to rank webpages
- Larry Page developed page rank in 1996
- Utilises graph theory of importance
  - Important websites are likely to receive more links from other websites

PageRank: Original Formula

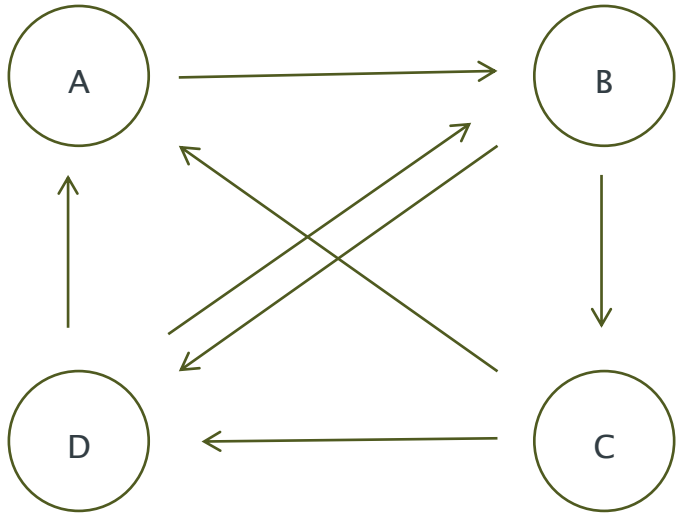
$$PR(u) = \sum_{v \in B_u} \frac{PR(v)}{L(v)}$$

- PageRank (*PR*) of page u is given by the summation of the PR of all pages in the set of all pages linking to page u ($v \in B_u$), divided by the number of $L(v)$ of links from page v
- Iterative formula, starting with rank $1/n$ for all n pages
- It can handle the scale of the web

Websites and Links



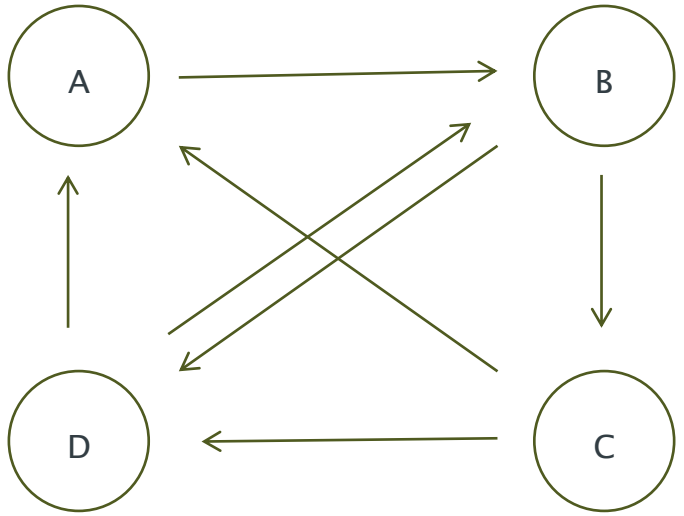
Websites and Links



	Iteration 1
A	1/n = 1/4
B	1/n = 1/4
C	1/n = 1/4
D	1/n = 1/4

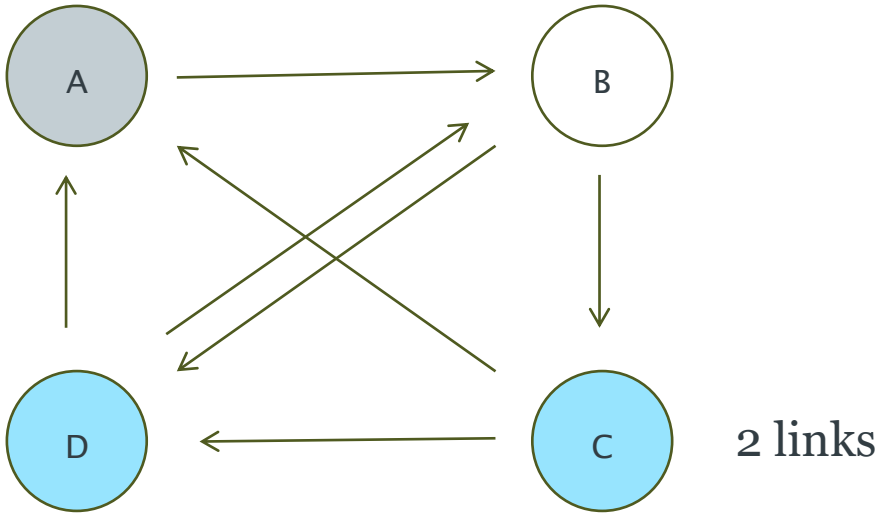
First iteration ranks all pages *1/n*, where *n* is the number of pages

Websites and Links



	Iteration 1
A	1/4
B	1/4
C	1/4
D	1/4

Websites and Links

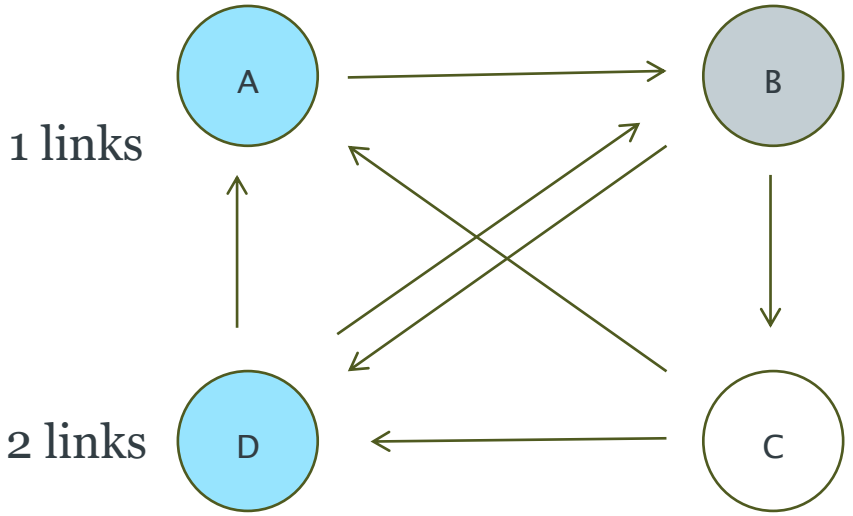


2 links

2 links

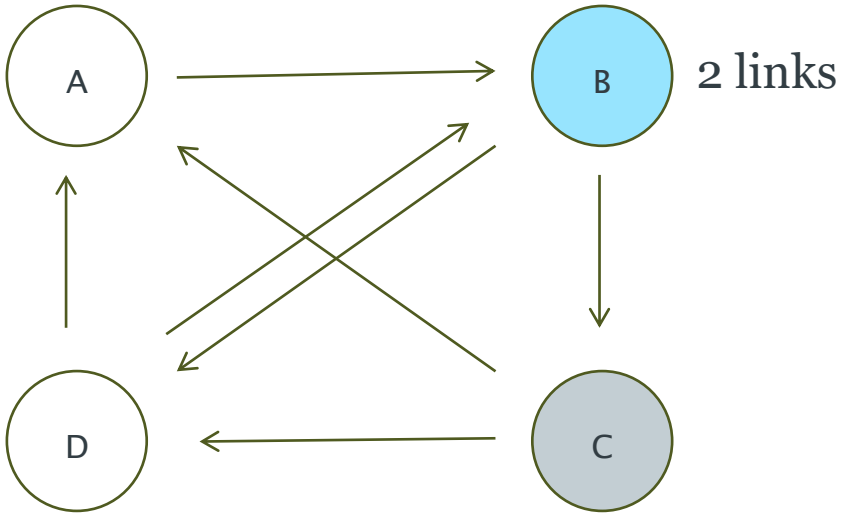
	Iteration 1	Iteration 2
A	1/4	¼ / 2 + ¼ / 2 = 2/8
B	1/4	
C	1/4	
D	1/4	

Websites and Links



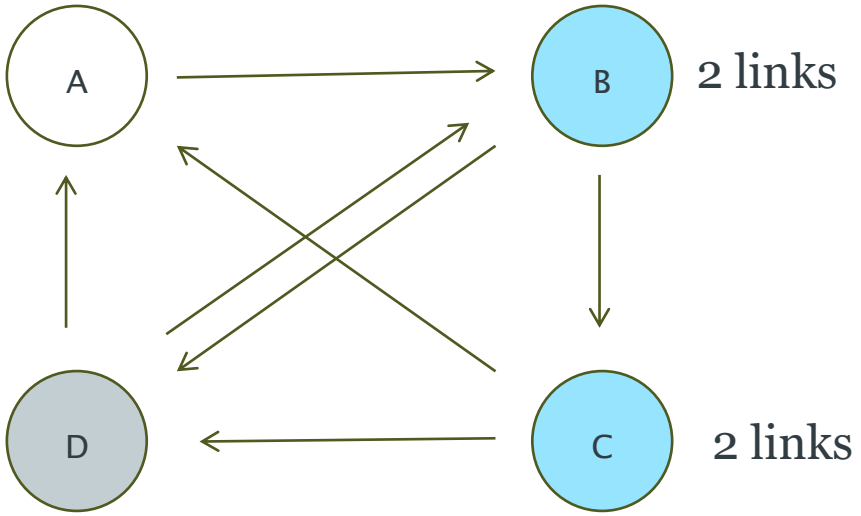
	Iteration 1	Iteration 2
A	1/4	¼ / 2 + ¼ / 2 = 2/8
B	1/4	¼ / 1 + ¼ / 2 = 3/8
C	1/4	
D	1/4	

Websites and Links



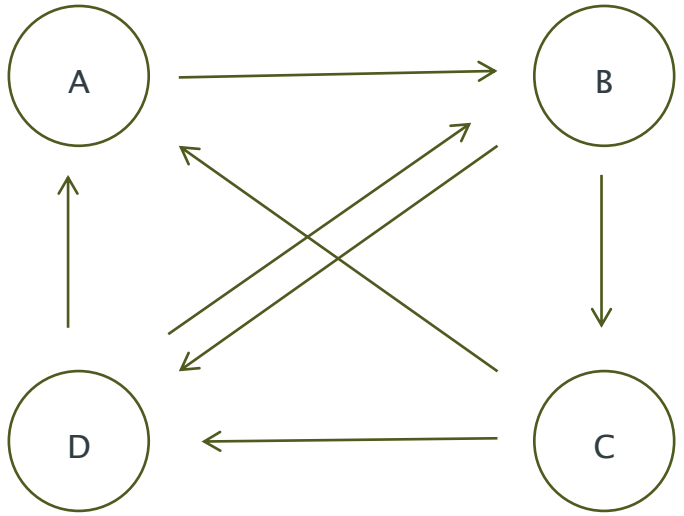
	Iteration 1	Iteration 2
A	1/4	¼ / 2 + ¼ / 2 = 2/8
B	1/4	¼ / 1 + ¼ / 2 = 3/8
C	1/4	¼ / 2 = 1/8
D	1/4	

Websites and Links



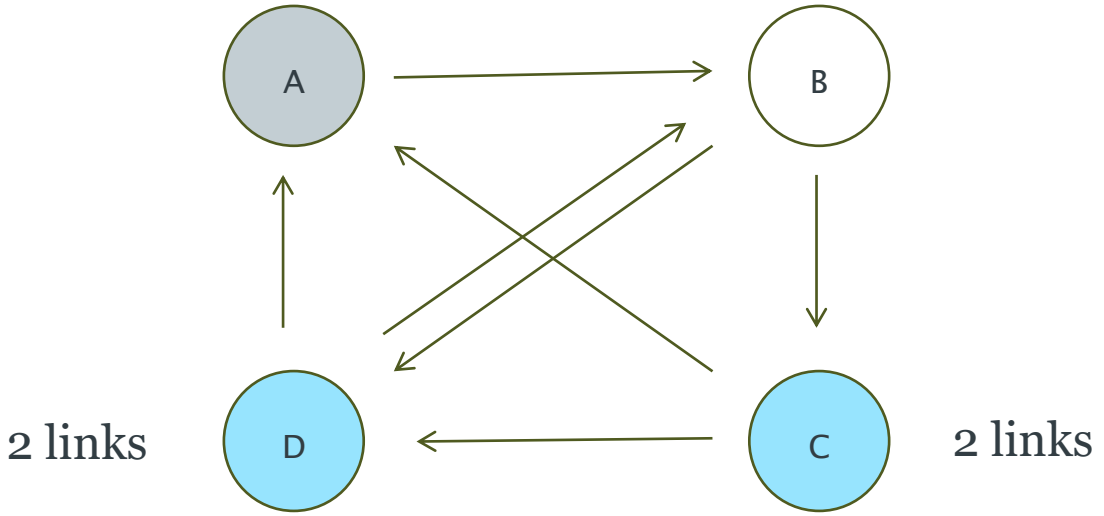
	Iteration 1	Iteration 2
A	1/4	¼ / 2 + ¼ / 2 = 2/8
B	1/4	¼ / 1 + ¼ / 2 = 3/8
C	1/4	¼ / 2 = 1/8
D	1/4	¼ / 2 + ¼ /2 = 2/8

Websites and Links



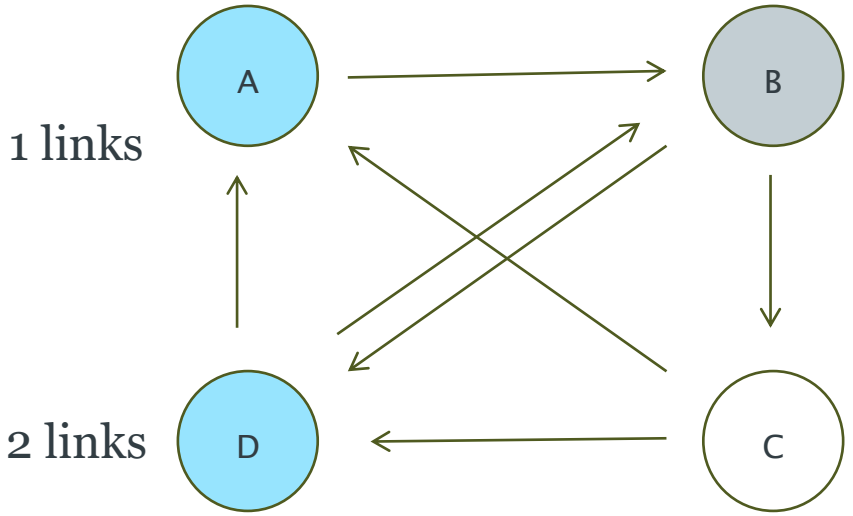
	Iteration 1	Iteration 2
A	1/4	2/8
B	1/4	3/8
C	1/4	1/8
D	1/4	2/8

Websites and Links



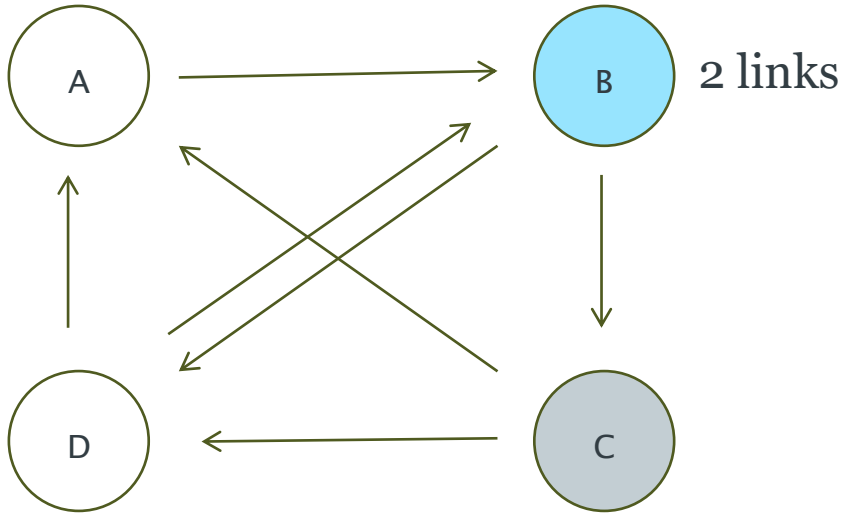
	Iteration 1	Iteration 2	Iteration 3		
A	1/4	2/8	1/8 / 2 + 2/8 / 2	=	3/16
B	1/4	3/8			
C	1/4	1/8			
D	1/4	2/8			

Websites and Links



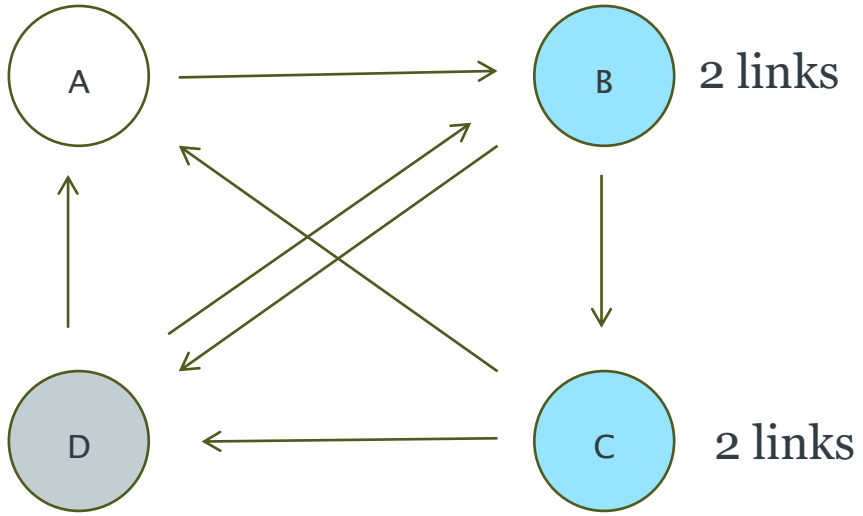
	Iteration 1	Iteration 2	Iteration 2		
A	1/4	2/8	1/8 / 2 + 2/8 / 2	=	3/16
B	1/4	3/8	2/8 / 1 + 2/8 / 2	=	6/16
C	1/4	1/8			
D	1/4	2/8			

Websites and Links



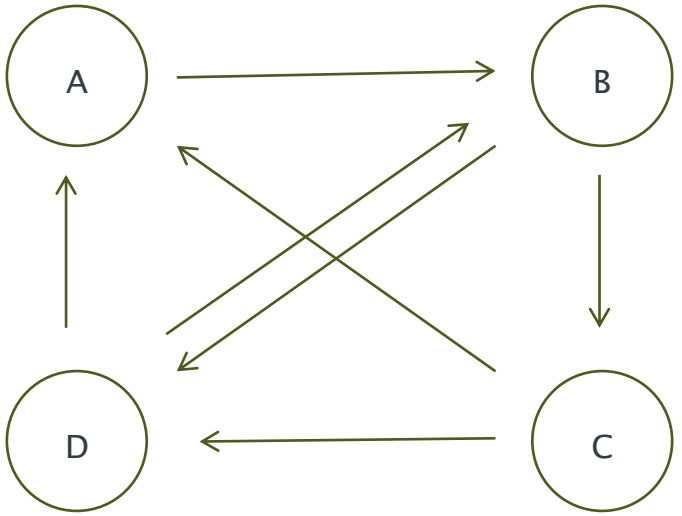
	Iteration 1	Iteration 2	Iteration 2		
A	1/4	2/8	1/8 / 2 + 2/8 / 2	=	3/16
B	1/4	3/8	2/8 / 1 + 2/8 / 2	=	6/16
C	1/4	1/8	3/8 / 2	=	3/16
D	1/4	2/8			

Websites and Links



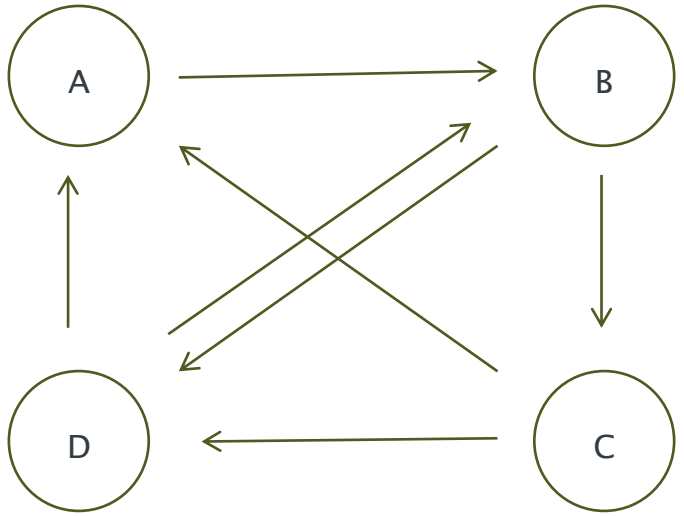
	Iteration 1	Iteration 2	Iteration 2		
A	1/4	2/8	1/8 / 2 + 2/8 / 2	=	3/16
B	1/4	3/8	2/8 / 1 + 2/8 / 2	=	6/16
C	1/4	1/8	3/8 / 2	=	3/16
D	1/4	2/8	3/8 / 2 + 1/8 / 2	=	4/16

Websites and Links



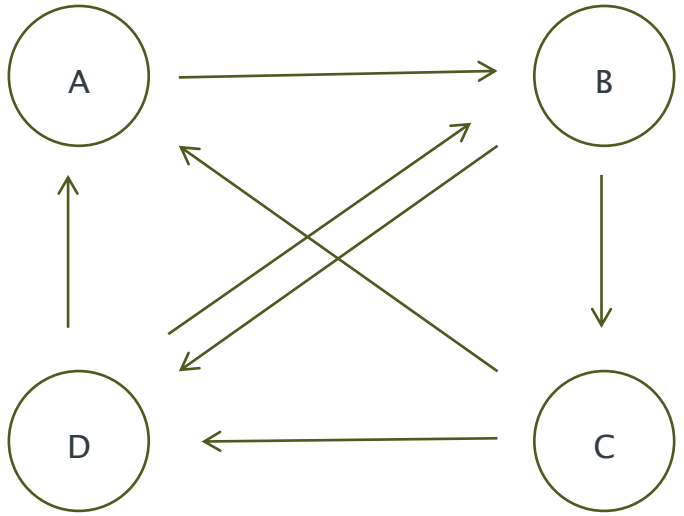
	Iteration 1	Iteration 2	Iteration 3
A	1/4	2/8	3/16
B	1/4	3/8	6/16
C	1/4	1/8	3/16
D	1/4	2/8	4/16

Websites and Links



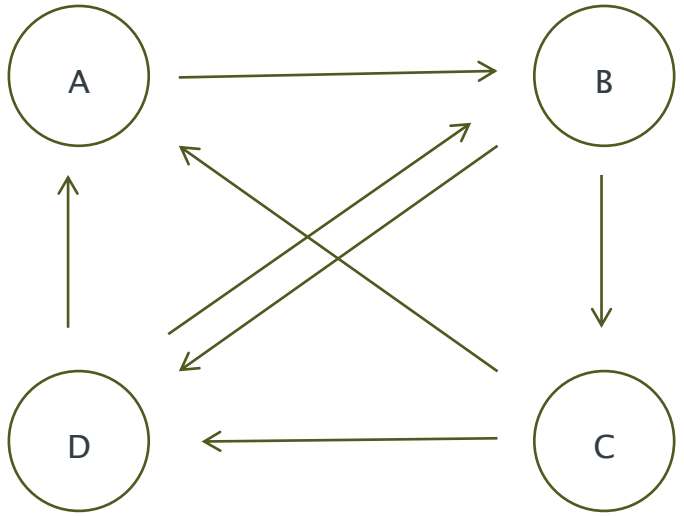
	Iteration 1	Iteration 2	Iteration 3	Rank
A	1/4	2/8	3/16	3
B	1/4	3/8	6/16	1
C	1/4	1/8	3/16	3
D	1/4	2/8	4/16	2

Websites and Links



	Iteration 3	Rank	Iteration 10	Rank
A	3/16	3	0.220458984375	3
B	6/16	1	0.3505859375	1
C	3/16	3	0.17041015625	4
D	4/16	2	0.258544921875	2

Websites and Links



	Iteration 1 Rank	Iteration 2 Rank	Iteration 3 Rank	Iteration 10 Rank
A	1	2	3	3
B	1	1	1	1
C	1	3	3	4
D	1	2	2	2

PageRank

- Assumes important websites are likely be linked to
- Similar to a voting:
  - where in links are votes
  - the quality of a vote is determined by the number of votes (in links)
  - your vote is worth more if you have a higher page rank

Google's PageRank Algorithm

1. Number of links
   - Pages with more in-links rank higher, than a page fewer in-links
2. Link Quality
   - A link from an important page is worth more than many links from relatively unknown sites
3. Link Context
   - The text in and around out links relates to the page they point at.
   - Ranking boosts on text styles

UNIVERSITY OF Southampton

Learning Outcomes

- Be able to calculate the simple PageRank for a particular graph
- Describe the results of PageRank and can interpret the ranking based on influence