




Web Content and Ranking

CS 510 Winter 2007

1



Issues with web pages

- Non-displayed content: tags, attributes, comments
- Scripts – variable names
- Ads
- Frames
- Boilerplate (e.g., navigation panes)
- Images
- Tables

CS 510 Winter 2007

2



Popularity-based ranking

Could use number of back-links

- easily manipulated
- not all links created equal, such as ones for navigating

CS 510 Winter 2007

3



Random surfer model

Random walk on the net

Continuously surfing; on each page, select a link at random to follow next

Interesting in stationary distribution

Over arbitrary stretches of time, what is the relative frequency with which each page is visited: its *page rank*

Page ranks sum to 1

CS 510 Winter 2007

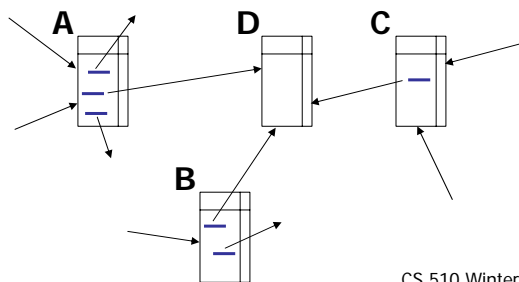
4

Page rank determined by ranks of referring pages

Have ranks r_A , r_B , r_C

$$r_D = \frac{1}{3}r_A + \frac{1}{2}r_B + r_C$$

- Each time you visit A, $\frac{1}{3}$ chance of visiting D
- Each time you visit B, $\frac{1}{2}$ chance of visiting D
- Each time you visit C, you will always visit D



CS 510 Winter 2007

5

Calculating page ranks

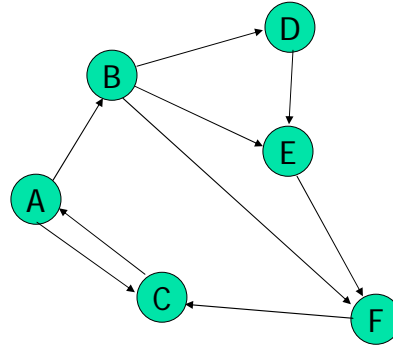
- Have a formula for each page, want to find a solution that satisfies all formulas
- Can get there from an arbitrary initial distribution by iterating it through the formulas

CS 510 Winter 2007

6

Example

$$\begin{aligned} A &= C \\ B &= \frac{1}{2}A \\ C &= \frac{1}{2}A + F \\ D &= \frac{1}{3}B \\ E &= \frac{1}{3}B + D \\ F &= \frac{1}{3}B + E \end{aligned}$$



CS 510 Winter 2007

7

Example: One iteration

Initial	Calculate	Result
$A = .16$	$A = .17$	$A = .17$
$B = .17$	$B = .08$	$B = .08$
$C = .17$	$C = .08 + .17$	$C = .25$
$D = .16$	$D = .06$	$D = .06$
$E = .17$	$E = .06 + .16$	$E = .22$
$F = .17$	$F = .05 + .17$	$F = .22$

CS 510 Winter 2007

8

Further iterations

Iteration	1	2	3	4	5	6	7	8	9
A =	.17	.25	.31	.38	.28	.28	.29	.30	.29
B =	.08	.08	.12	.15	.19	.14	.14	.12	.15
C =	.25	.31	.38	.28	.28	.25	.30	.29	.31
D =	.06	.02	.02	.04	.05	.06	.04	.04	.04
E =	.22	.09	.05	.06	.09	.11	.11	.09	.08
F =	.22	.25	.12	.09	.11	.16	.16	.16	.13

CS 510 Winter 2007

9

Convergence

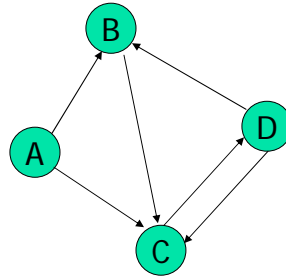
	9	10	11	12
A	.29	.31	.28	.29
B	.15	.14	.15	.14
C	.31	.28	.29	.29
D	.04	.05	.04	.05
E	.08	.09	.09	.09
F	.13	.13	.14	.14

CS 510 Winter 2007

10

Problem: traps

$$\begin{aligned}
 A &= 0 \\
 B &= \frac{1}{2}A + \frac{1}{2}D \\
 C &= \frac{1}{2}A + B + \frac{1}{2}D \\
 D &= C
 \end{aligned}$$



CS 510 Winter 2007

11

Iterations

Initial	1	2	3	4	5	6	7	8	9
A = .25	.00	.00	.00	.00	.00	.00	.00	.00	.00
B = .25	.24	.12	.25	.18	.19	.22	.18	.20	.20
C = .25	.51	.37	.38	.44	.37	.41	.41	.39	.41
D = .25	.25	.51	.37	.38	.44	.37	.41	.41	.39

CS 510 Winter 2007

12



Convergence

9	10	11
.00	.00	.00
.20	.19	.20
.41	.40	.40
.39	.41	.40

CS 510 Winter 2007

13



Bored random surfer

Surfer gets bored

- With probability d : follow a random link on current page
- With probability $(1-d)$: jump to a random page

CS 510 Winter 2007

14

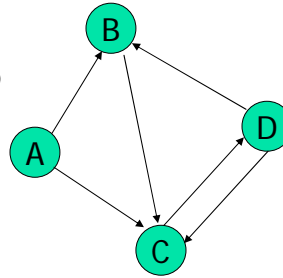
Previous graph, $d = 0.8$

$$A = .05$$

$$B = .8(\frac{1}{2}A + \frac{1}{2}D) + .05$$

$$C = .8(\frac{1}{2}A + B + \frac{1}{2}D) + .05$$

$$D = .8C + .05$$



CS 510 Winter 2007

15

Iterations

Initial	1	2	3	4	5	6	7
A = .25	.05	.05	.05	.05	.05	.05	.05
B = .25	.24	.16	.23	.20	.20	.21	.20
C = .25	.46	.37	.37	.40	.38	.39	.39
D = .25	.25	.42	.35	.35	.37	.35	.36

CS 510 Winter 2007

16



Consequences

- Hard to manufacture rank: need others to point to you
- If you have rank, you dilute your contributions to others with the more links you create