

Google Page Rank Algorithm

Omkar Dash
7th Semester
C.V.Raman College of
Engineering

AGENDUM



- Facts
- Understanding PageRank
- Simple Calculation of PageRank
- Analysis of PageRank Algorithm
- Case Discussion
- Practical Implementation
- References

FACTS



Page Rank:

- Developed by Larry Page and Sergey Brin in 1998
- Trademark of Google
- Patented by Stanford University
- Back bone of Google Search Technology

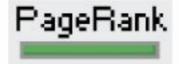
UNDERSTANDING PAGERANK



Page Rank Technology:-

- Ranks pages based on the number of other pages that link to it
- Gives an indication of the relative importance of a page.
- Hence, an appropriate <u>SERP</u> listing
- Calculated by nature and number of <u>back links</u>
- Scale: 0 10 (Google toolbar shows it)

www.toolbar.google.com



Definition of Page Rank

"We assume page A has pages T1...Tn which point to it (i.e., are citations). The parameter d is a damping factor which can be set between 0 and 1. We usually set d to 0.85. Also C(A) is defined as the number of links going out of page A.

The PageRank of a page A is given as follows:

$$PR(A) = (1-d) + d (PR(T1)/C(T1) + ... + PR(Tn)/C(Tn))$$

Note that the PageRanks form a probability distribution over web pages, so the sum of all web pages, PageRanks will be one"

CALCULATING PAGE RANK

DEFINITION OF TERMS

- PR: Shorthand for PageRank: the actual, real, page rank for each page as calculated by Google
- Toolbar PR: The PageRank displayed in the Google toolbar in your browser. This ranges from 0 to 10.



 Back link: If page A links out to page B, then page B is said to have a "back link" from page A.

Toolbar PageRank (log base 10)

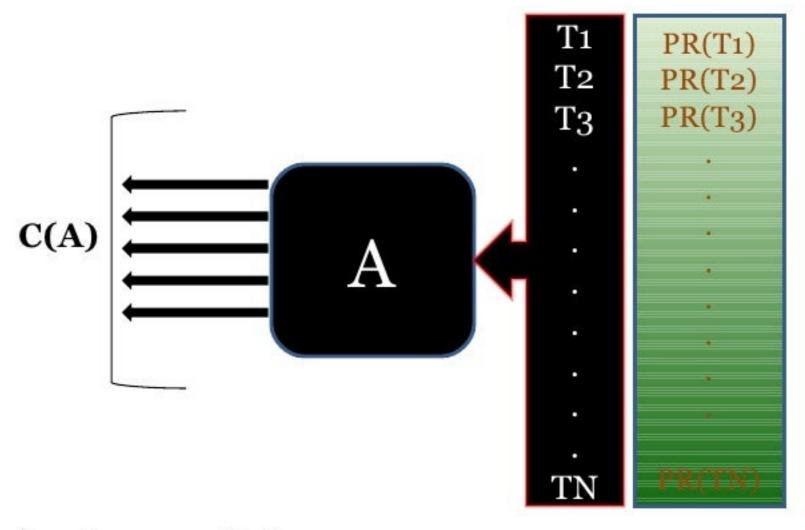
Real PageRank

	0 40
1	0 - 10
,	0 - 10
,	0 10

- **#** We, shall be calculating Real PR here.
- ♯ Toolbar PR is just an indication

CALCULATION OF PAGERANK





The PageRank of a page A is given as follows:

$$PR(A) = (1-d) + d (PR(T1)/C(T1) + ... + PR(Tn)/C(Tn))$$

$$PR(A) = (1-d) + d (PR(T1)/C(T1) + ... + PR(Tn)/C(Tn))$$

- The PR of each page depends on the PR of the pages pointing to it.
- We won't know what PR those pages have until the pages pointing to them have their PR calculated
 -and so on

Remember Chicken-Egg Story !!

Seems impossible to do this calculation!



Who came first?
Chicken Or Egg?

THERE IS A SOLUTION



- PageRank can be calculated using a simple iterative algorithm.
- It means,
 We can calculate a page's PR without knowing the final value of the PR of the other pages...

What we need to do :-

- Remember the each value we calculate
- Repeat the calculations lots of times

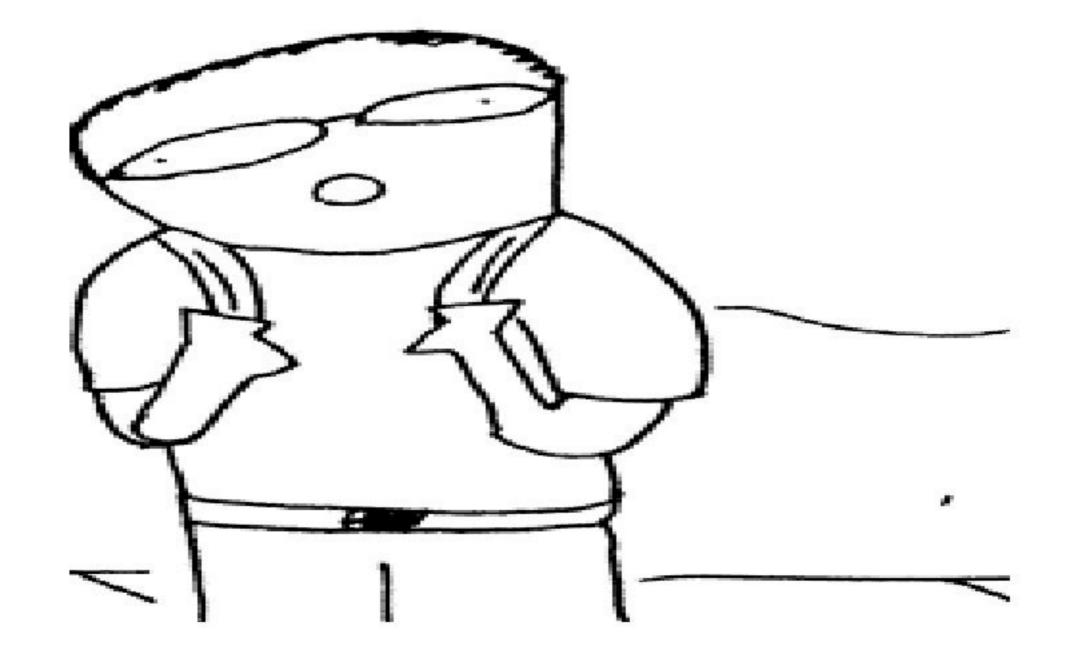
HOWIMANY

TIMES?

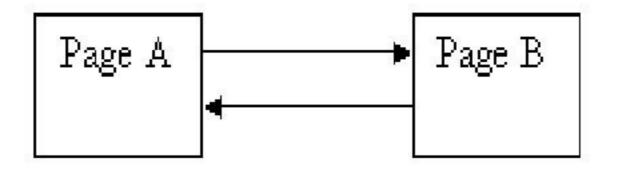
Until the numbers stop changing much.



i don't know...
how?



Lets Learn By Examples!



$$C(A) = 1$$
 and $C(B) = 1$

Let, us assume that PR = 1.0 & d = 0.85 (by theory !!)

$$PR(A) = (1 - d) + d(PR(B)/1)$$

$$PR(B) = (1 - d) + d(PR(A)/1)$$

i.e.

$$PR(A) = 0.15 + 0.85 * 1 = 1$$

$$PR(B) = 0.15 + 0.85 * 1 = 1$$

The numbers aren't changing at all! We made a lucky guess!!

OK BUT WHY SHOULD I ASSUME, PR = 1 ?

WHATIFNOT?

So, Lets Start with PR=0



$$PR(A) = 0.15 + 0.85 * 0 = 0.15$$

 $PR(B) = 0.15 + 0.85 * 0.15 = 0.2775$

Again,

$$PR(A) = 0.15 + 0.85 * 0.2775 = 0.385875$$

 $PR(B) = 0.15 + 0.85 * 0.385875 = 0.47799375$

And again

$$PR(A) = 0.15 + 0.85 * 0.47799375 = 0.5562946875$$

 $PR(B) = 0.15 + 0.85 * 0.5562946875 = 0.622850484375$

Inference: PR approaches 1

OK.. OK.. HOW CAN YOU SAY THAT

PR WON'T CROSS 1?

NO.. IT WONT

LET'S SEE HOW!



Let us assume : PR(A) = 40, PR(B) = 40

First calculation:

$$PR(A) = 0.15 + 0.85 * 40 = 34.15$$

$$PR(B) = 0.15 + 0.85 * 34.15 = 29.1775$$

And again:

$$PR(A) = 0.15 + 0.85 * 29.1775 = 24.950875$$

$$PR(B) = 0.15 + 0.85 * 24.950875 = 21.35824375$$

...... PR will approach and settle down @ 1

CODE FOR PR



```
#! | usr | bi n | perl
   print "Content-Type: text/html/n/n/n";
   $damp = 0.85;
   a = 0; b = 0;
   $i = 40; # loop 40 times
   # forward links
   # a -> b - 1 outgoing link
   # b -> a - 1 outgoing link
   # i.e. "backward" links (what's pointing to me?)
   \# a \leq b
   \# b \le a
   print "I've rounded to 5 decimal places to make the output easier to read
   n";
   while ($i--) {
        printf("a: % 5f b: % 5f\n", $a, $b);
       \$a = (1 - \$damp) + \$damp * (\$b);

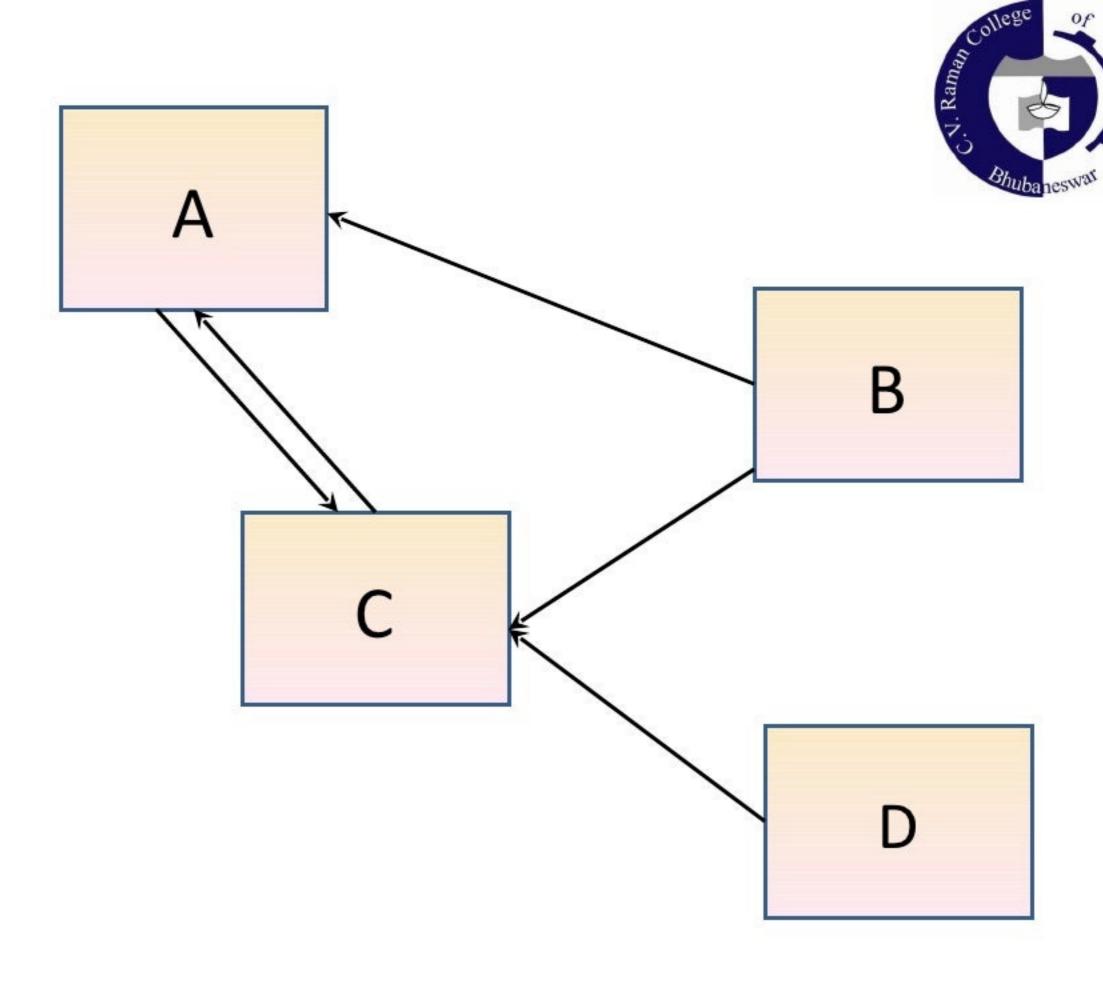
\$b = (1 - \$damp) + \$damp * (\$a);
   printf("Average pagerank = %.4f\n", ($a + $b) / 2);
```

a: 0.00000 0.00000	b:	a: 0.99823 0.99850	b:
a: 0.15000 0.27750	b:	executing	Ō.
a: 0.38588 0.47799	b:	a: 0.99908 0.99922	₽:
a: 0.55629 0.62285	b:	a: 0.99933 0.99943	b:
a: 0.67942 0.72751	b:	a: 0.99952 0.99959	b:
a: 0.76838 0.80313	b:	a: 0.99965 0.99970	b:
a: 0.83266 0.85776	b:	a: 0.99975 0.99979	b:
a: 0.87909 0.89723	b:	a: 0.99982 0.99985	
a: 0.91265 0.92575	b:	a: 0.99987 0.99989	b:
a: 0.93689 0.94635	b:	a: 0.99991 0.99992	b:
a: 0.95440 0.96124	b:	a: 0.99993 0.99994	
a: 0.96705 0.97200	b:	a: 0.99995 0.99996	b:
a: 0.97620 0.97977	b:	a: 0.99996 0.99997	b:
a: 0.98280 0.98538	b:	a: 0.99997 0.99998	
a: 0.98757 0.98944	b:	a: 0.99998 0.99998	b:
a: 0.99102 0.99237	b:	a: 0.99999 0.99999	b:
a: 0.99351 0.99449	b:	a: 0.99999 0.99999	
a: 0.99531 0.99602	b:	Average pagerank a: 0,999999	b:
a: 0.99661 0.99712	b:	a: 0.99999 1.00000	b:
		4 00000	

PRINCIPLE



- It doesn't matter where we start our guess!
- Once the PageRank calculations have settled down: the "normalized probability distribution" (the average PageRank for all pages) will be 1.0



Engineering

CODE FOR GETTING PR

```
Bhubaneswar
```

```
#!/usr/bin/perl
print "Content-Type: text/html\n\n\n";
$damp = 0.85;
n = 1 - danp;
$a = 0;
$b = 0;
$c = 0;
$d = 0;
\$i = 40; \#loop 40 times
# forward links
#a->b, c - 2 outgoing links
#b->c - 1 outgoing link
#c->a - 1 outgoing link
#d->a - 1 outgoing link
#i.e. "backward" links (what's pointing to me?)
```



```
# a <= c
\# b <= a
# c <= a, b, d
# d D not hi ng
while ($i--) {
    printf(
        "a: %5f b: %5f c: %5f d: %5f\n",
        $a, $b, $c, $d
    );
    $a = $norm + $damp * $c;
    $b = $norm + $damp * ($a/2);
    c = snorm + sdamp * (sa/2 + sb + sd);
    $d = $norm
printf("Average pagerank = %4f\n", ($a + $b + $c + $d) /
  4);
```

- a: 0.00000 b: 0.00000 c: 0.00000 d: 0.00000
- a: 0.15000 b: 0.21375 c: 0.39544 d: 0.15000
- a: 0.48612 b: 0.35660 c: 0.78721 d: 0.15000
- a: 0.81913 b: 0.49813 c: 1.04904 d: 0.15000
- a: 1.04169 b: 0.59272 c: 1.22403 d: 0.15000
- a: 1.19042 b: 0.65593 c: 1.34097 d: 0.15000
- a: 1.28982 b: 0.69818 c: 1.41912 d: 0.15000
- a: 1.35626 b: 0.72641 c: 1.47136 d: 0.15000
- a: 1.40065 b: 0.74528 c: 1.50626 d: 0.15000
- a: 1.43032 b: 0.75789 c: 1.52959 d: 0.15000
- a: 1.45015 b: 0.76632 c: 1.54518 d: 0.15000
- a: 1.46341 b: 0.77195 c: 1.55560 d: 0.15000
- a: 1.47226 b: 0.77571 c: 1.56257 d: 0.15000
- a: 1.47818 b: 0.77823 c: 1.56722 d: 0.15000
- a: 1.48214 b: 0.77991 c: 1.57033 d: 0.15000
- a: 1.48478 b: 0.78103 c: 1.57241 d: 0.15000
- a: 1.48655 b: 0.78178 c: 1.57380 d: 0.15000
- a: 1.48773 b: 0.78228 c: 1.57473 d: 0.15000
- a: 1.48852 b: 0.78262 c: 1.57535 d: 0.15000
- a: 1.48904 b: 0.78284 c: 1.57576 d: 0.15000

```
a: 1.48940 b: 0.78299 c: 1.57604 d: 0.15000
```

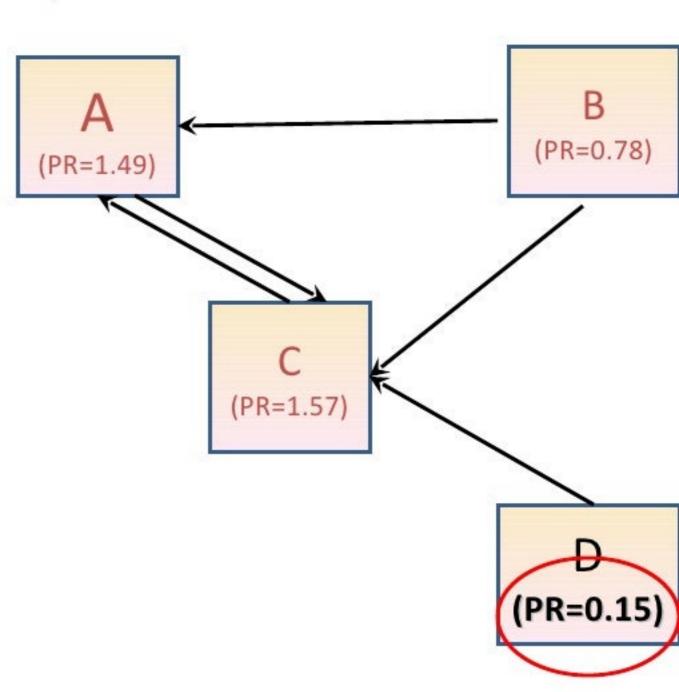
Average pagerank = 1.0000

..... just graphically representing it!



Its okay if A, B, C have Page Ranks, But how come D has?

It has got no in bound links!!



a: 1.48940 b: 0.78299 c: 1.57604 d: 0.15000

a: 1.48963 b: 0.78309 c: 1.57622 d: 0.15000

a: 1.48979 b: 0.78316 c: 1.57635 d: 0.15000

a: 1.48990 b: 0.78321 c: 1.57643 d: 0.15000

a: 1.48997 b: 0.78324 c: 1.57649 d: 0.15000

a: 1.49001 b: 0.78326 c: 1.57652 d: 0.15000

a: 1.49004 b: 0.78327 c: 1.57655 d: 0.15000

a: 1.49007 b: 0.78328 c: 1.57656 d: 0.15000

a: 1.49008 b: 0.78328 c: 1.57657 d: 0.15000

a: 1.49009 b: 0.78329 c: 1.57658 d: 0.15000

a: 1.49009 b: 0.78329 c: 1.57659 d: 0.15000

a: 1.49010 b: 0.78329 c: 1.57659 d: 0.15000

a: 1.49011 b: 0.78329 c: 1.57660 d: 0.15000

a: 1.49011 b: 0.78330 c: 1.57660 d: 0.15000



Let's revisit the basic equation:

•
$$PR(A) = (1-d) +$$

 $d(PR(T1)/C(T1) + ... + PR(Tn)/C(Tn))$

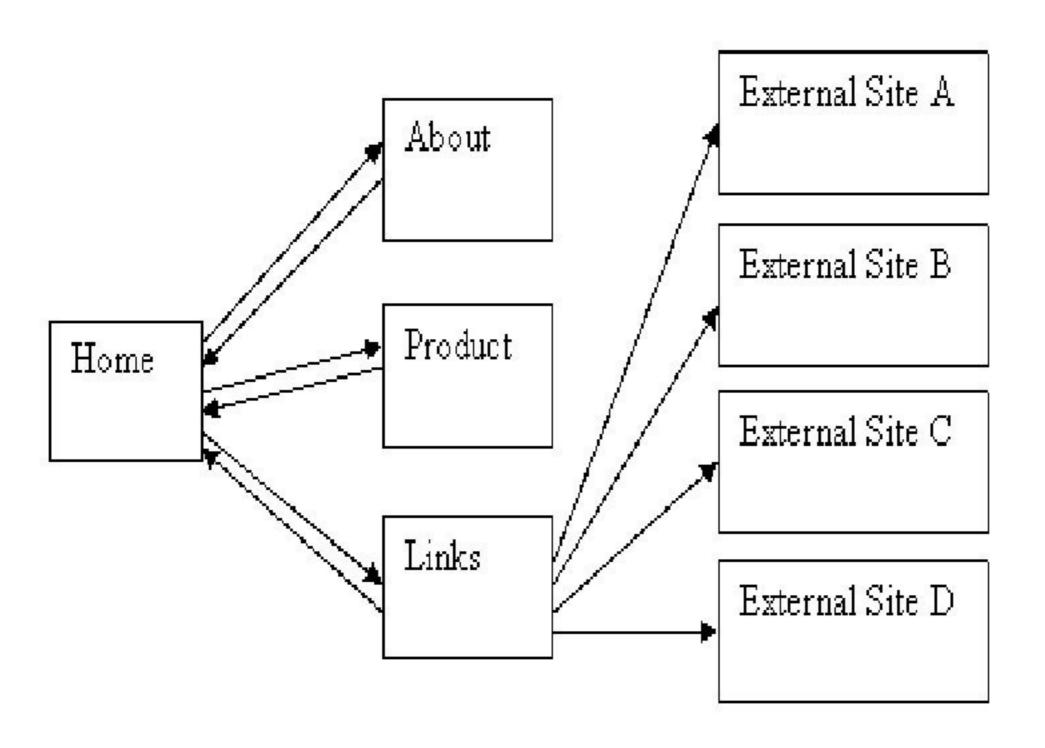
- So, for Page D, no back links means the equation looks like this: PR(A)=(1-d)+d*(0)=0.15
- Every page has at least a PR of 0.15 to share out.



A

Sample Case...



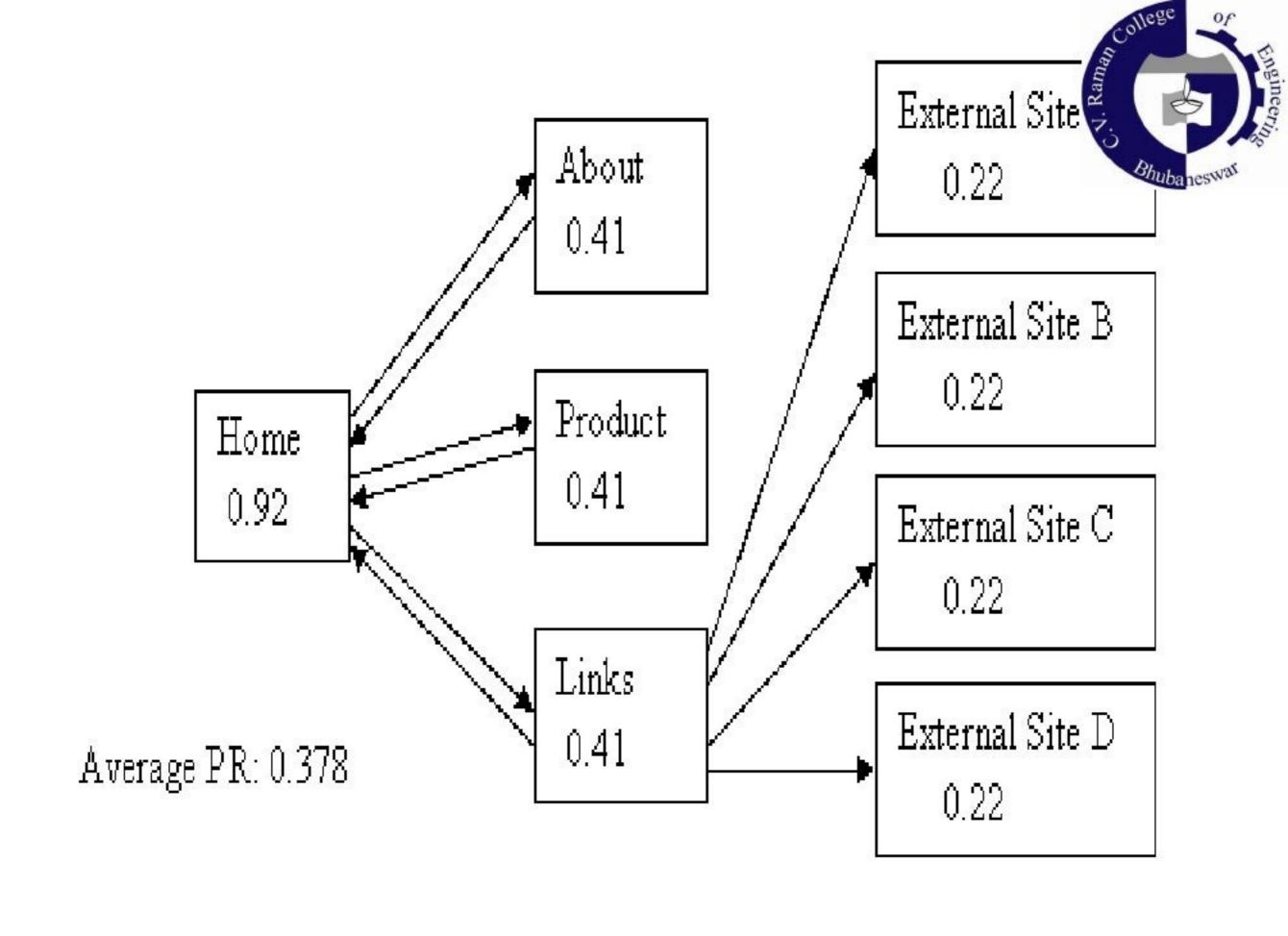


```
College College
#!/usr/bin/perl
print "Content-Type: text/html\n\n\n";
$danp = 0.85;
a = b = c = d = e = f = g = h = 0;
$i = 40; # loop 40 times
# forward links
# a -> b, c, d - 3 outgoing links - home
#b->a - 1 outgoing link - about
# c -> a - 1 outgoing link - products
# d -> a, e, f, g, h - 5 outgoing links - links
# e, f, g, h - nothing - external sites
# i.e. "backward" links (what's pointing to me?)
# a <= b, c, d
# b, c, d <= a
# e, f, g, h <= d
whi I eÊ($i --) Ê{
   ÊÊÊÊprint f
   (ÊÊÊ' a: Ê% 5f Êb: Ê% 5f Êc: Ê% 5f Êd: Ê% 5f Êe: Ê% 5f Êf : Ê% 5f Êg: Ê% 5f Êh: Ê% 5f \ n",
   ÊÊÊÊÊÊÊÊ$a, Ê$b, Ê$c, Ê$d, Ê$e, Ê$f, Ê$g, Ê$h
   ÊÊÊÊ):
   ÊÊÊÊ$aÊ=Ê1Ê-Ê$danpÊ+Ê$danpÊ*Ê($bÊ+Ê$cÊ+Ê$d/5);
   ÊÊÊÊ$bÊ=Ê1Ê-Ê$danpÊ+Ê$danpÊ*Ê($a/3);
   ÊÊÊÊ$cÊ=Ê1Ê-Ê$danpÊ+Ê$danpÊ*Ê($a/3);
   ÊÊÊÊ$dÊ=Ê1Ê-Ê$danpÊ+Ê$danpÊ*Ê($a/3);
   ÊÊÊÊ$eÊ=Ê$f Ê=Ê$gÊ=Ê$hÊ=Ê1Ê-Ê$danpÊ+Ê$danpÊ*Ê($d/5);
   printf
( "Aver ageÊpager ankÊ=Ê% 4f \ n", Ê( $aÊ+Ê$bÊ+Ê$cÊ+Ê$dÊ+Ê$eÊ+Ê$f Ê+Ê$gÊ+Ê$h) Ê/Ê8);
```

```
a: 0.00000 b: 0.00000 c: 0.00000 d: 0.00000 e: 0.00000 f: 0.00000 g: 0.00000 h: 0.00000
a: 0.15000 b: 0.19250 c: 0.19250 d: 0.19250 e: 0.18273 f: 0.18273 g: 0.18273 h: 0.18273
a: 0.50998 b: 0.29449 c: 0.29449 d: 0.29449 e: 0.20006 f: 0.20006 g: 0.20006 h: 0.20006
a: 0.70070 b: 0.34853 c: 0.34853 d: 0.34853 e: 0.20925 f: 0.20925 g: 0.20925 h: 0.20925
a: 0.80176 b: 0.37716 c: 0.37716 d: 0.37716 e: 0.21412 f: 0.21412 g: 0.21412 h: 0.21412
a: 0.85530 b: 0.39233 c: 0.39233 d: 0.39233 e: 0.21670 f: 0.21670 g: 0.21670 h: 0.21670
a: 0.88366 b: 0.40037 c: 0.40037 d: 0.40037 e: 0.21806 f: 0.21806 g: 0.21806 h: 0.21806
a: 0.89869 b: 0.40463 c: 0.40463 d: 0.40463 e: 0.21879 f: 0.21879 g: 0.21879 h: 0.21879
a: 0.90666 b: 0.40689 c: 0.40689 d: 0.40689 e: 0.21917 f: 0.21917 g: 0.21917 h: 0.21917
a: 0.91088 b: 0.40808 c: 0.40808 d: 0.40808 e: 0.21937 f: 0.21937 g: 0.21937 h: 0.21937
a: 0.91311 b: 0.40872 c: 0.40872 d: 0.40872 e: 0.21948 f: 0.21948 g: 0.21948 h: 0.21948
a: 0.91430 b: 0.40905 c: 0.40905 d: 0.40905 e: 0.21954 f: 0.21954 g: 0.21954 h: 0.21954
a: 0.91493 b: 0.40923 c: 0.40923 d: 0.40923 e: 0.21957 f: 0.21957 g: 0.21957 h: 0.21957
a: 0.91526 b: 0.40932 c: 0.40932 d: 0.40932 e: 0.21958 f: 0.21958 g: 0.21958 h: 0.21958
a: 0.91543 b: 0.40937 c: 0.40937 d: 0.40937 e: 0.21959 f: 0.21959 g: 0.21959 h: 0.21959
a: 0.91553 b: 0.40940 c: 0.40940 d: 0.40940 e: 0.21960 f: 0.21960 g: 0.21960 h: 0.21960
a: 0.91558 b: 0.40941 c: 0.40941 d: 0.40941 e: 0.21960 f: 0.21960 g: 0.21960 h: 0.21960
a: 0.91560 b: 0.40942 c: 0.40942 d: 0.40942 e: 0.21960 f: 0.21960 g: 0.21960 h: 0.21960
a: 0.91562 b: 0.40942 c: 0.40942 d: 0.40942 e: 0.21960 f: 0.21960 g: 0.21960 h: 0.21960
a: 0.91562 b: 0.40943 c: 0.40943 d: 0.40943 e: 0.21960 f: 0.21960 g: 0.21960 h: 0.21960
```

a: 0.91563 b: 0.40943 c: 0.40943 d: 0.40943 e: 0.21960 f: 0.21960 g: 0.21960 h: 0.21960 a: 0.91563 b: 0.40943 c: 0.40943 d: 0.40943 e: 0.21960 f: 0.21960 g: 0.21960 h: 0.21960 a: 0.91563 b: 0.40943 c: 0.40943 d: 0.40943 e: 0.21960 f: 0.21960 g: 0.21960 h: 0.21960 a: 0.91563 b: 0.40943 c: 0.40943 d: 0.40943 e: 0.21960 f: 0.21960 g: 0.21960 h: 0.21960 a: 0.91563 b: 0.40943 c: 0.40943 d: 0.40943 e: 0.21960 f: 0.21960 g: 0.21960 h: 0.21960 a: 0.91563 b: 0.40943 c: 0.40943 d: 0.40943 e: 0.21960 f: 0.21960 g: 0.21960 h: 0.21960 a: 0.91563 b: 0.40943 c: 0.40943 d: 0.40943 e: 0.21960 f: 0.21960 g: 0.21960 h: 0.21960 a: 0.91563 b: 0.40943 c: 0.40943 d: 0.40943 e: 0.21960 f: 0.21960 g: 0.21960 h: 0.21960 a: 0.91563 b: 0.40943 c: 0.40943 d: 0.40943 e: 0.21960 f: 0.21960 g: 0.21960 h: 0.21960 a: 0.91563 b: 0.40943 c: 0.40943 d: 0.40943 e: 0.21960 f: 0.21960 g: 0.21960 h: 0.21960 a: 0.91563 b: 0.40943 c: 0.40943 d: 0.40943 e: 0.21960 f: 0.21960 g: 0.21960 h: 0.21960 a: 0.91563 b: 0.40943 c: 0.40943 d: 0.40943 e: 0.21960 f: 0.21960 g: 0.21960 h: 0.21960 a: 0.91563 b: 0.40943 c: 0.40943 d: 0.40943 e: 0.21960 f: 0.21960 g: 0.21960 h: 0.21960 a: 0.91563 b: 0.40943 c: 0.40943 d: 0.40943 e: 0.21960 f: 0.21960 g: 0.21960 h: 0.21960 a: 0.91563 b: 0.40943 c: 0.40943 d: 0.40943 e: 0.21960 f: 0.21960 g: 0.21960 h: 0.21960 a: 0.91563 b: 0.40943 c: 0.40943 d: 0.40943 e: 0.21960 f: 0.21960 g: 0.21960 h: 0.21960 a: 0.91563 b: 0.40943 c: 0.40943 d: 0.40943 e: 0.21960 f: 0.21960 g: 0.21960 h: 0.21960 a: 0.91563 b: 0.40943 c: 0.40943 d: 0.40943 e: 0.21960 f: 0.21960 g: 0.21960 h: 0.21960

Average pagerank = 0.3778

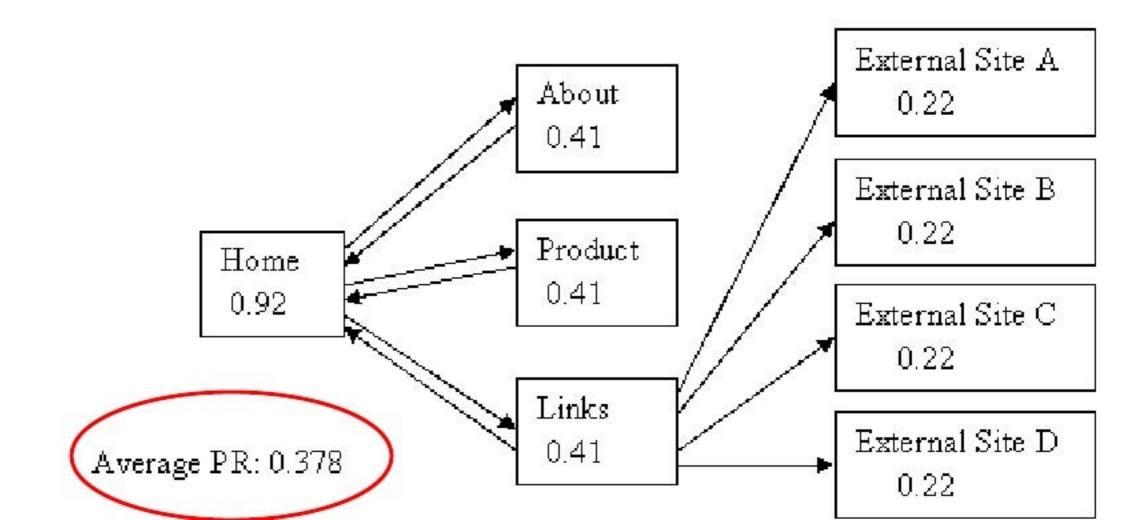


☐ The home page has the highest PR —after all, it has the most incoming



But what's happened to the average?

It's only 0.378!!!

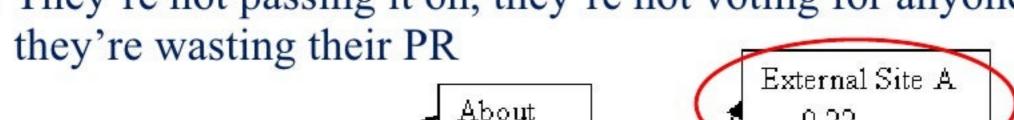


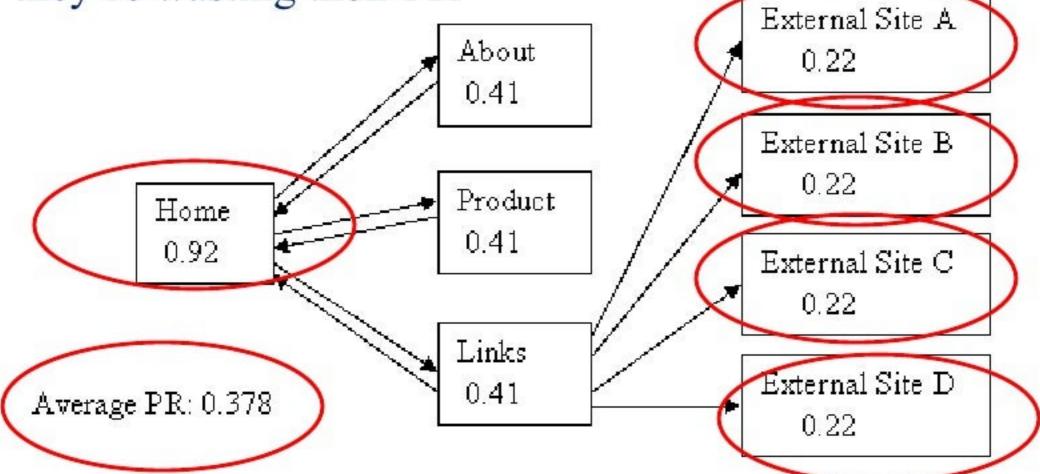
Well, Everything is just fine!!

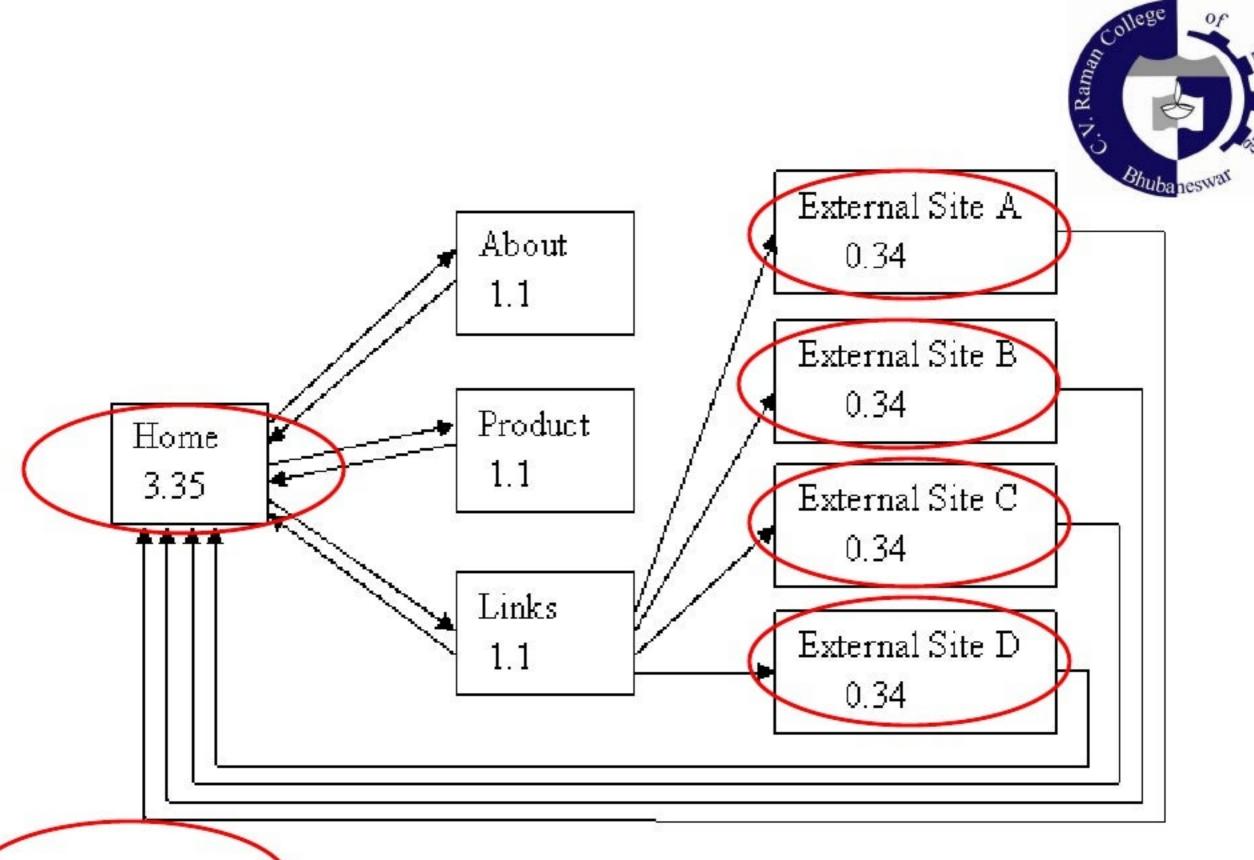
Lets, take a look at the "external site" pages - What's happening to their PageRank?

Bhubaneswar

☐ They're not passing it on, they're not voting for anyone,



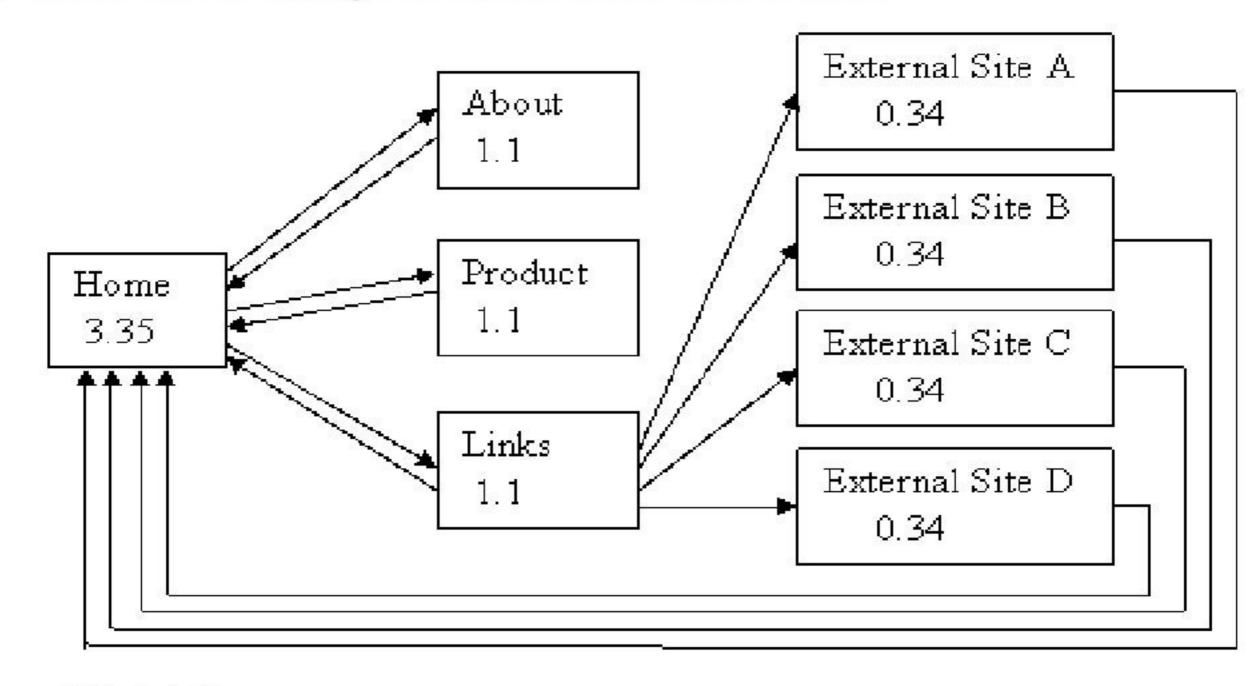




Average PR: 1.000

- ☐ That's better It does work after all!
- **And look at the PR of our home page!**
- All those incoming links sure make a difference



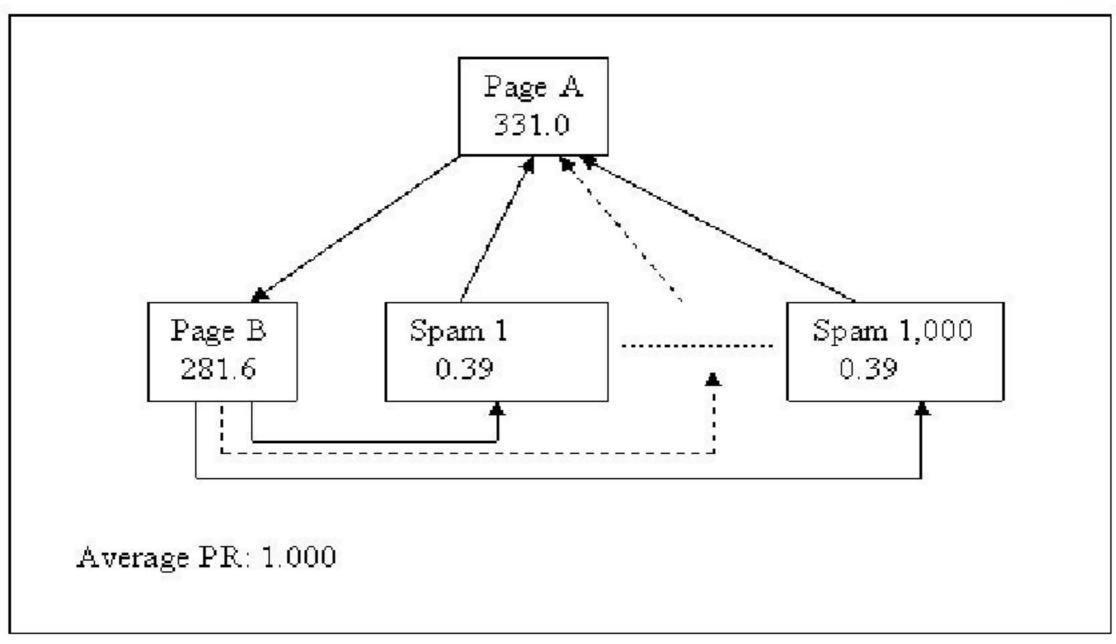


Average PR: 1,000



Just as an experiment, let's see if we can get 1,000 pages pointing to our home page, but only have one link leaving it...





Yes, those spam pages are pretty worthless but they sure add up!



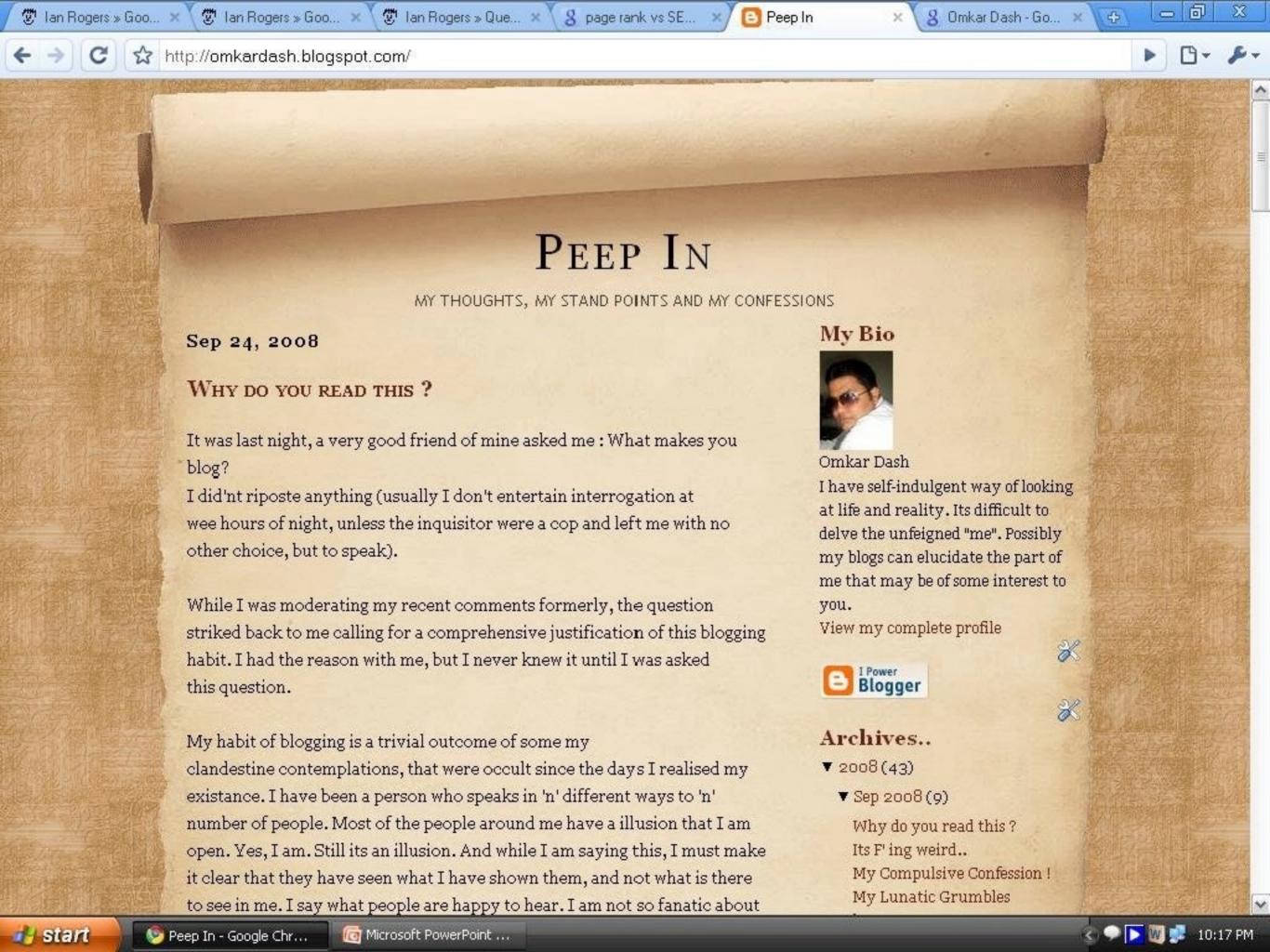
OBSERVATION

- Regardless, the number of pages, average PR will always be
 1.0 at best.
- But a hierarchical layout can strongly concentrate votes, and therefore pump the PR into the home page!



- It is a technique used by some disreputable sites (mostly adult content sites).
- But its not advisable anyway !!
- If Google's robots decide you're doing this..





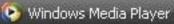




Project 10^100: Tell Google how you want to change the world.

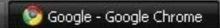
Google.co.in offered in: Hindi Bengali Telugu Marathi Tamil Gujarati Kannada Malayalam Punjabi

Advertising Programs - About Google - Go to Google.com @2008 - Privacy













I am Omkar Dash from CVRCE (I.T), Bhubaneswar. I was one of the lucky 130 students who got through campus recruitment procedure of TCS with success. ... www.chetanasinterview.com/articles/3637/1/TCS-INTERVIEW---15-MAR-2008---BHUBANESWAR/Page1.html - 64k - Cached - Similar pages

Omkar - MvBlogLog

Omkar, Omkar FOAF Omkar hCard Omkar vCard · Omkar Omkar Dash, 21, Male Bhubaneswar. Yahoo Yahoo dash.omkar. Member Since: Aug '08 Last Login: Online Now ... www.mybloglog.com/buzz/members/mybloglog39f5c3893b4933b0bca2/ - 48k -Cached - Similar pages

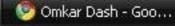
Names Directory: Omkar

Common surnames for Omkar: Omkar Adkar Omkar Agrahari Omkar Ambardekar Omkar Ambulkar Omkar Anand Omkar Bach Omkar Babu Omkar Bhadcayla Omkar Bhadwat

















Names Directory: **Omkar**

Common surnames for Omkar: Omkar Adkar Omkar Agrahari Omkar Ambardekar Omkar Ambulkar Omkar Anand Omkar Dash Omkar Babu Omkar Bhadsavle Omkar Bhagwat ... www.namesdir.com/F_0mkar - 6k - Cached - Similar pages

dash omkar - eSnips Profile - 2 visits - 28 Apr

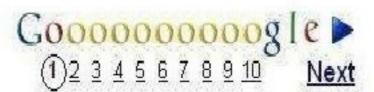
dash.omkar. Lives in: Joined eSnips on Nov. 3 2007. Personal details: Full name: dash.omkar; Age: 30; Total folder visitors: 16; Most popular folder: www.esnips.com/user/dashomkar - 20k - Cached - Similar pages

::Student transfer :: Student transfer from 1.4,2008 on wards Sl ...

1545, OMKAR DASH, SHISHIR KUMAR DASH, 5654, 1, 22652, 15.04.2008. 1546, KUMARI. SUHASREE CHANDA, SUBIR CHANDA, 4980, IV, 22653, 22.04.2008 ... www.kvnazira.com/uploads/april.htm - 17k - Cached - Similar pages

(Peep In)10 visits - 22 Sep

Omkar Dash: I have self-indulgent way of looking at life and reality. Its difficult to delve the unfeigned "me". Possibly my blogs can elucidate the part of ... omkardash blogspot.com/ - 63k - Cached - Similar pages



Omkar Dash

Search

Search within results | Language Tools | Search Tips | Dissatisfied? Help us improve

Google Home - Advertising Programs - Privacy - About Google











A BETTER SERP

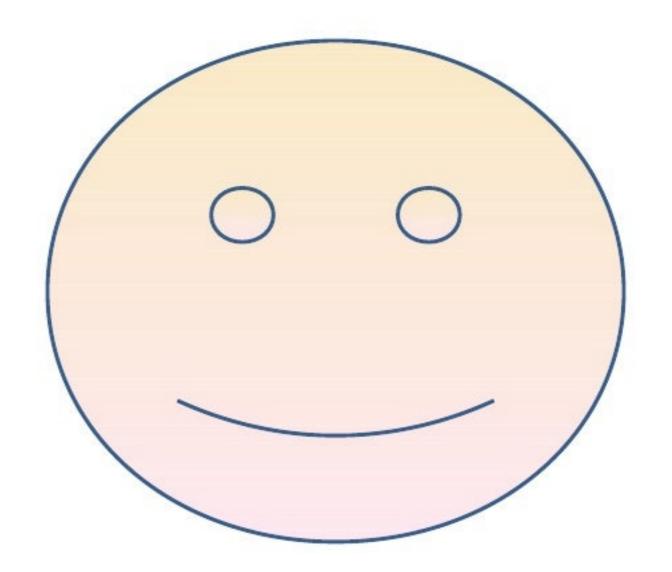


- Meta tags should be properly used
- Contents should be unique and high quality
- Back links to your pages but that should be relevant

REFERENCES



- □ Sergey Brin and Lawrence Page, "Anatomy of a Large-Scale
 Hyper textual Web Search Engine"
- http://www.ianrogers.net/google-page-rank
- PageRank: http://en.wikipedia.org/
- www.google.stanford.edu



Thank You

Queries Please!

