

Broadening (cont.)



- ◆ It is also helpful because many search engines also use measures of ***similarity*** among keywords in order to give added ***bonuses***

Example

- ◆ If we look for Disney and within a web page there is Winnie the Pooh, the score of our page can have a little bonus



What is missing...?

- ◆ Beyond the techniques of textual spamdex, there is also another problem even more fundamental:
- ◆ ***What keywords*** to choose!



How can we do it?

- ◆ We need to know what users want



Smart methogs

- ◆ Google Ads Keyword Planner
(a.k.a. Keyword Tool *gk (\$\$...))

Select an active account

Surfside PPC 

Discover new keywords



START WITH KEYWORDS

START WITH A WEBSITE

Enter products or services closely related to your business



beach decor



coastal decor



+ Add another keyword



Include brand names in results

Enter a domain to use as a filter



https://

Try not to be too specific or general. For example, "meal delivery" is better than "meals" for a food delivery business






[Learn more](#)

Using your site will filter out services, products, or brands that you don't offer

GET RESULTS

18:22

Want to get up to 2x more keywords instead? [Subscribe to Keyword Tool Pro now!](#)

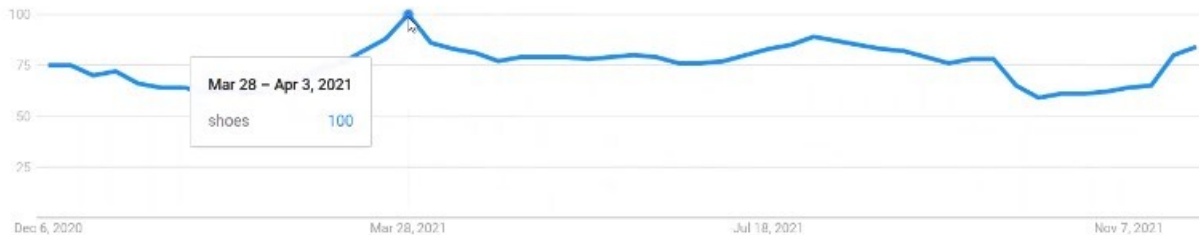
<input type="checkbox"/> ▼ Keywords ?	Search Volume ?	Trend ?	CPC (USD) ?	Competition ?
<input type="checkbox"/> shoes near me	135,000	 +123%	\$0.75	98 (High)
<input type="checkbox"/> shoes for crews	135,000	 +22%	\$0.46	96 (High)
<input type="checkbox"/> shoes for men	550,000	 -45%	\$1.52	100 (High)
<input type="checkbox"/> shoes for women	3,350,000	 -45%	\$1.22	99 (High)
<input type="checkbox"/> shoes carnival	1,500,000	 0%	\$0.07	26 (Low)
<input type="checkbox"/> shoes on sale				
<input type="checkbox"/> shoes stores near me				
<input type="checkbox"/> shoes black friday				
<input type="checkbox"/> shoes at walmart				
<input type="checkbox"/> shoes adidas				
<input type="checkbox"/> shoes and more				
<input type="checkbox"/> shoes amazon				
<input type="checkbox"/> shoes air force				
<input type="checkbox"/> shoes and brews				
<input type="checkbox"/> shoes at target				

Do You Need This Information?
Keyword Tool Pro Provides Search Volume Data For English
Keywords. Try It Now!



● shoes

United States, Past 12 months



Interest by subregion ?

Subregion ▾ ⬇ ⏪ ⏩ ⏴ ⏵



1	Mississippi	100	<div></div>
2	Louisiana	98	<div></div>
3	Alabama	97	<div></div>
4	New York	93	<div></div>

Problems of term spamming

- ◆ We said it at the very beginning when talking about body spam:
- ◆ Text spamming generally changes the content of the pages, so users are affected (likely in a bad way!!)

Think...

- ◆ We «power up» a page by inserting keywords, piece of other pages etc etc
- ◆ The page gets higher in search engines
→ Users click and land on the modified page → content corrupted or not so relevant → (usability) he gets angry!!



Content hiding (examples)

◆ `<body background="white">`

...

``

text to disappear...

``

...

`</body>`



Redirection



- ◆ Also called "302 technique"
- ◆ Easy way:
- ◆ **`<meta http-equiv="refresh" content="0;url=pippo.html">`**

Redirection (cont.)



- ◆ More effective: use javascript (!)
- ◆ Because search engines have a hard time to understand Javascript (and in many cases they just ignore it!)
- ◆ Example:

```
<script language="javascript"> <!--  
location.replace("pippo.html")  
--></script>
```

Cloaking...

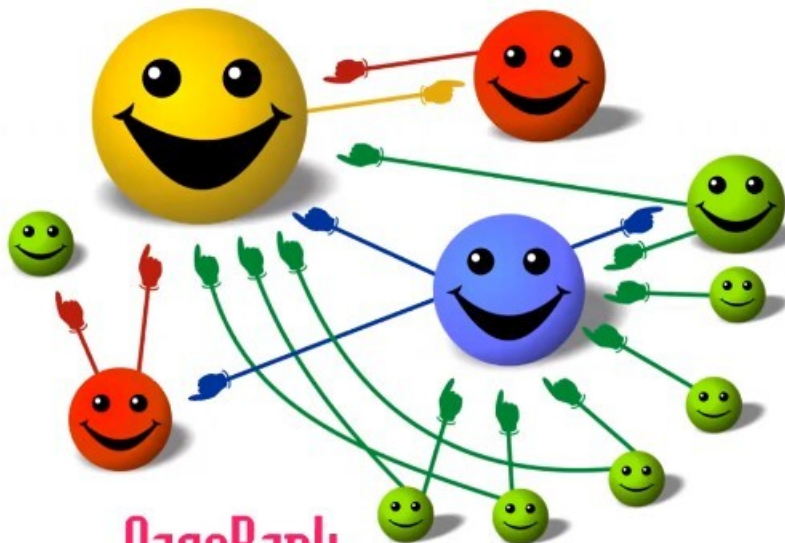


Let's pass now...

- ◆ ... to the other component, the ***hypertextual*** one
- ◆ Complementary to the textual one, and contributes with a good deal of points derived from the ***network topology***



The super-famous Pagerank



PageRank

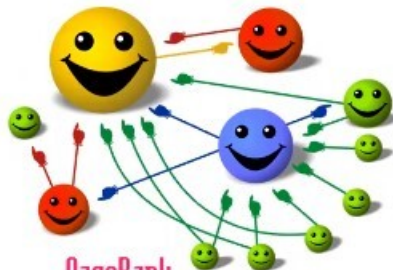
Pagerank in more detail...

◆ (d_w = number of outgoing links)

$$\pi_v = \sum_{(w,v) \in E} \frac{\pi_w}{d_w}$$

Diagram illustrating the PageRank formula. The term π_v is circled. The summation is over edges $(w,v) \in E$, with arrows pointing to w and v . The term $\frac{\pi_w}{d_w}$ is also circled, with an arrow pointing to d_w .

$$\sum_v \pi_v = 1$$



Intuition

$$\pi_v = \sum_{(w,v) \in E} \frac{\pi_w}{d_w}$$

$$\sum_v \pi_v = 1$$



Reformulation...

- ◆ ***Markov chains*** and ***random walks***
(the
«drunkard's walk»)

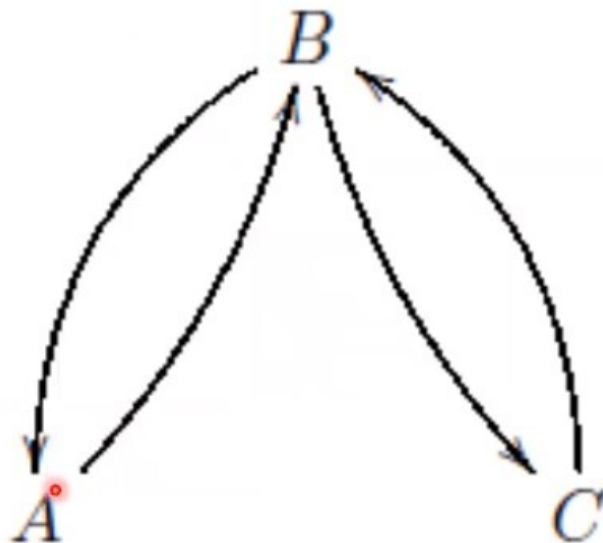


In the web drunkard → monkey
("random surfer")

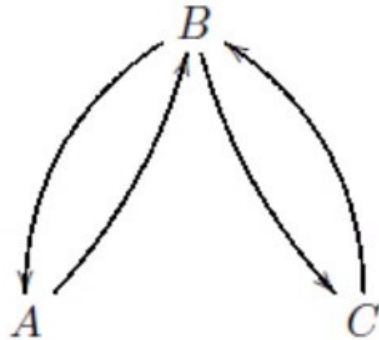


Example

◆ What Pagerank?



Example (cont.)



- ◆ Either we calculate, or we reason by symmetry:
- ◆ $A=C$, B has double flow than A and C, so (if the total liquid is 1):
- ◆ The pagerank of A and C is $\frac{1}{4}$, those of B is $\frac{1}{2}$

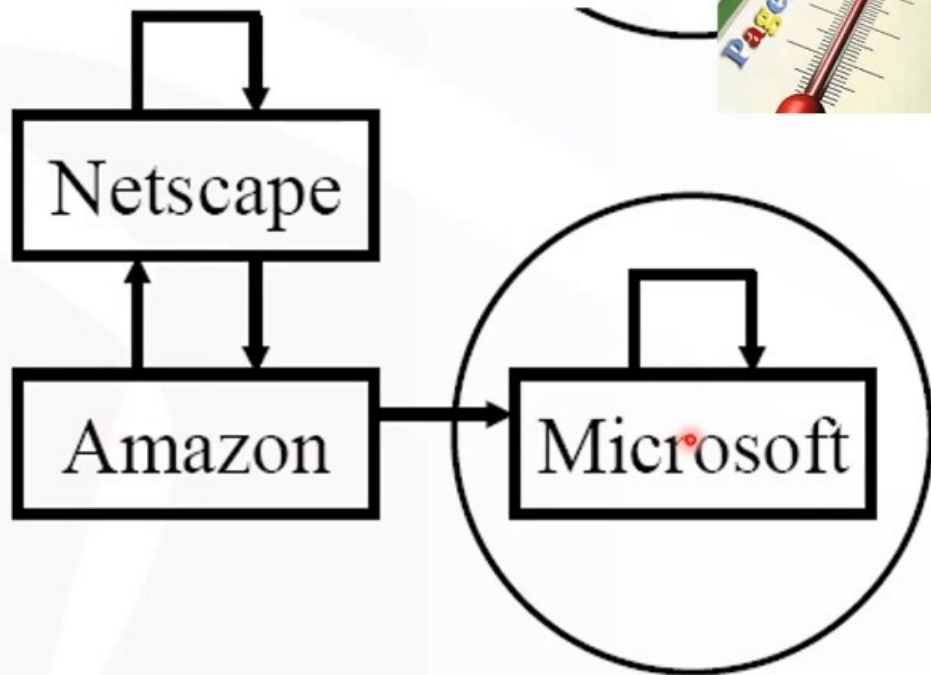


Spider traps...

- ◆ Unfortunately are very frequent
- ◆ For instance just think of an event calendar...



Spider traps... and Pagerank (??!!)



Solutions....??



Let's change....!



- ◆ The formula has been patched
(***teleportation***)!

$$\pi_v = (1 - \epsilon) \left(\sum_{(w,v) \in E} \frac{\pi_w}{d_w} \right) + \frac{\epsilon}{N}$$



Note on elegance... Totalrank

- ◆ Computational cost: almost the same as Pagerank

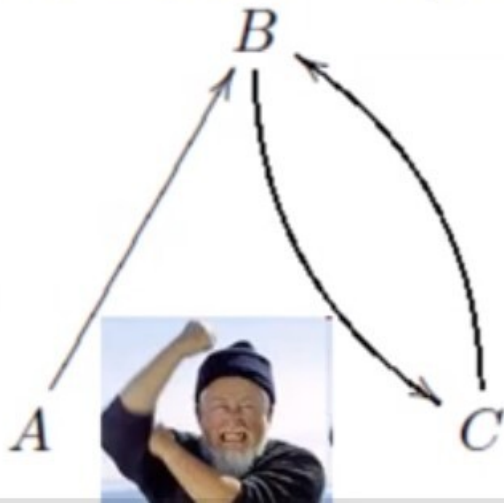
$$\mathbf{T} = \int_0^1 \mathbf{r}(\alpha) d\alpha$$

But, wait...!!



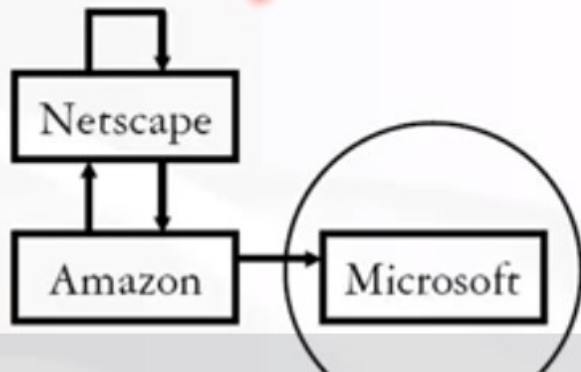
◆ Here comes C, which changes a link!

◆ Pagerank of B and C = $\frac{1}{2}$, A = 0 (!!!)

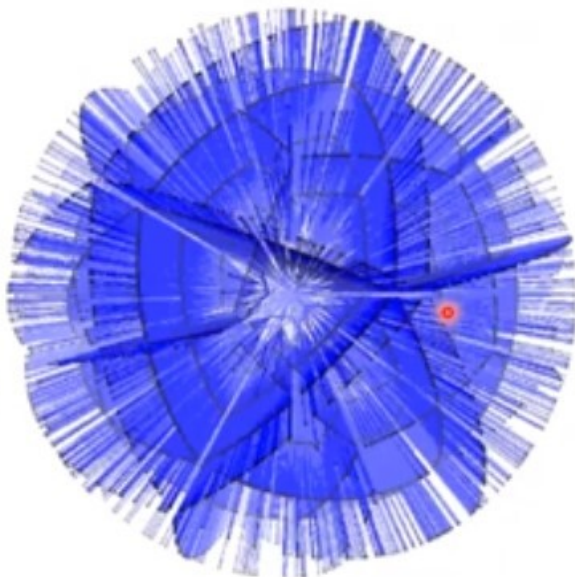
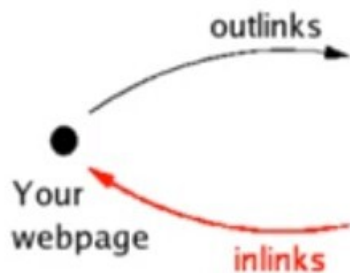


Soluzione?

◆ **A MESS!!**



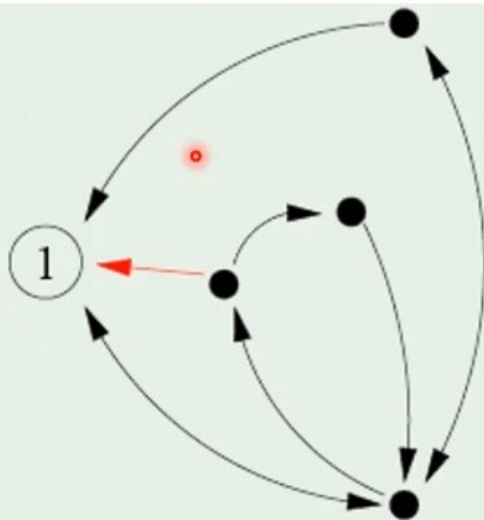
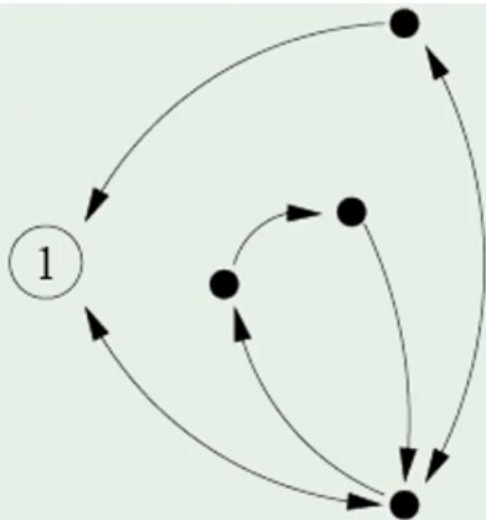
Strategic move: Adding inlinks



The rest of the Web

Example

◆ **0.196** at first, passing to



Infiltration

- ◆ We «infiltrate» in various sites and try to insert links to our site
- ◆ For instance inside directories, blogs, wikis, comment sections etc



Honey pot

◆ Yummy!!



Honey pot (cont.)

- ◆ Create “yummy” content, and then naturally receive incoming links
- ◆ We can do it by smart paste & copy from content of other sites (!)



Link Exchange



Resurrection

◆ When a domain «dies»... catch it!!

