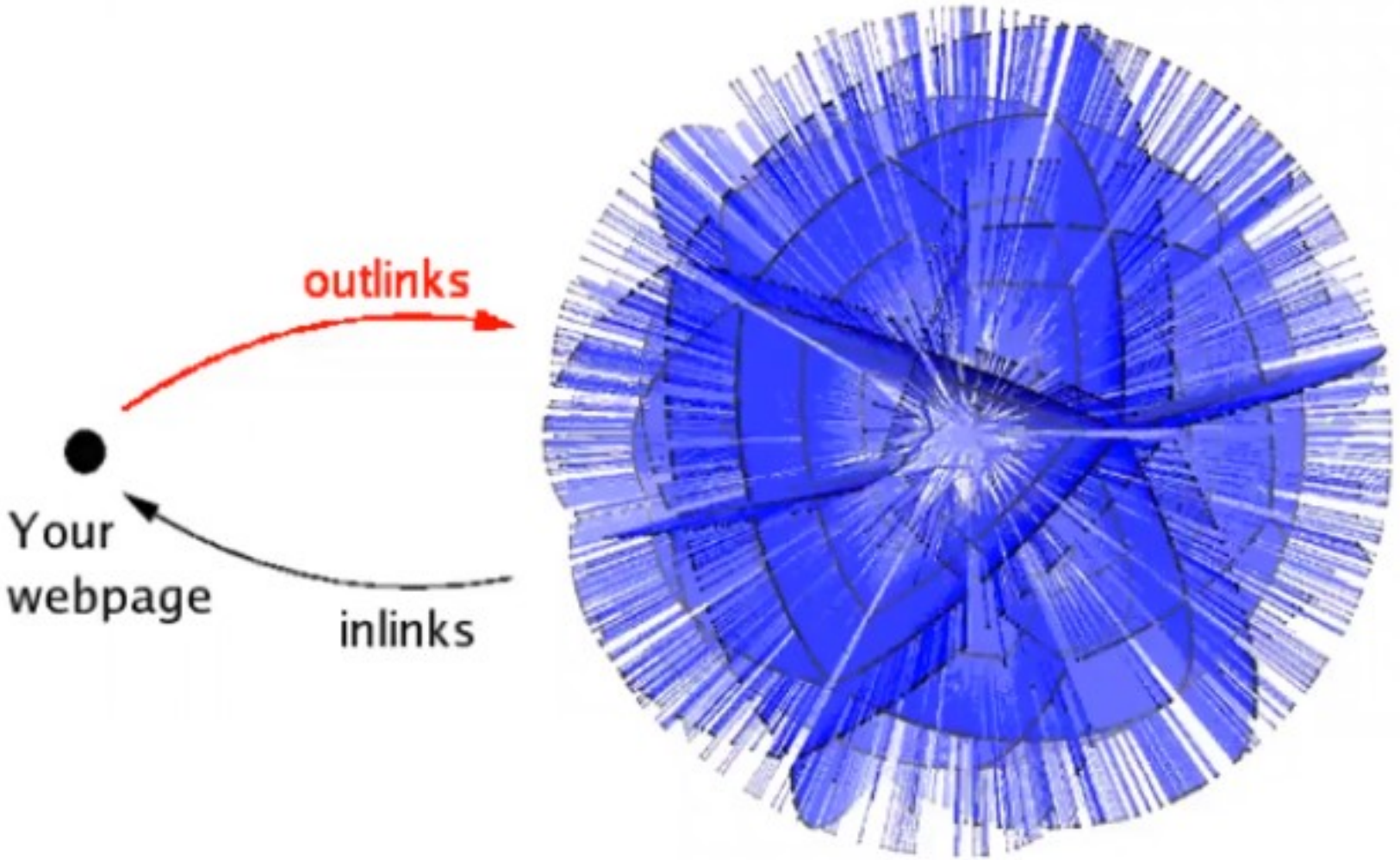
Related aspect…

- "Marketing 2.0"

Let's now see the other direction of links

Strategic move: Adding outlinks



What happens in this case?

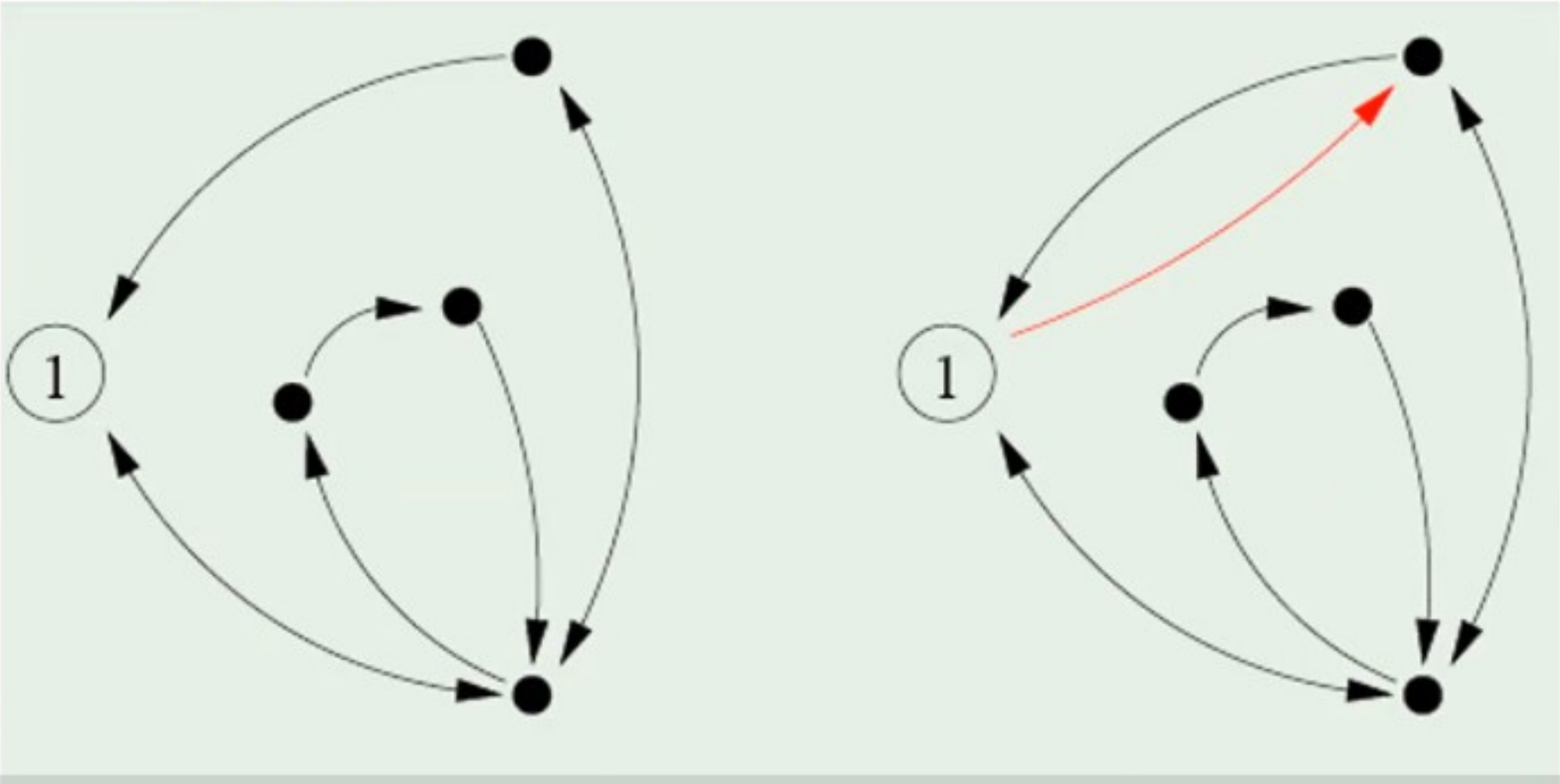
- Just reason by symmetry
- The pagerank gets lower (<=) because some flow goes out, we cannot augment flow with an outgoing channel

Property: solidity

- Property of ranking measure, to be "***solid***": adding outlinks to a page doesn't cause an increase of the hypertextual score
- → The spamdex effort cannot just be ***local***

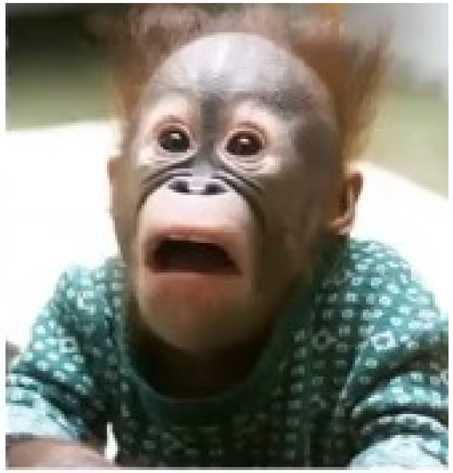
Other example

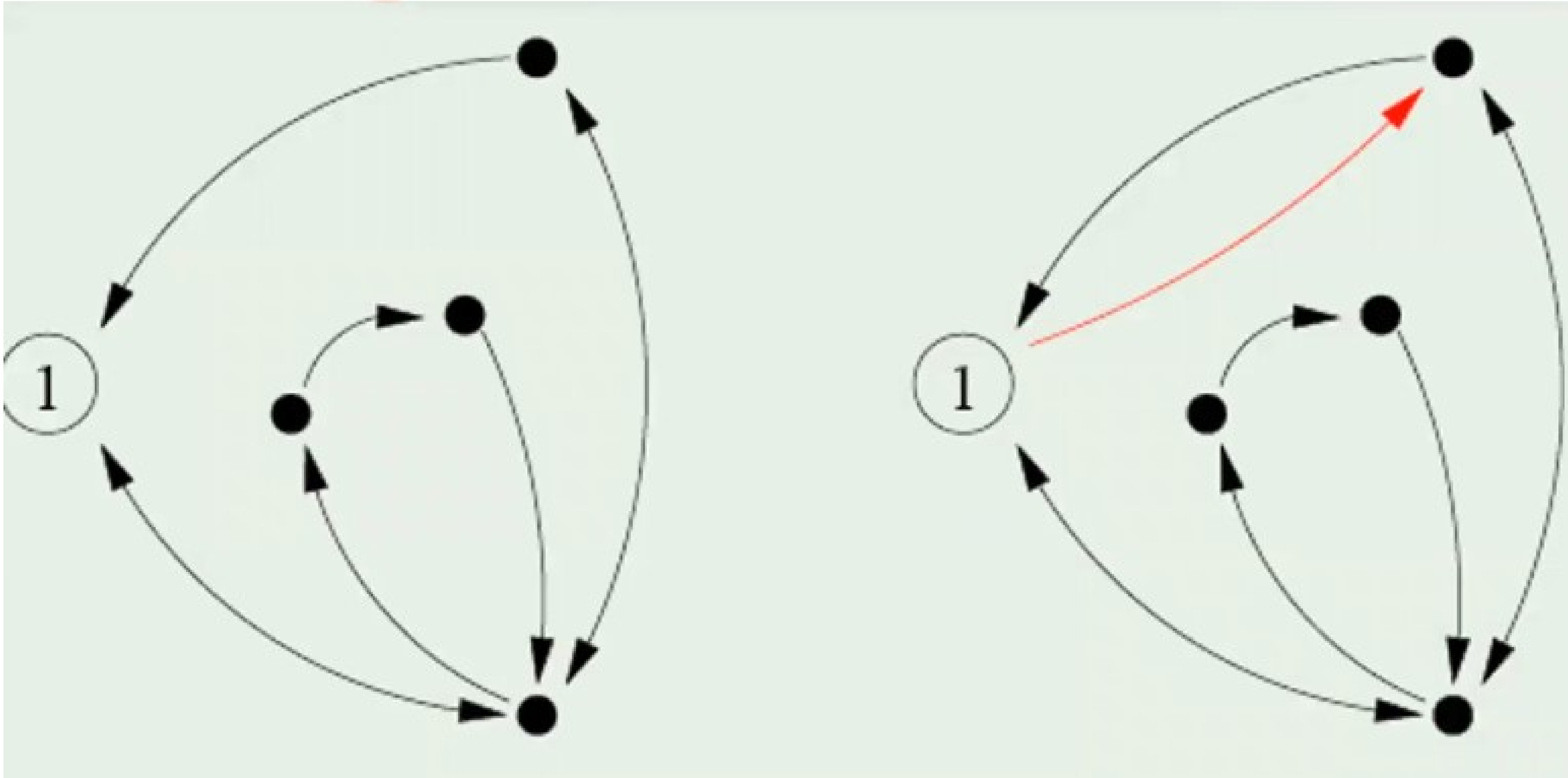
- **0.196** initially, passing to
- **0.211**



0.196 → 0.211

!!

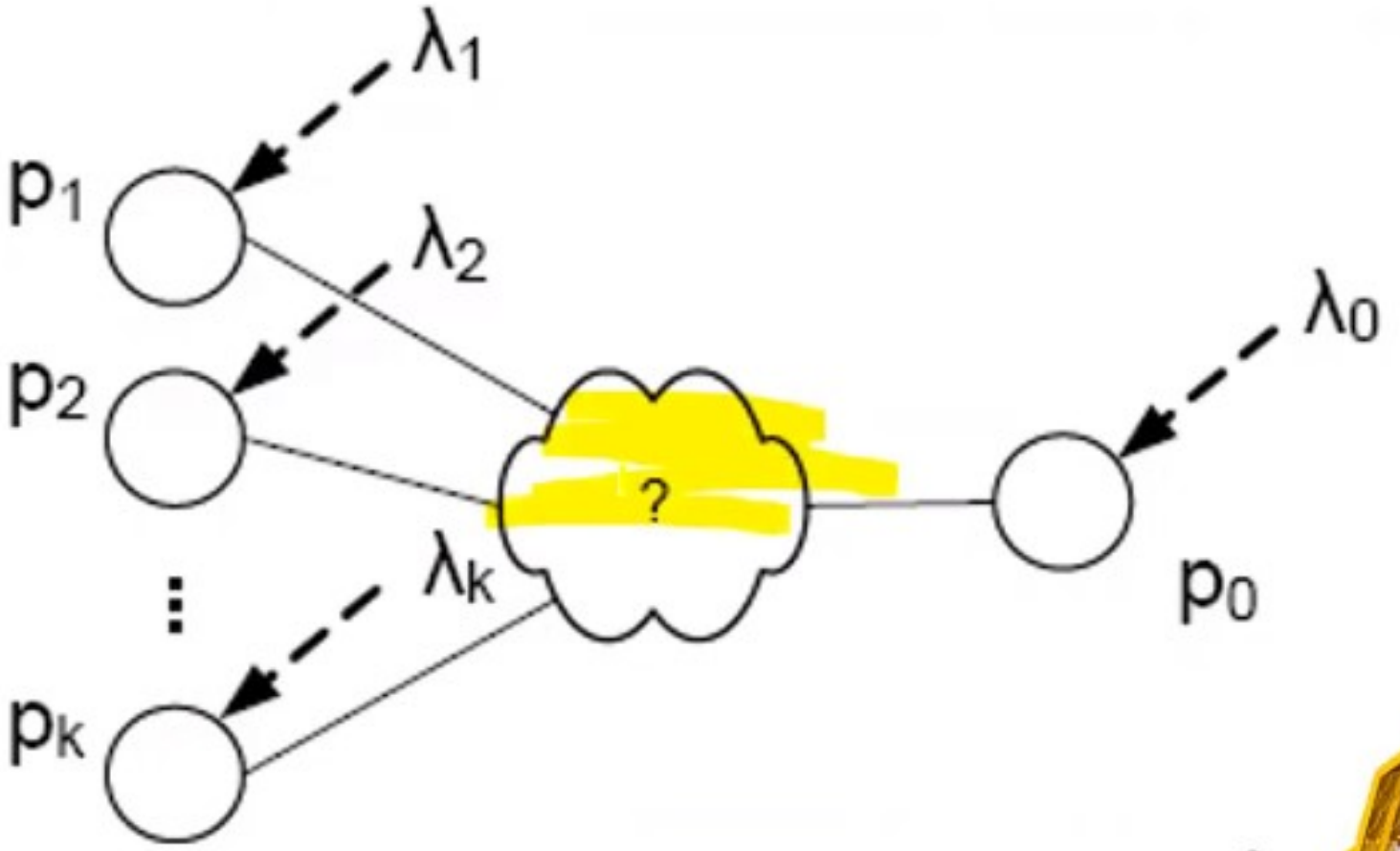




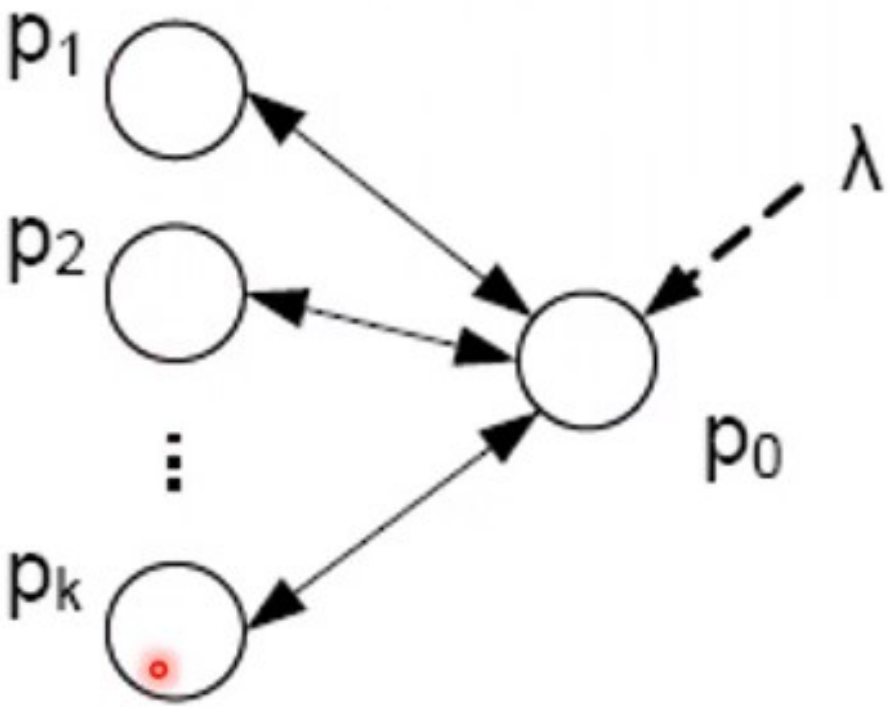
The Spam Farm

- A special link structure devoted to increase the hypertextual score

Structure of a generic farm

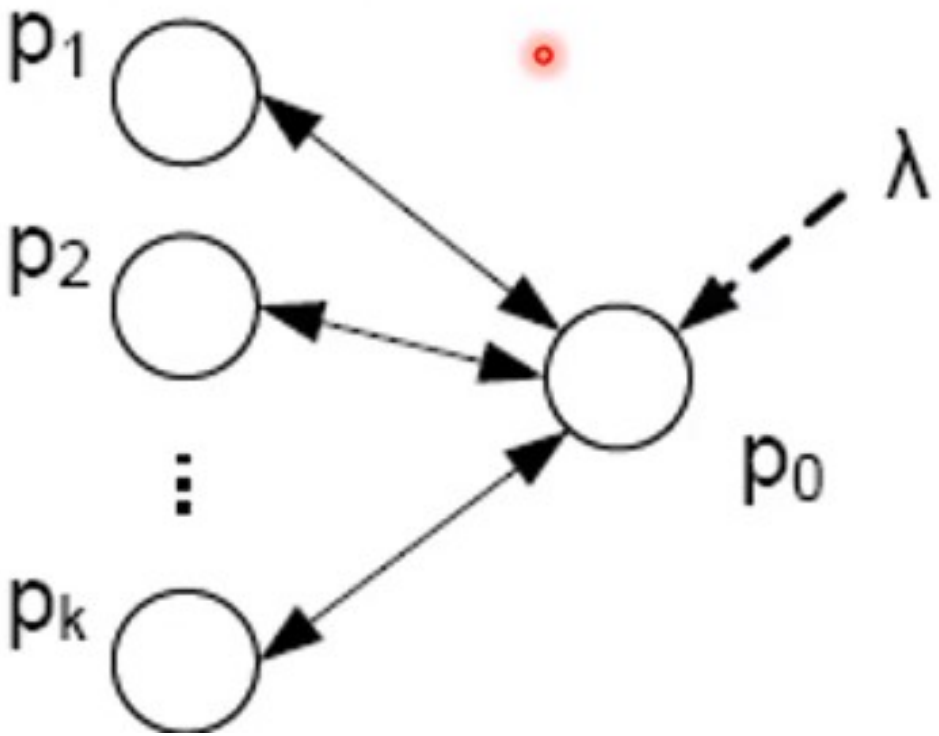


Optimal spam farm with a target page?



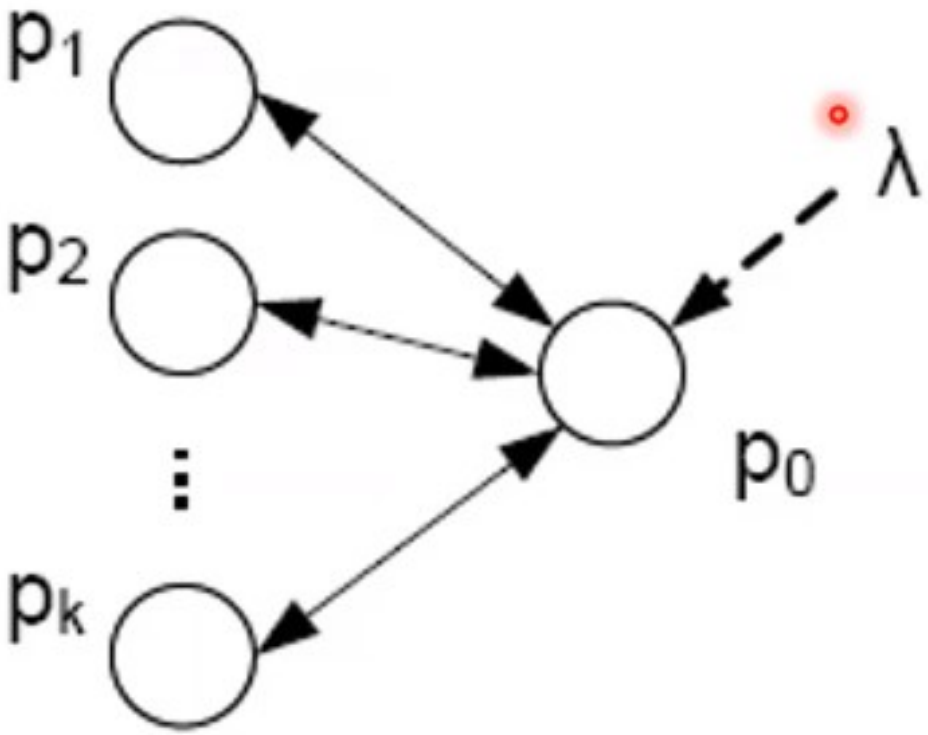
Good properties of this structure

- It uses the ***least possible number of links***, while still keeping…

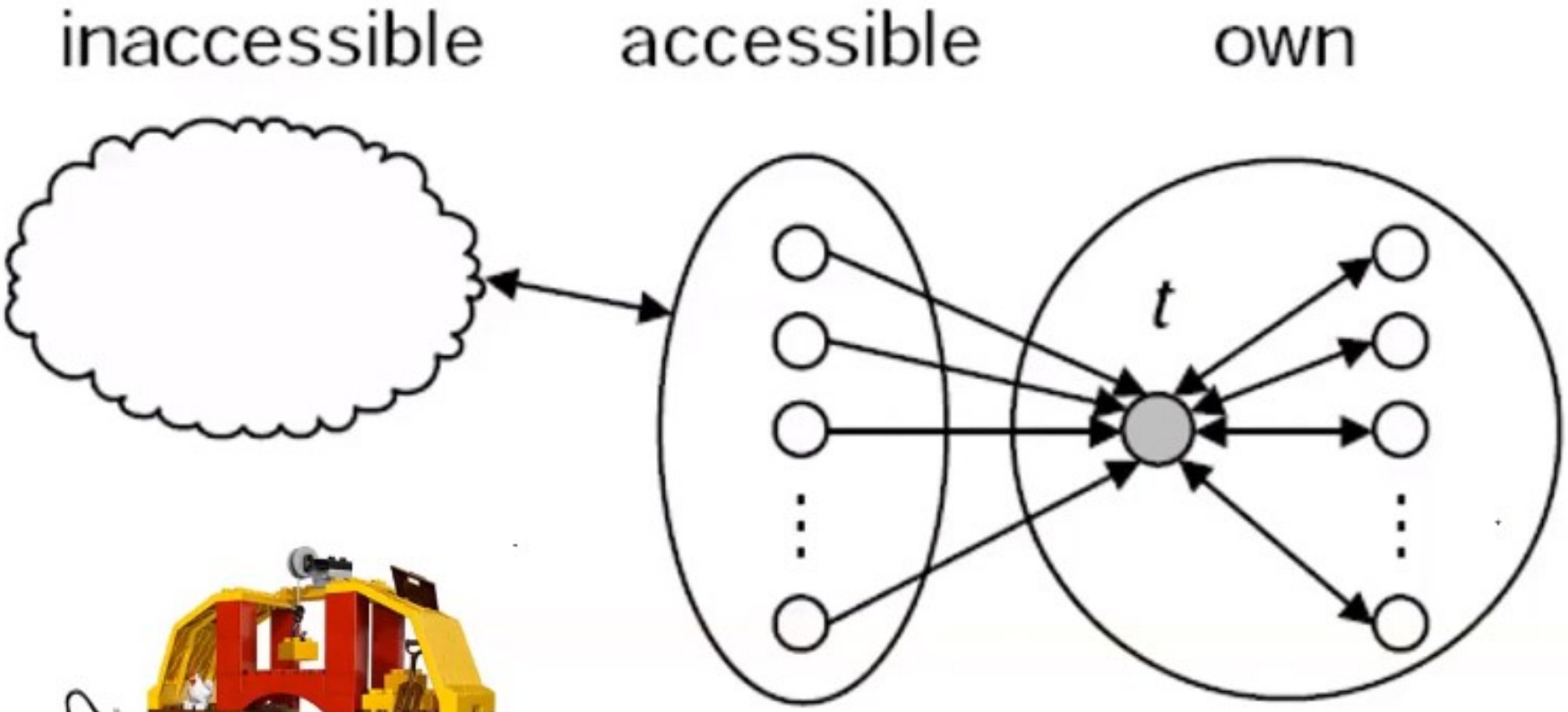


… a very important property

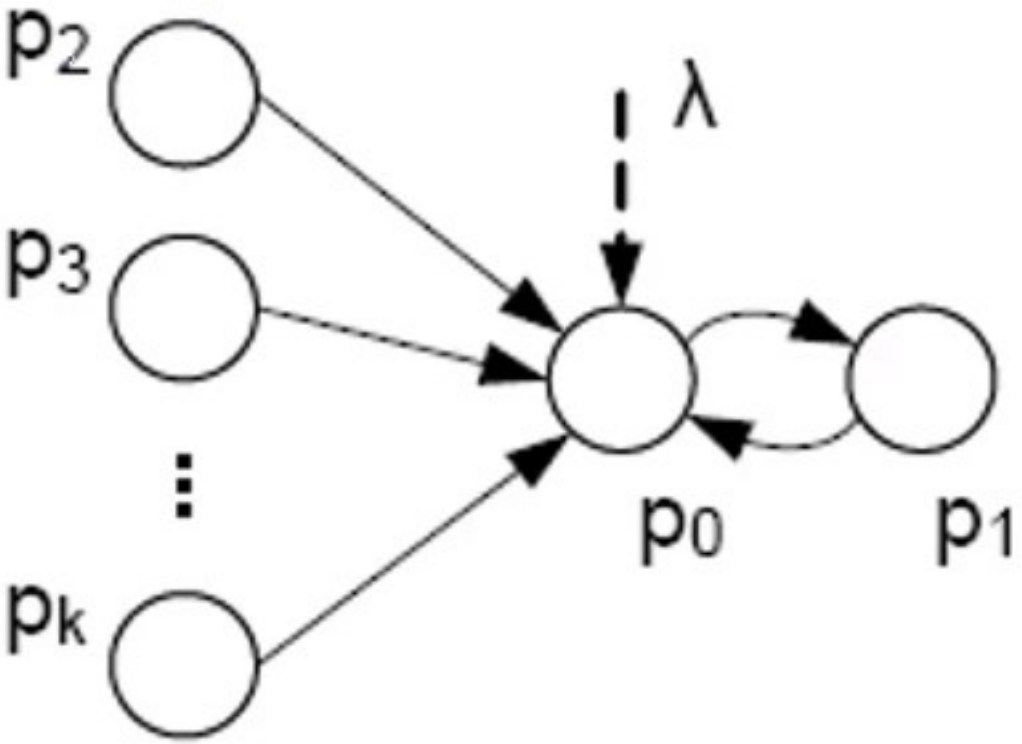
- ***Reachability***



Optimal structure generalized



Another optimal spam farm (sacrificing reachability)



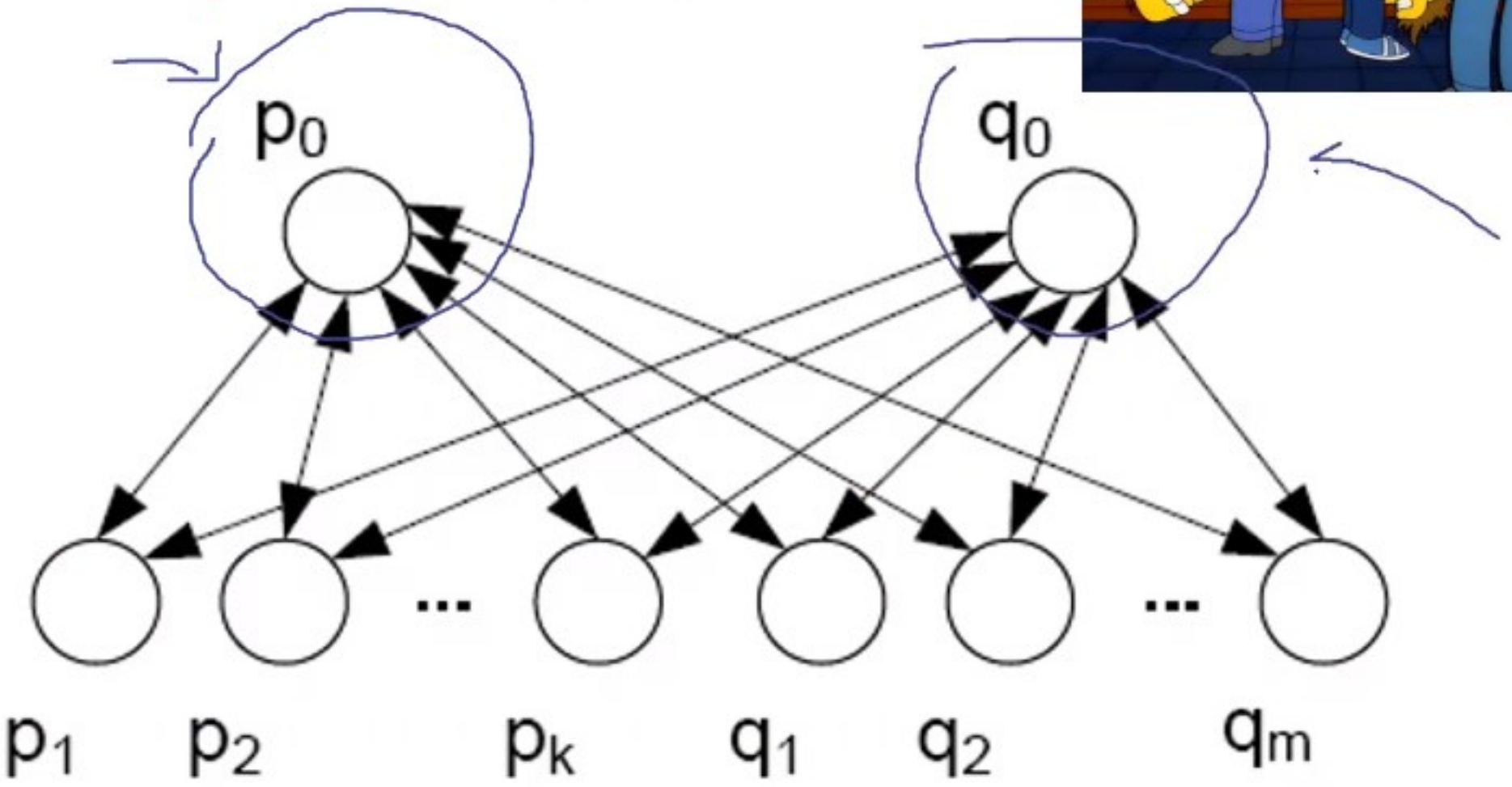
Important aspect

- Are we alone? Cannot we join forces instead with an ***alliance*** ?
- The idea: join our spam farm with someone else's

The Alliance Problem

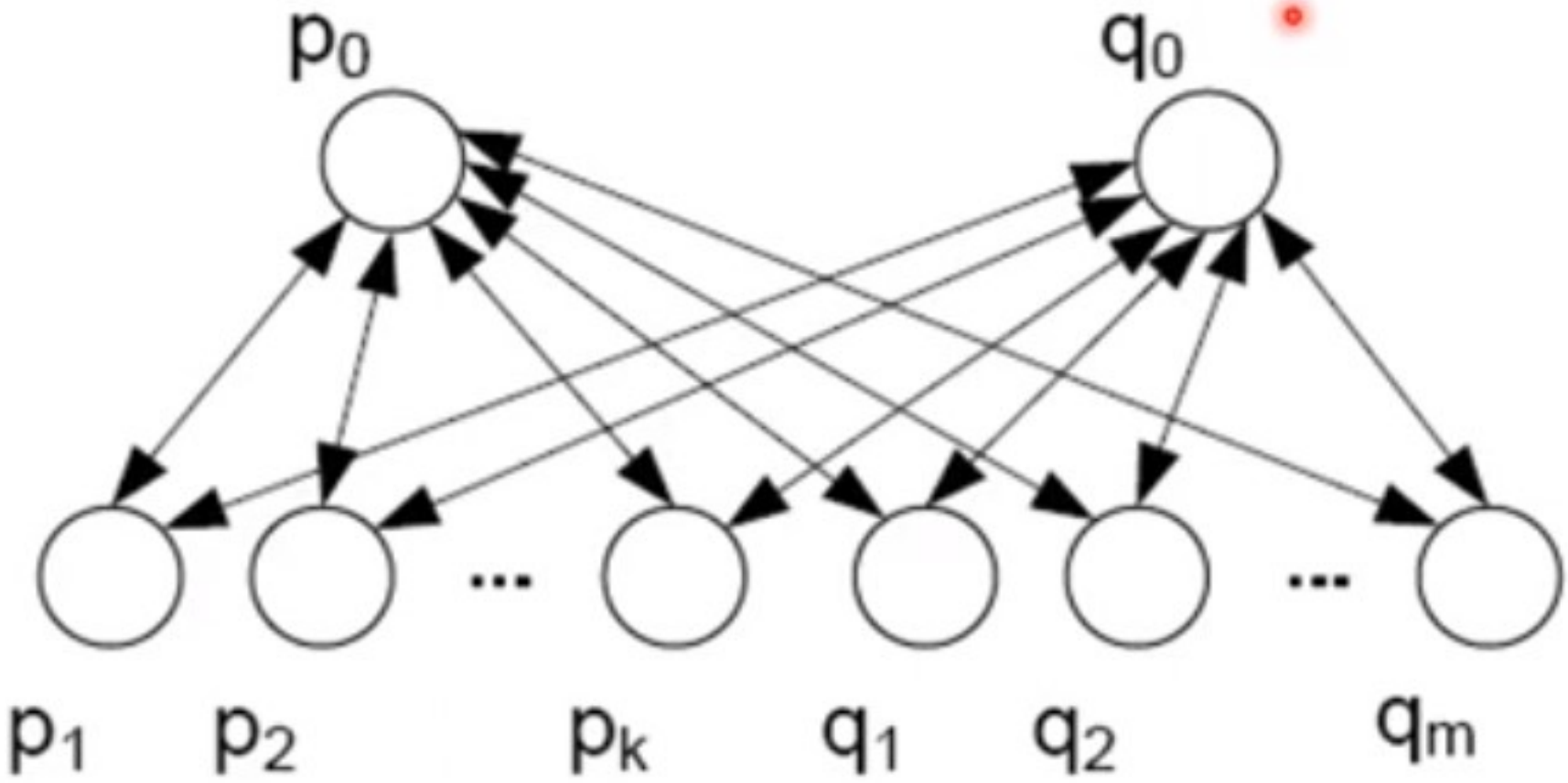
- What are the best ways to do it?

Scheme: «deep» alliance



How much pagerank?

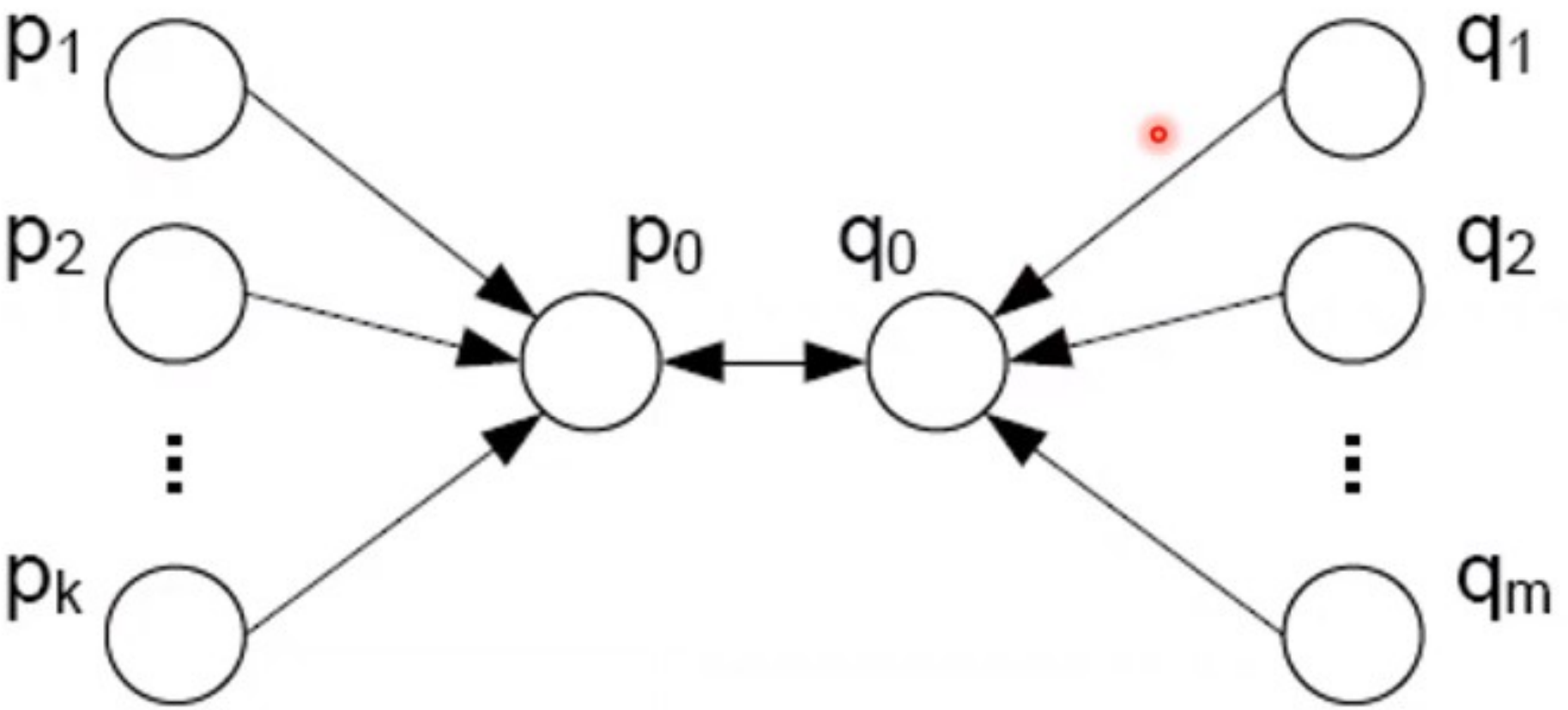
- → the average of the two pageranks!



So…

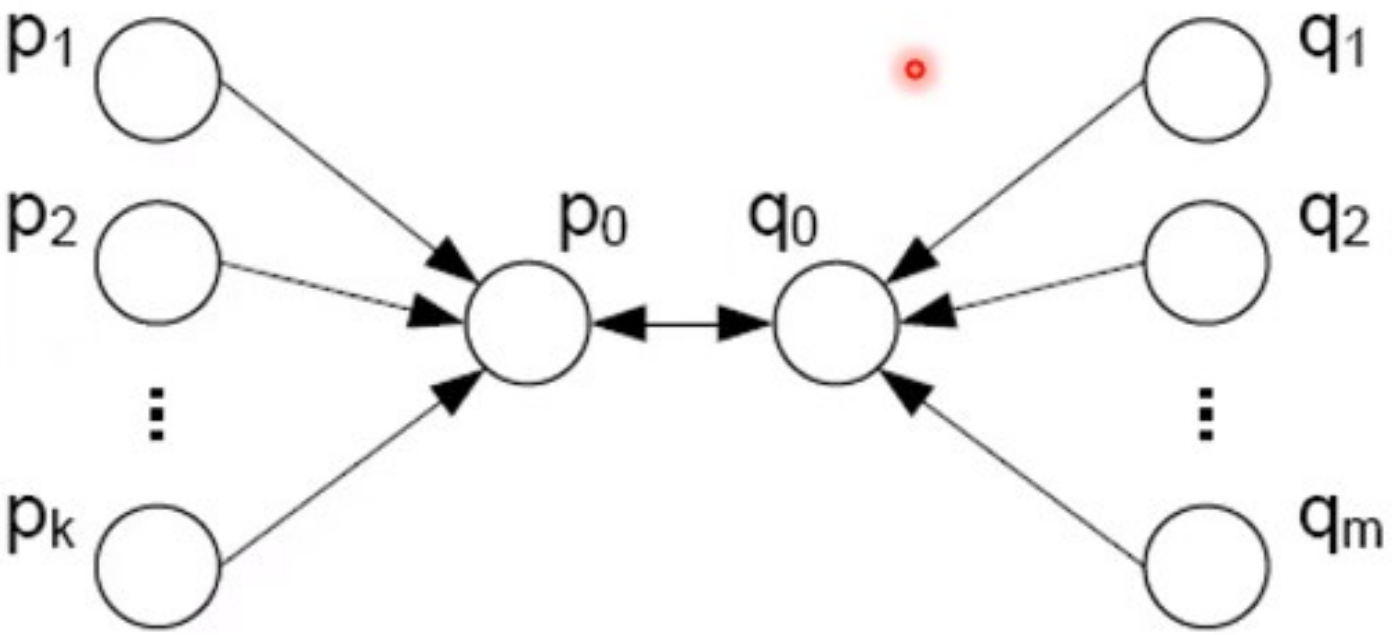
- Useful alliance scheme to share the load and get a more ***robust*** configuration: we do our best and then we share goods and bads

Scheme: «superficial» alliance

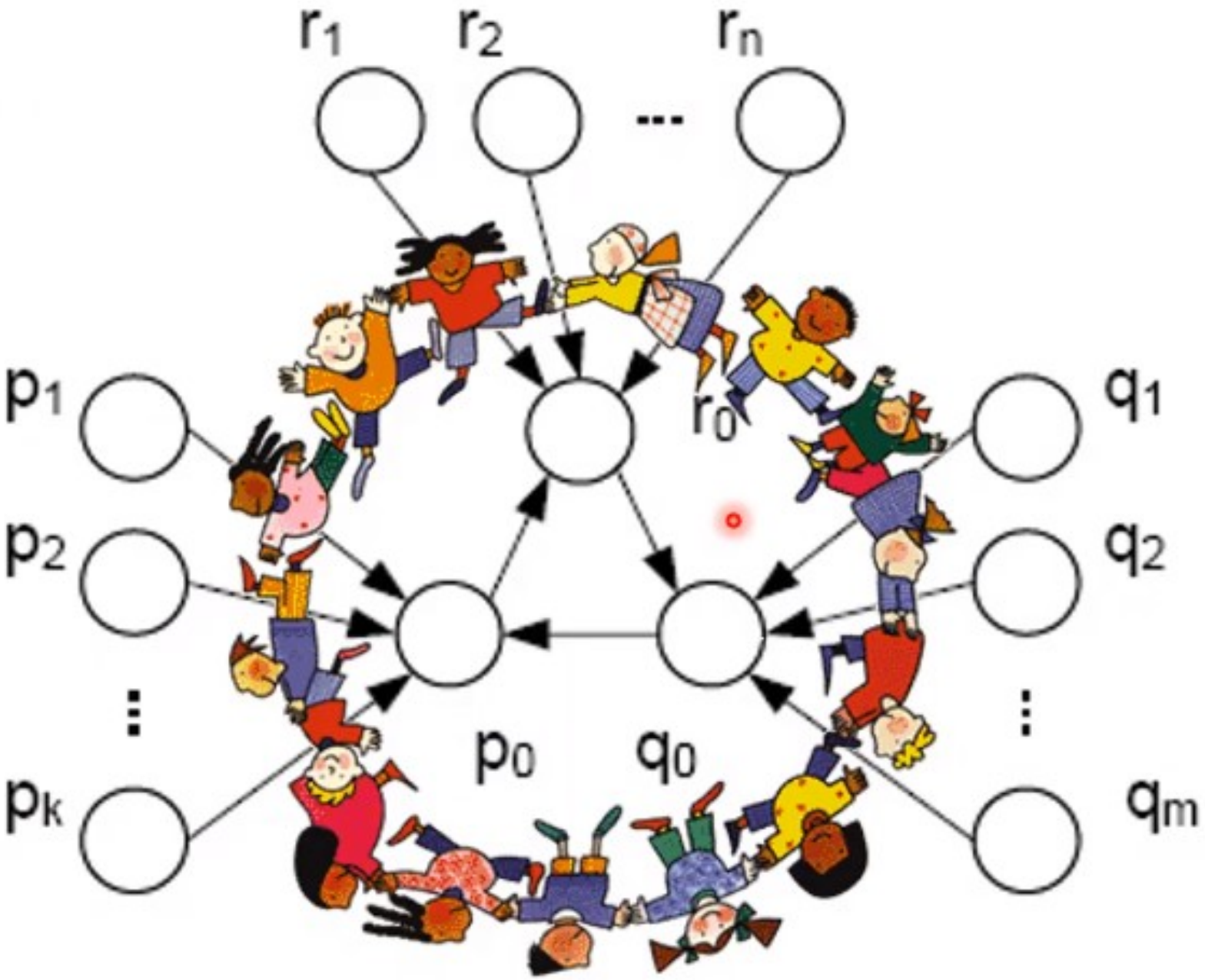


How much pagerank?

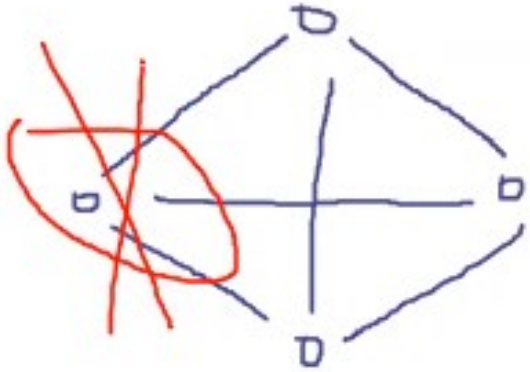
- **MORE THAN THE MAX BETWEEN THE TWO (!!!!)**
  (bonus proportional to k and m)

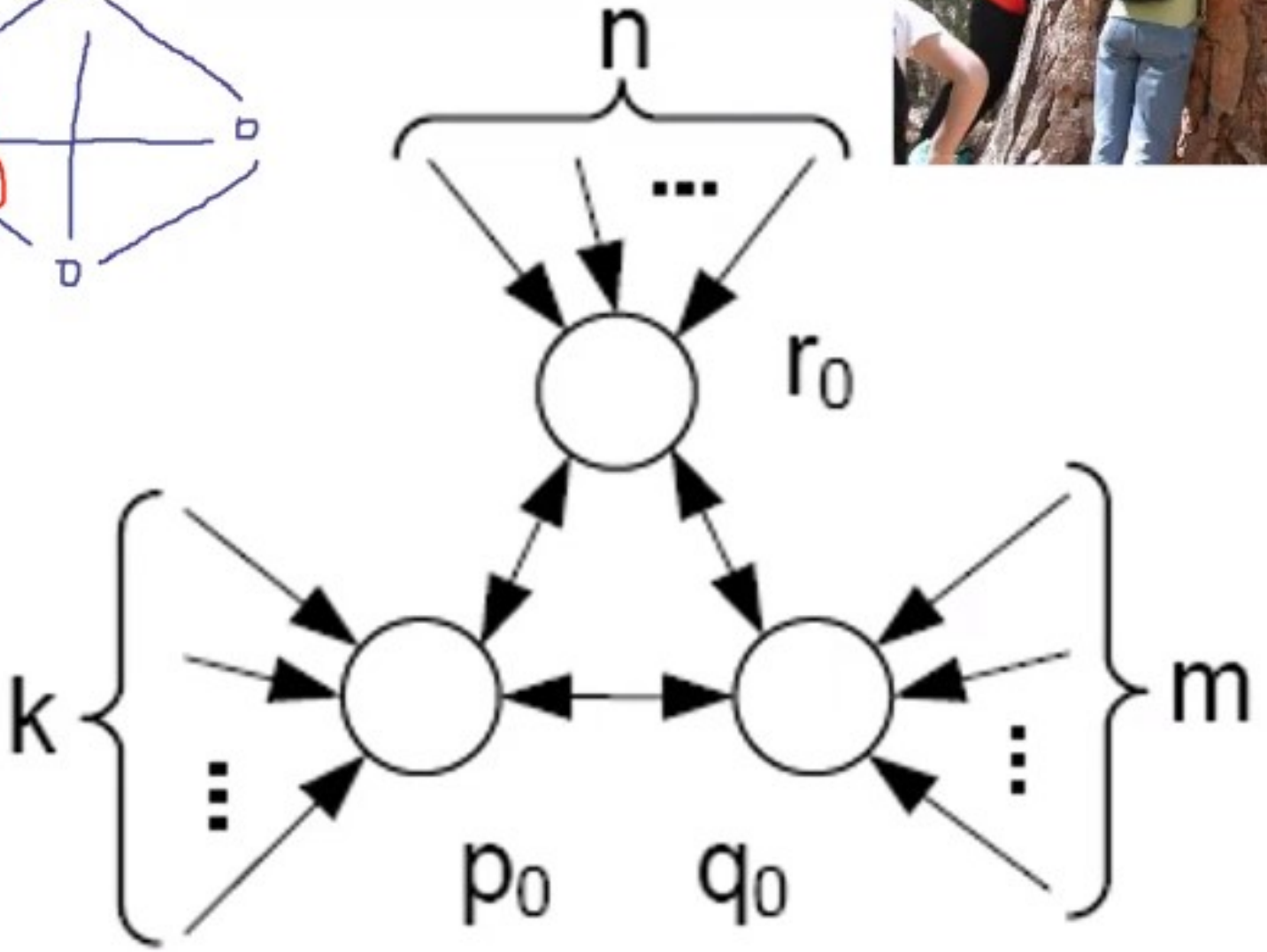


Ring



Complete core



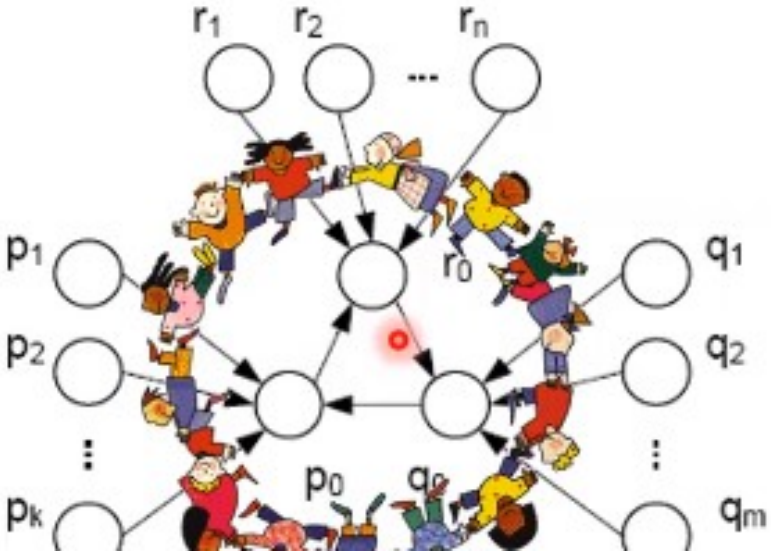


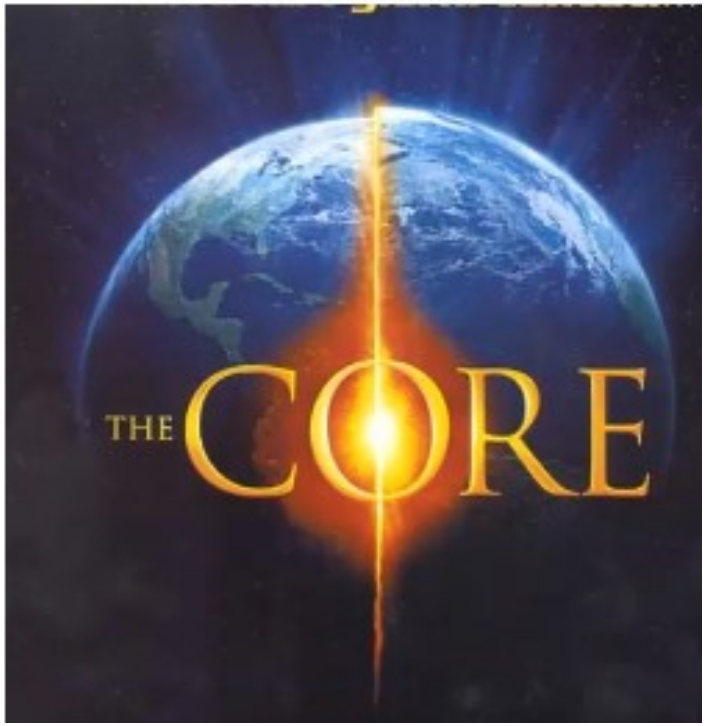
Countermeasures!

- In this case, it doesn't look so difficult: just find the ring / complete core structures among different sites ☺

Other cores?

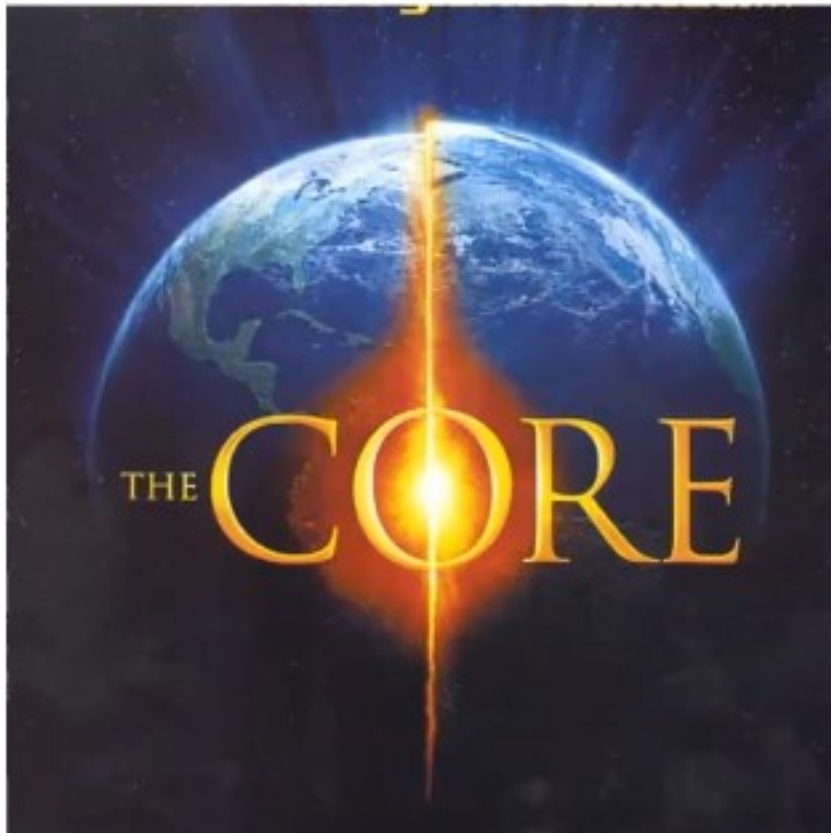
- We just need to have a ***strongly connected*** graph among the target pages (that is to say, from every page I can arrive to any other)





And so…

- … how many strongly connected graphs there can be?
- If they are only a few ones, then the countermeasures will work well…
- So how many for an alliance (size N)?

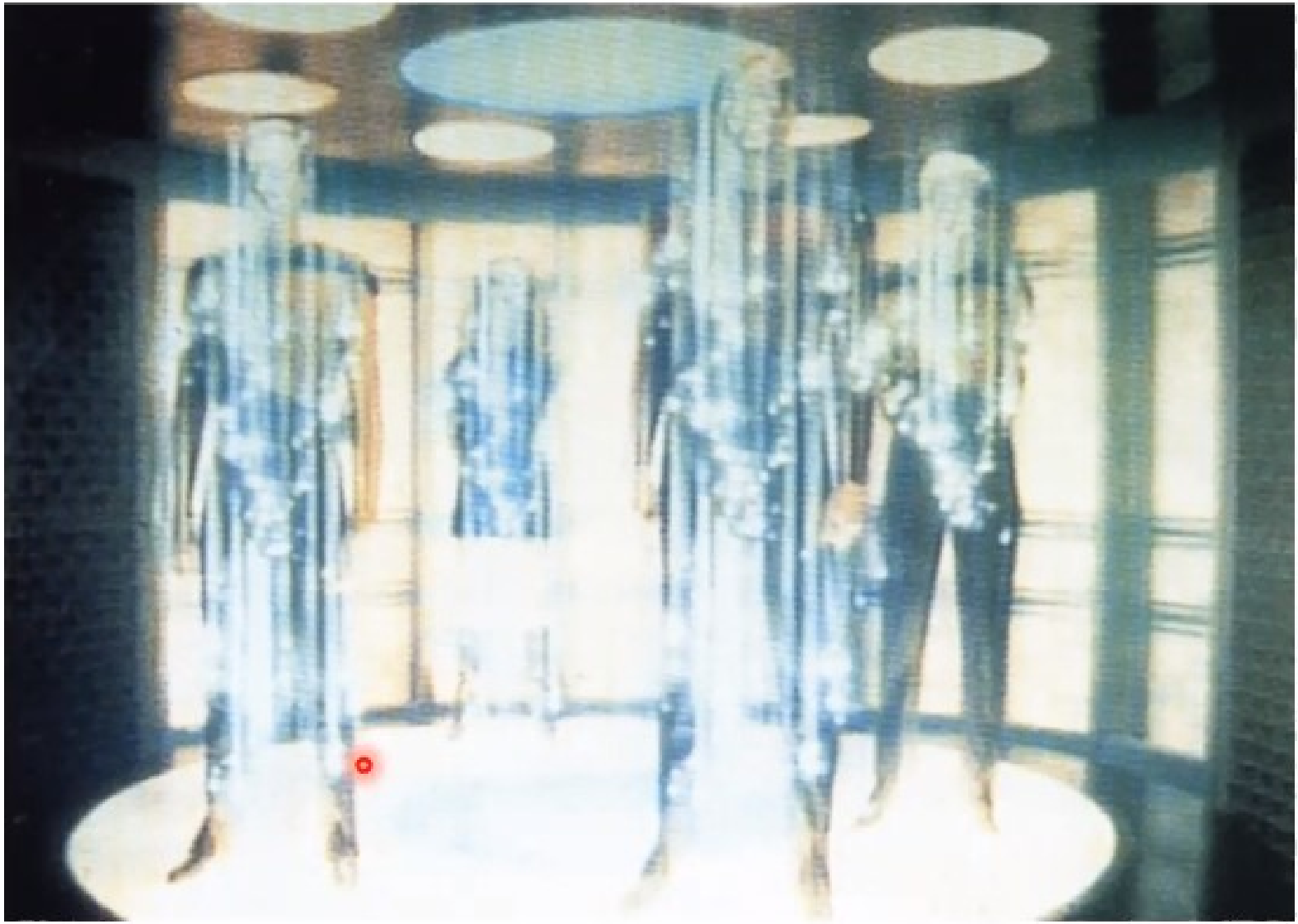


The "A003030 sequence"...

- N=3 : 18
- N=4 : 1606
- N=5 : 565080
- N=6 : 734774776
- N=7 : 3523091615568

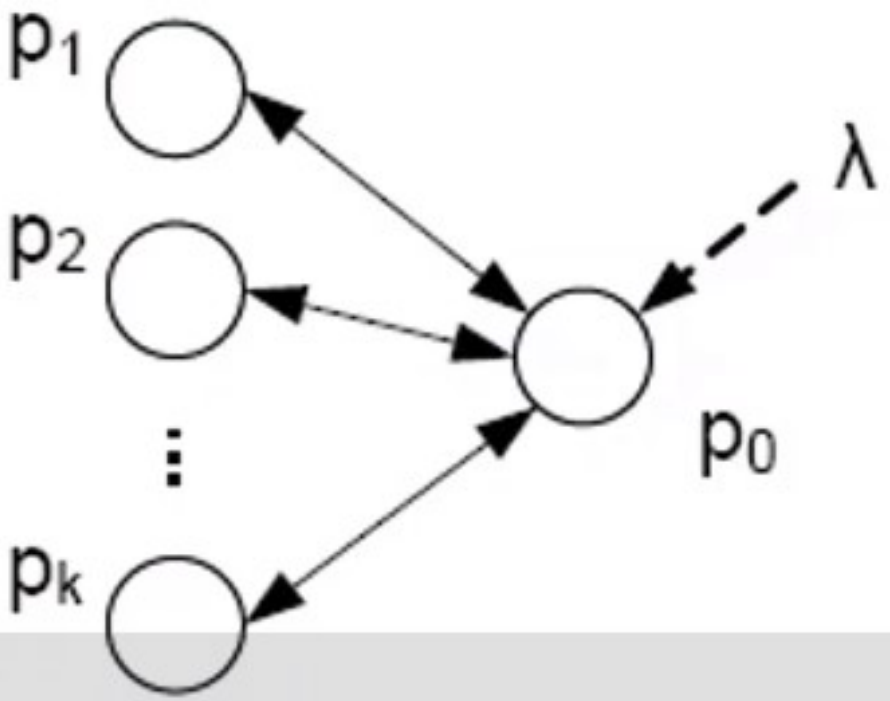
N=8 : 6351920938966417 6
N=9 : 4400410978376102609280
N=10:1190433705317814685295399296
N=11:12704638649578287993184246767488
...

Let’s see two of them…



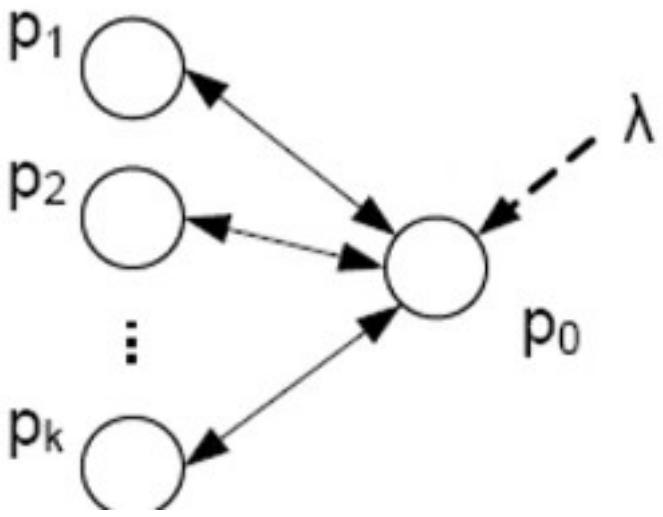
Example.....

- Remember the «teleport» component in pagerank



This way...

- We could measure for a web page / web site what is the effect of teleportation
- A value called ***relative spam mass***
- → if it is too big, maybe something is wrong (too many «secondary» pages are offering contributions)



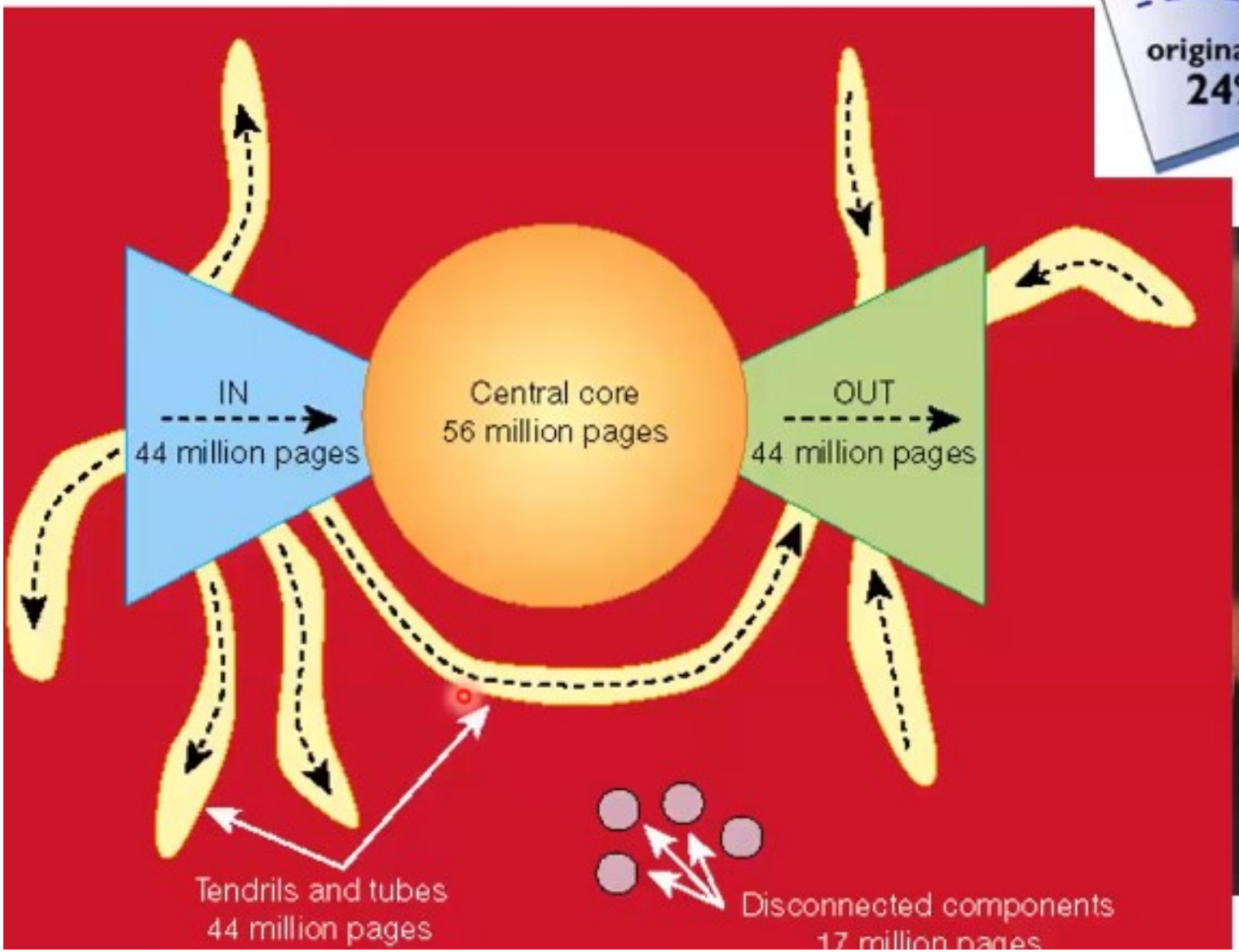
Result?

- Success rate to find spam:
- **95%-100%** ☺

Other example of countermeasure

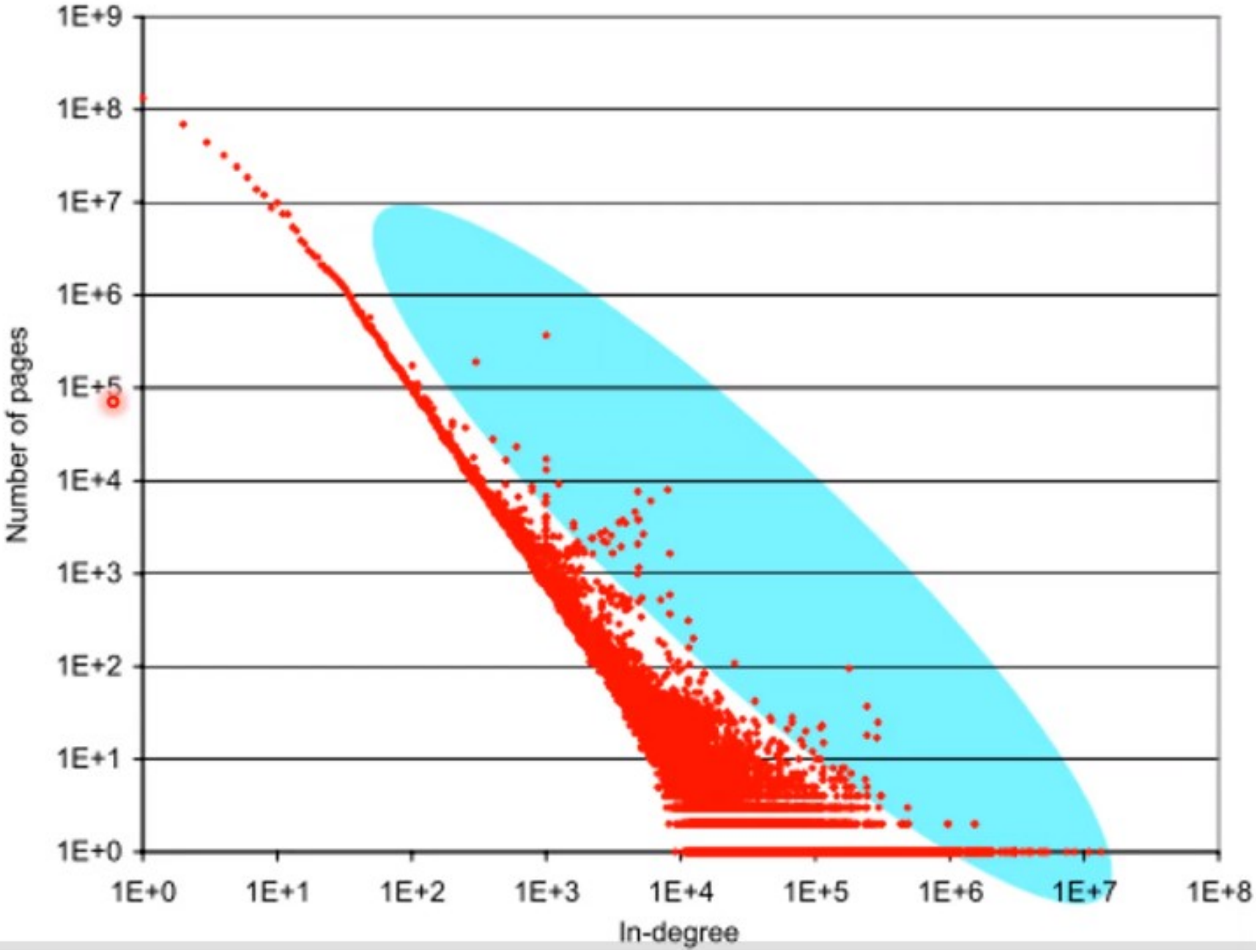
- Does the web have a structure? Or is it just chaotic?

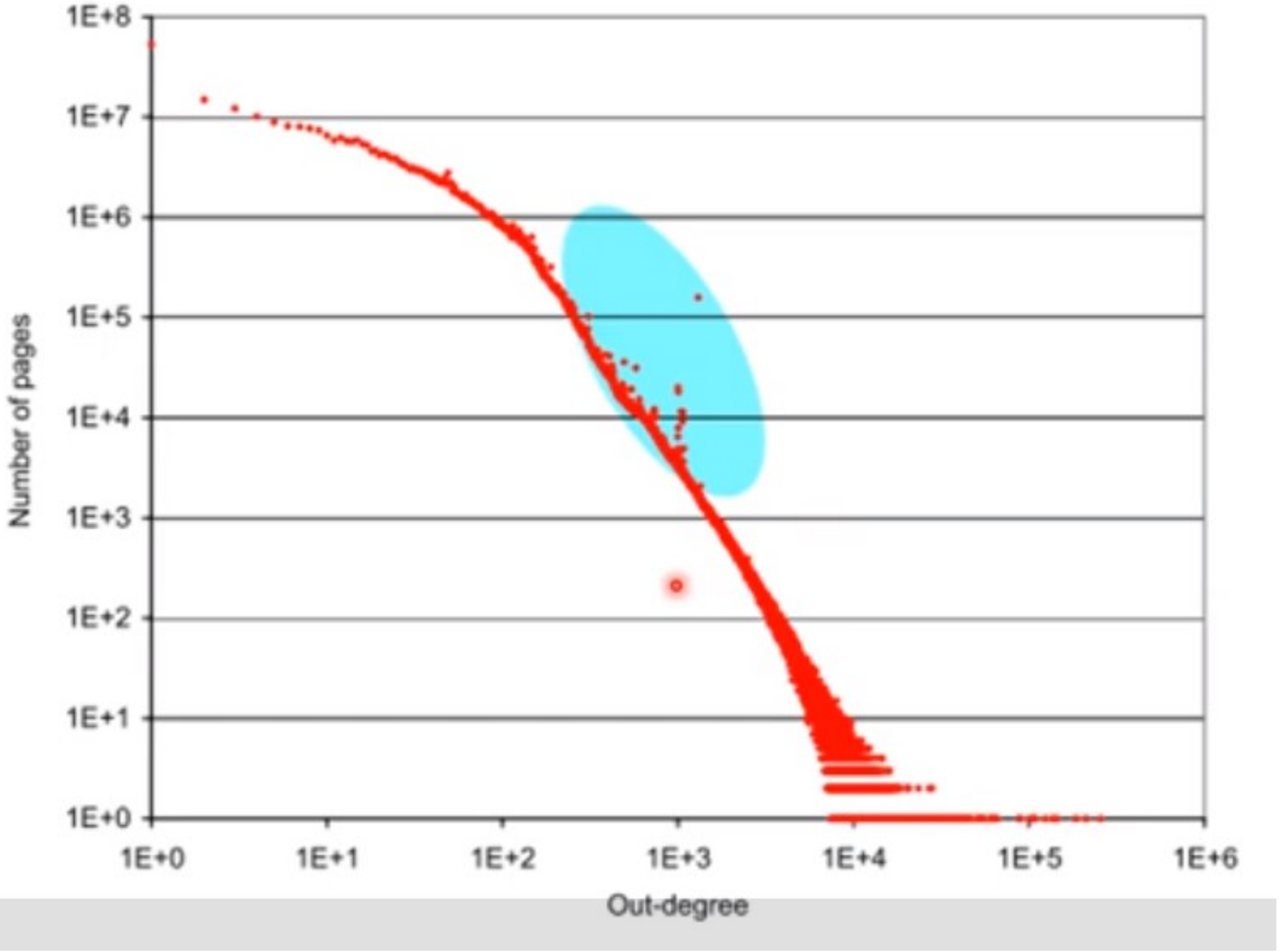

pages
22%
origination
24%
core
30%
24%
termination

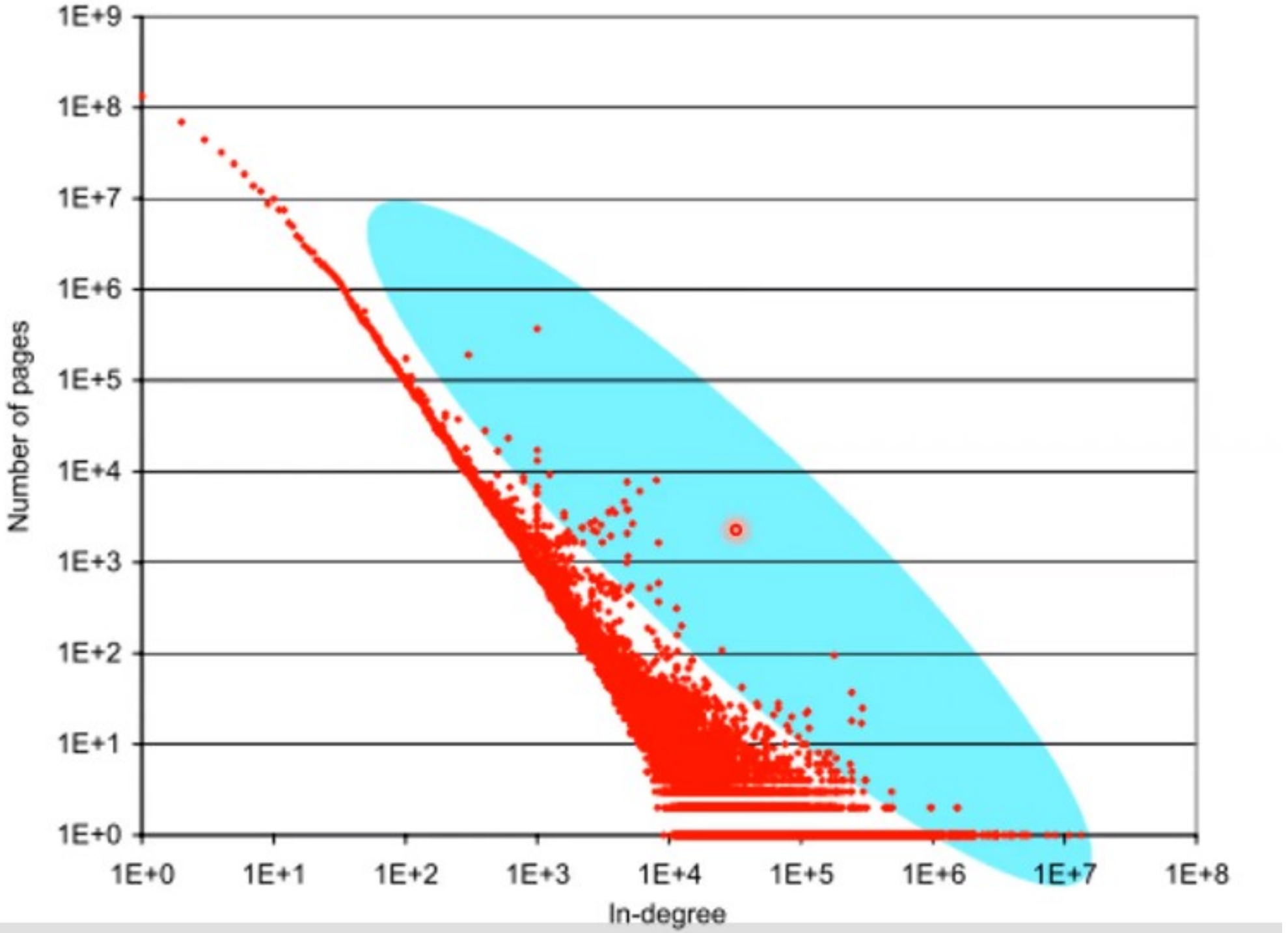

IN
44 million pages
Central core
56 million pages
OUT
44 million pages
Tendrils and tubes
44 million pages
Disconnected components

Super-Powerful countermeasure!

- Base idea: analyze the «shape» of a web site: if this is too much different
  - from the average, then something is wrong
- All this (challange) in an ***efficient way*** (!!)


1E+9
1E+8
1E+7
1E+6
1E+5
1E+4
1E+3
1E+2
1E+1
1E+0
Number of pages
1E+0
1E+1
1E+2
1E+3
1E+4
1E+5
1E+6
1E+7
1E+8
In-degree


1E+8
1E+7
1E+6
1E+5
1E+4
1E+3
1E+2
1E+1
1E+0
Number of pages
1E+0
1E+1
1E+2
1E+3
1E+4
1E+5
1E+6
Out-degree


1E+9
1E+8
1E+7
1E+6
1E+5
1E+4
1E+3
1E+2
1E+1
1E+0
Number of pages
1E+0
1E+1
1E+2
1E+3
1E+4
1E+5
1E+6
1E+7
1E+8
In-degree