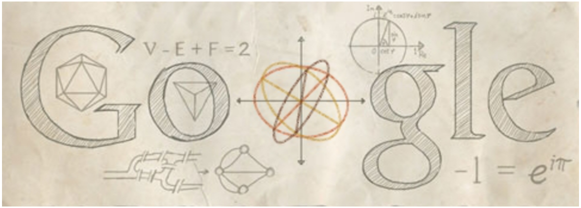
The Mathematics of Google

Firas Rassoul-Agha

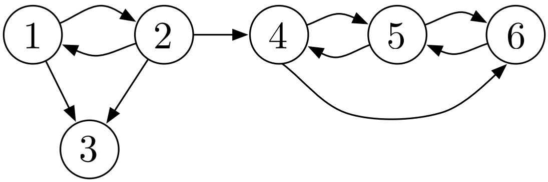
Department of Mathematics
University of Utah

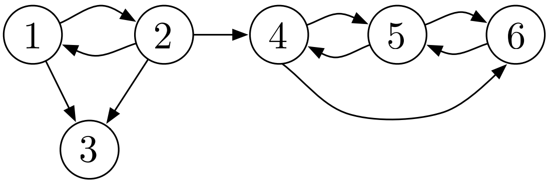
November 15, 2014



Consider a world tiny web consisting of 6 sites.

Site 1 is John's homepage. John teaches calculus and his site points to sites 2 and 3. Site 3 is the course syllabus and does not point to any other sites. Site 2 is the calculus course website. It points back at John's homepage and also at the course syllabus. It also points at Emily's webpage, site 4. (Emily is the TA for the course.) Site 5 belongs to a friend of Emily's, Jack. It points at Emily's website and at Jack's old website, site 6. Emily's website points at both Jack's pages, the new one 5 and the old one 6. Jack's old webpage 6 points at his new webpage.





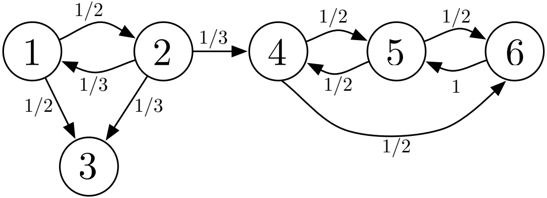
Question: How can one rank these sites?

Of course, the ranking should reflect the fact that the more links a site gets the more important it is.

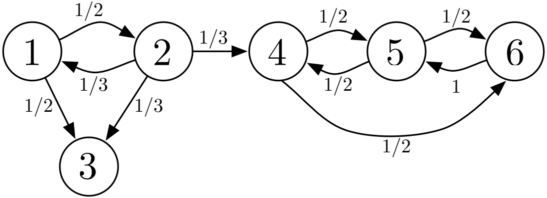
Also, links from important sites should count more.

And if a site has too many links, its importance should drop.

Consider the game where from any given site you click equally likely on any of the links on that site and go there.



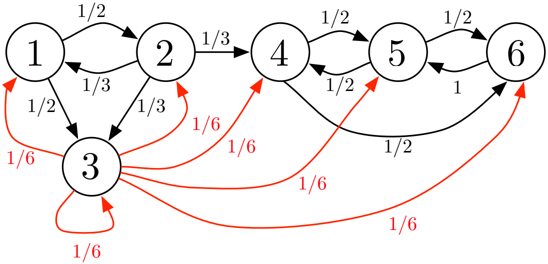
Consider the game where from any given site you click equally likely on any of the links on that site and go there.



What happens if you eventually reach site 3?

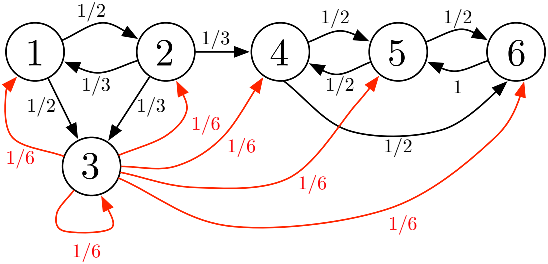
Solution: when a site does not have any links, start afresh.

Go equally likely to any site on the web.



Solution: when a site does not have any links, start afresh.

Go equally likely to any site on the web.



What happens once you reach 4, 5, or 6?

Solution: modify the game a tiny bit.

Solution: modify the game a tiny bit.

If you are at any site other than 3, flip a coin that lands heads with probability 90%.

Solution: modify the game a tiny bit.

If you are at any site other than 3, flip a coin that lands heads with probability 90%.

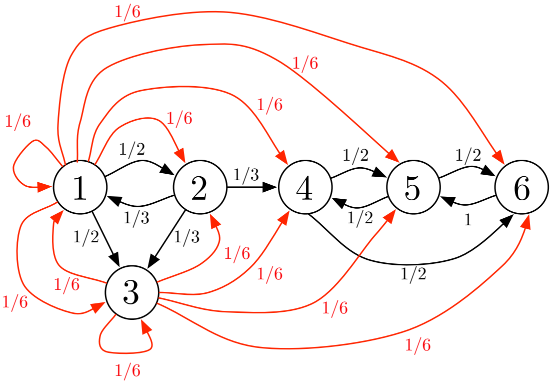
If it lands heads, move as before: equally likely, but only to the links on the page.

Solution: modify the game a tiny bit.

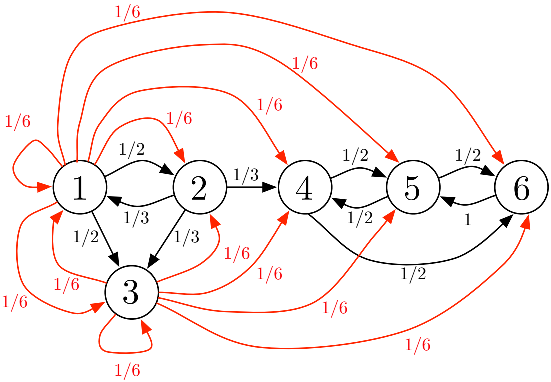
If you are at any site other than 3, flip a coin that lands heads with probability 90%.

If it lands heads, move as before: equally likely, but only to the links on the page.

If it lands tails, move to any of the web's sites equally likely.



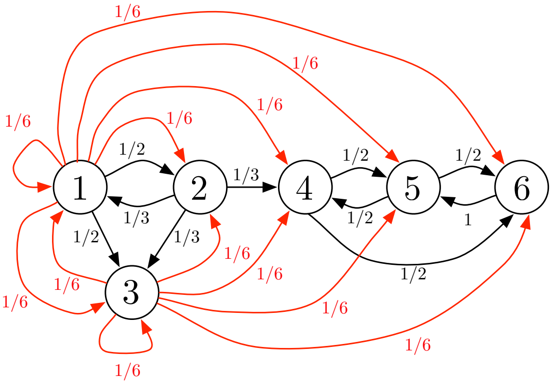
(with similar red arrows from sites 2, 4, 5, and 6.)



(with similar red arrows from sites 2, 4, 5, and 6.)

The Rule: unless you are at 3, with probability 90% follow the black arrows and with probability 10% follow the red arrows.

Now, no more problems.



(with similar red arrows from sites 2, 4, 5, and 6.)

The Rule: unless you are at 3, with probability 90% follow the black arrows and with probability 10% follow the red arrows.

Now, no more problems. No Site Left Behind!

Imagine now you assign to everyone on earth one of the 6 sites. (The initial assignment can be made in any way you wish. Assign 1 to everybody, or assign sites equally, or any other way you want!)

Imagine now you assign to everyone on earth one of the 6 sites. (The initial assignment can be made in any way you wish. Assign 1 to everybody, or assign sites equally, or any other way you want!)

Next, let each person play the above game: every second each person moves according to the rules of the game and ends up at the next site. They wait for a second and then repeat the processes to move to another site, and so on.

Imagine now you assign to everyone on earth one of the 6 sites. (The initial assignment can be made in any way you wish. Assign 1 to everybody, or assign sites equally, or any other way you want!)

Next, let each person play the above game: every second each person moves according to the rules of the game and ends up at the next site. They wait for a second and then repeat the processes to move to another site, and so on.

Theorem: If you let this processes go on for a long time (say, a day or so), the fraction of people at each site will converge to a fixed (non random!) number between 0 and 1.

Imagine now you assign to everyone on earth one of the 6 sites. (The initial assignment can be made in any way you wish. Assign 1 to everybody, or assign sites equally, or any other way you want!)

Next, let each person play the above game: every second each person moves according to the rules of the game and ends up at the next site. They wait for a second and then repeat the processes to move to another site, and so on.

Theorem: If you let this processes go on for a long time (say, a day or so), the fraction of people at each site will converge to a fixed (non random!) number between 0 and 1.

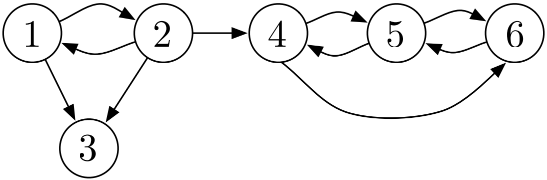
This final distribution of "mass" does not depend on the initial assignments, nor on the random evolution the games!

It only depends on the architecture of the network.

Now: use these numbers to rank the sites.

The higher the mass, the more important the site is.

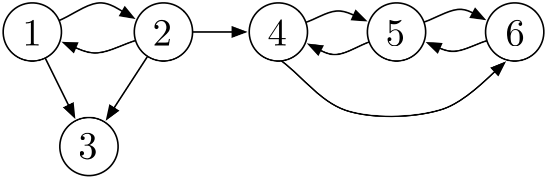
This is the basic idea behind Google's ranking method!



Now: use these numbers to rank the sites.

The higher the mass, the more important the site is.

This is the basic idea behind Google's ranking method!



5	6	4	3	2	1
0.3751	0.2862	0.2060	0.0540	0.0415	0.0372

Questions?