The University of Georgia

Mathematics Education
EMAT 4680/6680 Mathematics with Technology
Jim Wilson, Instructor

Google Page Rank

by

Josh Traxler

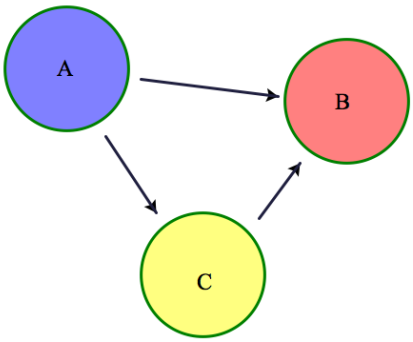

Figure 1: Amber, Bob, and Cal link direction graph.

We now look at some of the math behind the page rank algorithm, which is fundamental to how the google search process works.

Have you ever wondered what makes google tick? You put in a phrase, and like an oracle, it tells you where you need to be. It turns out this sorcery actually has it's base in mathematics. In this essay we're going to explore a bit of the innovative mathematical algorithms that puts the magic into our favorite search engine!

There is of course a LOT of mathematics that goes into it, but the revolutionary idea that shot it ahead of the other search engines was initially pagerank. The basic idea is that the more links there are that connect to a page, the higher the pagerank becomes. Of course, we are only counting links from pages that contain our key word in them.

Pagerank works a lot like a voting system, where a webpage has to give away all of its votes through the links it provides, with its current pagerank determining the number of votes it can give out. For example, if the webpage has a pagerank of 20 and has 5 links to other webpages, each of these webpages receives 4 of its votes. Imagine you are in a room filled with people, and you are told that you have to give choose one or more people to give all your money to, giving of those people and equal share. But everyone else in the room is also forced to do this. This is similar to how one round of pagerank works.

But how did those pages get their ranks to give out in the first place? Well, each page begins with a pagerank of one. Then after one round of link voting, we use the new pageranks for our next round of voting. So it is an iterative process. If a page doesn't link to any other pages, then it is as if it links to every other webpage out there.

For example, say we have three pages that we will name Amber, Bob, and Cal. Each of them start with one pagerank.

Amber links to Bob and Cal. Bob links to noone. Cal links to Bob. This can be seen using a link graph in figure 1.

So, Amber sends half a vote to both Bob and Cal. Cal sends a whole vote to Bob. Since Bob links to noone, he sends his votes to everyone or half of a vote to both Amber and Cal. This can be seen using a pagerank graph in figure 2. Notice that there are some subtle but distinct differences in how pagerank rules differ from what page is linked to what page, namely that Bob's page rank is

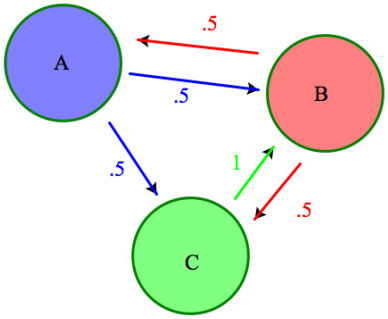

Figure 2: Amber, Bob, and Cal pagerank graph.

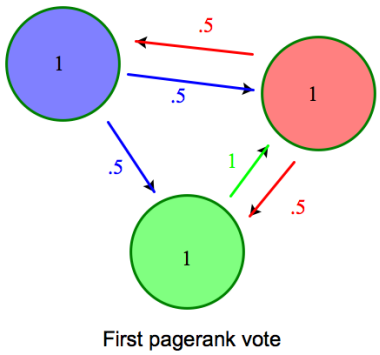

Figure 3: Pagerank progression

distributed to his two friends.

Each vote will be based on this page rank distribution. We see that the first vote in figure 3 looks almost identical to our pagerank graph. At the end of this first round of voting, Amber has half a vote, Bob has 1.5 votes, Cal has 1 vote. The total votes amounts to their new pageranks. Notice that the sum of the ranks never change in size, they are just redistributed. So, now we have a better idea of the relative popularity of the pages: Amber, Bob and Cal. So notice that a link from Bob, generally means more than a link from Amber, since Bob has 1.5 votes to give instead of half.

So, in our second round of voting. Amber sends .25 votes to Bob and Cal. Bob sends .75 votes to Amber and Cal. Cal sends 1 vote to Bob. So, at the end of the second voting session, Amber has .75 votes, Bob has 1.25 votes, and Cal has 1 vote.

In our third round of voting Amber sends 3/8 votes to Bob and Cal, Bob send 5/8 votes to Amber and Cal, and Cal sends a vote to Bob. At this end of this round of voting, Amber has 5/8 votes, Bob has 1.375 votes and Cal has 1 vote.

This is the basic algorithm that google employs to determine the relative popularity of pages. Though this a rather intensive task to determine the pagerank

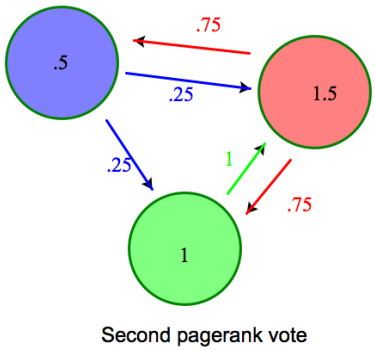

Figure 4: Pagerank progression

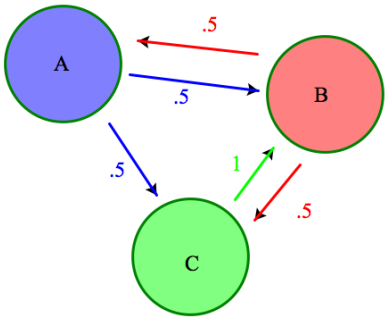

Figure 5: Pagerank progression

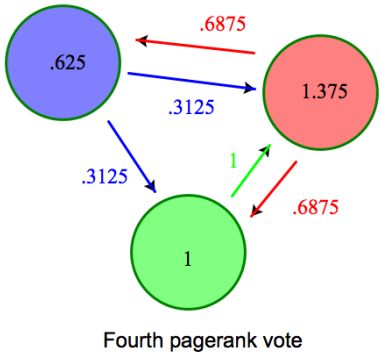

Figure 6: Pagerank progression

for these few pages. Surely there is an easier method? Well, it turns out that matrices perform the exact task that we've been describing!

Take the matrices M and N, where a+b+c=d+e+f=g+h+i=1. Notice that each column adds up to one!

M= $\begin{pmatrix} a & d & g \\ b & e & h \\ c & f & i \end{pmatrix}$

N= $\begin{pmatrix} j \\ k \\ l \end{pmatrix}$

One way of conceptualizing multiplication of the type $M * N$, is to say that each column of M determines how the contents of the cells of N, ie. j, k, and l, are redistributed. For example, take the cell k. d of k will go to cell j, e of k will go back to cell k, and f of k will go to cell l. Recall that each column adds up to 1, so we aren't making the sum of our three cells in N any larger. Our result of M*N is

$$\begin{pmatrix} a & d & g \\ b & e & h \\ c & f & i \end{pmatrix} \begin{pmatrix} j \\ k \\ l \end{pmatrix} = \begin{pmatrix} a*j+d*k+g*l \\ b*j+e*k+h*l \\ c*j+f*k+i*l \end{pmatrix}$$

So we see more clearly now that the quantity in j being redistributed into itself, k, and l the proportions of a, b and c. Similarly for k and l. We see that this is exactly the sort of thing that happened in our pagerank examples with Amber, Bob and Cal!

Let's see what happens when we represent the pagerank using matices.

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

Looks good so far! Let's see what happens when we look at the individual voting processes, starting with the first, where every page starts with a rank of 1.

$$\begin{pmatrix} 0 & \frac{1}{2} & 0 \\ \frac{1}{2} & 0 & 1 \\ \frac{1}{2} & \frac{1}{2} & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0+\frac{1}{2}+0 \\ \frac{1}{2}+0+1 \\ \frac{1}{2}+\frac{1}{2}+0 \end{pmatrix} = \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2}+1 \\ \frac{1}{2}+\frac{1}{2} \end{pmatrix} = \begin{pmatrix} \frac{1}{2} \\ \frac{3}{2} \\ 1 \end{pmatrix}$$

We end with exactly what were hoping to, the results from working out the graph directly! Which is to say that after the first voting sequence, Amber has one vote, Bob has 1.5 votes, and Cal has 1 vote. Now if we were see the results of the second voting, we would merely need to multiply our voting distribution matrix by our results. This is because the proportions of Amber's vote, and everyone else's vote, are all the same.

$$\begin{pmatrix} 0 & \frac{1}{2} & 0 \\ \frac{1}{2} & 0 & 1 \\ \frac{1}{2} & \frac{1}{2} & 0 \end{pmatrix} \begin{pmatrix} \frac{1}{2} \\ \frac{3}{2} \\ 1 \end{pmatrix} = \begin{pmatrix} .75 \\ 1.25 \\ 1 \end{pmatrix}$$

We see that a calculator or a computer with a matrix utility in it could make this process very streamlined! Remember that results of the first vote were:

$$\begin{pmatrix} 0 & \frac{1}{2} & 0 \\ \frac{1}{2} & 0 & 1 \\ \frac{1}{2} & \frac{1}{2} & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

So, another way to find the results of the second vote would be to multiply the pagerank matrix by this! This gives:

$$\begin{pmatrix} 0 & \frac{1}{2} & 0 \\ \frac{1}{2} & 0 & 1 \\ \frac{1}{2} & \frac{1}{2} & 0 \end{pmatrix} \begin{pmatrix} 0 & \frac{1}{2} & 0 \\ \frac{1}{2} & 0 & 1 \\ \frac{1}{2} & \frac{1}{2} & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 & \frac{1}{2} & 0 \\ \frac{1}{2} & 0 & 1 \\ \frac{1}{2} & \frac{1}{2} & 0 \end{pmatrix}^2 \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

In fact, if we wish to know the results of of any nth test, we may simply find the results of the matrix product:

$$\begin{pmatrix} 0 & \frac{1}{2} & 0 \\ \frac{1}{2} & 0 & 1 \\ \frac{1}{2} & \frac{1}{2} & 0 \end{pmatrix}^n \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

This method really streamlines the process! If we choose n to be 100, which would normally take a long time to calculate, we get approximately:

$$\begin{pmatrix} \frac{2}{3} \\ \frac{4}{3} \\ 1 \end{pmatrix}$$

What an amazing application of matrices! Though, as you can imagine, in a real google algorithm, there are a lot more than just three pages! So much more robust calculations are being made. Though even if we were to use hundreds of pages, the basic process would be the same. This algorithm was a key innovation allowed for google's success. Google is characterized by a great many innovations, but this algorithm is no doubt emblematic of their early success.

Page rank method and more found at $MathematicsatGoogle$ by Javier Tordable. http://static.googleusercontent.com/media/research.google.com/en/us/pubs/archive/38331.pdf