# How does Google work?

*Javier Rodríguez-Laguna, UC3M*

April 7, 2008

This text is a didactic review of the contents of the original paper about PageRank, the relevance assignment algorithm which gave birth to the google search engine. You can find the original paper linked in this webpage.

Why is google so successful? There are many other search engines in the web... The ordering of the results is the key. Let's say that you search for "algebra". Then you get thousands of results. Google puts order into such a list, and gives the results to you in order of "relevance". Now the real question comes: how does google know about the relevance of a webpage?

Before google, the main algorithm was to find out the number of "backlinks". A backlink is a link coming from another webpage. This algorithm is not bad, but it leads to trouble in many cases. First of all, because not all backlinks are equally important. If your webpage has a backlink from, let's say, the New York Times main page, it is not the same as if it is from an obscure webpage that nobody visits... The other peril of the mere "backlink count" is that companies really need to come high in search engines. Therefore, they can make a lot of fake webpages with links to their main page in order to fool search engines...

So the main idea is that of "relevance" of a webpage. A webpage is relevant if it is linked by relevant webpages. This definition is quite *circular*... no? Yes, it is. But we prefer to say that it is *self-consistent*!

Let us try to be more precise. We consider a webpage $i$, whose relevance is known to be $R_i$, and linked to $L_i$ other webpages. Now we say that it gives relevance $cR_i/L_i$ to each of these webpages, where $c$ is a certain normalization factor. In a sense, you can think that a webpage is allowed to assign to other webpages a fraction $c$ of its own relevance. How much is $c$? Let's leave that open by now. Then, the relevance of a webpage is the sum of the relevances given to it by its backlinks, is it clear? This leads to a set of equations:

$$R_i = c \sum_{\langle i,j \rangle} \frac{R_j}{L_j}$$

That's a set of equations we can try to solve. Uff... let's give an example right now! Consider three webpages, numbered one to three, in the following





way: 1 points to 2, 2 points to 3, and 3 points to 1 [figure]. Then, each $L_i = 1$, and the equations read:

$$R_1 = cR_3 \qquad R_2 = cR_1 \qquad R_3 = cR_2$$

From here we get $R_1 = c^3 R_1$, and the same for the others. If $c \neq 1$, then all three are zero! So, $c$ must be one: all relevance is assigned to the linked pages. And then, $R_1 = R_2 = R_3$. All webpages have the same relevance. How much? Does not matter, only the order is important. We can fix a scale by definition, saying that $R_1 = 1$. At this moment you might think: what is the full idea of $c$, if it is going to be one? OK, not always!

Let us consider another example. Pages 1 and 2 are linked among themselves, and page 2 also links page 3, which does not contain any links [Figure]. Find the relevances here.

In this case the equations are: $R_1 = cR_2/2$, $R_2 = cR_1$ and $R_3 = cR_2/2$. If we try to solve these equations for $c = 1$ we get only the trivial solution! So, solving for general $c$, we get

$$\begin{pmatrix} 0 & 1/2 & 0 \\ 1 & 0 & 0 \\ 0 & 1/2 & 0 \end{pmatrix} \begin{pmatrix} R_1 \\ R_2 \\ R_3 \end{pmatrix} = c^{-1} \begin{pmatrix} R_1 \\ R_2 \\ R_3 \end{pmatrix}$$

This is an eigenvalue equation, if we solve it we get the characteristic equation ($\lambda = c^{-1}$)

$$\begin{vmatrix} -\lambda & 1/2 & 0 \\ 1 & -\lambda & 0 \\ 0 & 1/2 & -\lambda \end{vmatrix} = 0$$

Solving that determinant we get $-\lambda^3 + \lambda/2 = 0$. This gives the solutions $\lambda = 0$, $\lambda = \pm 1/\sqrt{2}$. Out of these values for $\lambda$, which value shall we choose? Clearly, $\lambda = 0$ does not make sense, since $c = \infty$! Also, $\lambda = -1/\sqrt{2}$ also is wrong: relevances can't be negative. So only $\lambda = 1/\sqrt{2}$, i.e.: $c = \sqrt{2}$ does make sense.

OK, then, we get $c = \sqrt{2}$. Now let's find the relevances. Substituting in the original equations, if we set $R_1 = 1$ (global scale!) we get $R_1 = 1$, $R_2 = \sqrt{2}$, $R_3 = 1$.