

Worked PageRank Example

Posted on [2-January-2012](#) by [mathscinotes](#)

Introduction

One of my sons asked me if I could work through a PageRank calculation example a couple of different ways (algebraic and iterative). It was an interesting exercise and I thought it would be worth documenting here. I used Mathcad for both my algebraic and iterative solutions.

The Wikipedia has an excellent [definition](#) of the PageRank algorithm, which I will quote here.

PageRank is a link analysis algorithm, named after Larry Page[1] and used by the Google Internet search engine, that assigns a numerical weighting to each element of a hyperlinked set of documents, such as the World Wide Web, with the purpose of "measuring" its relative importance within the set.

Businesses and individuals who want their site to rank more highly in Google's search results could look at the services of a [Digital Marketing Company](#) to help boost their position.

PageRank is computed using a relatively simple function (see Equation 1), but a number of web-based examples treat the weighting of inbound links from sites external to a particular group of pages as a special case. I did not see any explicit calculation examples, so I thought I would include this calculation here, you'll more than definitely want to take this calculation into consideration if you're wanting to [learn to how to build a website](#) as well as wishing to attract traffic to your new site.

Background

PageRank views the web as graph, with inbound links being viewed as measure of the significance of a web page. Equation 1 shows the PageRank equation. Note that this equation can be solved several different ways. One approach involves eigenvalues, which my son does not know about yet. The equation can be solved algebraically or iteratively. I will use both approaches for this example.

Eq. 1
$$PR(p_i) = \frac{1-d}{N} + d \cdot \sum_{p \in M(p_i)} \frac{PR(p_j)}{L(p_j)}$$

where

- N is the number of pages.
- d is called the damping factor and it is an arbitrary weighting factor.
- $PR(p_i)$ is the PageRank of page p_i .
- $L(p_j)$ is the number of outbound links from page p_j .
- $M(p_i)$ is the set of links to page p_i .

Equation 2 shows a matrix form of Equation 1. The matrix form is most likely the form used "out in the wild."

Eq. 2
$$\mathbf{R}(t+1) = d \cdot \mathcal{M} \cdot \mathbf{R}(t) + \frac{1-d}{N} \cdot \mathbf{1}$$

where

- \mathbf{R} is the PageRank vector.
- $\mathbf{1}$ is a column vector with all elements equal to 1.
- t is a discrete time variable (really a sequence number).

Note that many examples of PageRank are computed using a variant of Equations 1 and 2 that multiplies the PageRank value by the number of pages ($N \cdot \text{PageRank}$). Equation 3 illustrates Equation 2 modified with the substitution $\mathbf{R}' = N \cdot \mathbf{R}$. Equation 1 can be modified similarly. This is the equation that Ian used for his examples. Since I am going to duplicate his results, I will multiply my results by N .

Eq. 3	$N \cdot \mathbf{R}(t+1) = N \cdot d \cdot \mathcal{M} \cdot \mathbf{R}(t) + N \cdot \frac{1-d}{N} \cdot \mathbf{1}$
	$(N \cdot \mathbf{R}(t+1)) = d \cdot \mathcal{M} \cdot (N \cdot \mathbf{R}(t)) + (1-d) \cdot \mathbf{1}$
	$\therefore \mathbf{R}'(t+1) = d \cdot \mathcal{M}' \cdot \mathbf{R}'(t) + (1-d) \cdot \mathbf{1}, \text{ where } \mathbf{R}' = N \cdot \mathbf{R}$

There has been quite a bit written about the nuances of this equation because of its importance in determining a web page's position in a list of search results. I am not concerned about those details here. I am focused here on the calculation of the PageRank for a specific set of pages.

Analysis

Example

My son was using [Ian Roger's excellent site](#) for learning about the details of PageRank. The question he had is on Example 10, which assigns a PageRank of 1 to an external page. Figure 1 shows the Example 10's web page configuration. Ian's PageRank results are shown in the boxes, which represent web pages. I want to show the details on obtaining Ian's results as an illustration of how to handle an external link from a page with a defined PageRank.

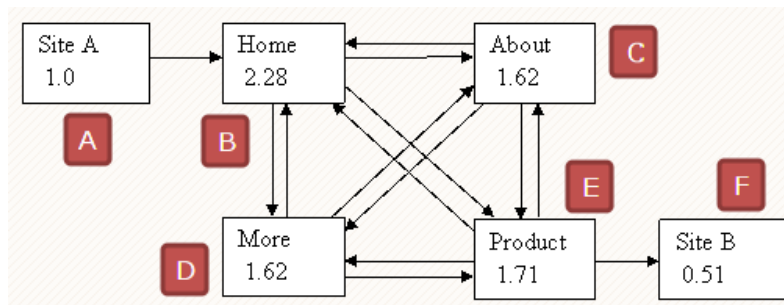


Figure 1: PageRank Example from Ian Roger's Website.

I have added the letters A-F to Figure 1, which constitute the variable names I will use in my solutions.

Algebraic Solution

When Ian uses a link from an external site, he sets the PageRank value to 1. Here is his rationale.

We'll assume there's an external site that has lots of pages and links with the result that one of the pages has the average PR of 1.0.

Algebraically, this is easy to handle. Figure 2 shows my solution implemented in Mathcad.

$$\begin{aligned}
 &\text{prA} := 1 \quad \text{prB} := 1 \quad \text{prC} := 1 \quad \text{prD} := 1 \quad \text{prE} := 1 \quad \text{prF} := 1 \\
 &d := 0.85 \\
 &\text{given} \\
 &\text{prB} = 1 - d + d \cdot \left(\text{prA} + \frac{\text{prC}}{3} + \frac{\text{prD}}{3} + \frac{\text{prE}}{4} \right) \\
 &\text{prC} = 1 - d + d \cdot \left(\frac{\text{prB}}{3} + \frac{\text{prD}}{3} + \frac{\text{prE}}{4} \right) \\
 &\text{prD} = 1 - d + d \cdot \left(\frac{\text{prB}}{3} + \frac{\text{prC}}{3} + \frac{\text{prE}}{4} \right) \quad + \\
 &\text{prE} = 1 - d + d \cdot \left(\frac{\text{prB}}{3} + \frac{\text{prC}}{3} + \frac{\text{prD}}{3} \right) \\
 &\text{prF} = 1 - d + d \cdot \left(\frac{\text{prE}}{4} \right) \\
 &\text{Find}(\text{prB}, \text{prC}, \text{prD}, \text{prE}, \text{prF}) = \begin{pmatrix} 2.28 \\ 1.62 \\ 1.62 \\ 1.71 \\ 0.51 \end{pmatrix}
 \end{aligned}$$

Figure 2: Algebraic Solution for Example 10.

Iterative Solution

Figure 3 shows how I setup my iterative solution. To force page A to have a PageRank of 1, I needed to remove page A from the R vector and M matrix, but add it back in so that page A's contribution can be included. Again, it is a slight modification of Equation 3 so that I can force page A to have a PageRank of 1.

$$\begin{aligned}
 &d := 0.85 \quad h(m,n) := 1 \quad \text{Sum the numbers in a vector.} \\
 &R := \text{matrix}(5,1,h) \quad N := \text{rows}(R) \quad N = 5 \quad \text{sum}(v) := \begin{cases} s \leftarrow 0 \\ \text{for } i \in 0.. \text{rows}(v) - 1 \\ s \leftarrow s + v_i \\ s \end{cases} \\
 &A := \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix} \quad \begin{array}{l} A \text{ defines the graph for this example.} \\ \text{Each one means a connection.} \end{array} \\
 &\text{normalize}(A) := \begin{cases} \text{for } i \in 0.. \text{cols}(A) - 1 \\ \quad \text{for } j \in 0.. \text{rows}(A) - 1 \\ \quad \quad M_{i,j} \leftarrow 0 \\ \text{for } i \in 0.. \text{cols}(A) - 1 \\ \quad c_i \leftarrow \text{sum}(A_{\langle i \rangle}) \\ \text{for } i \in 0.. \text{cols}(A) - 1 \\ \quad \text{for } j \in 0.. \text{rows}(A) - 1 \\ \quad \quad M_{j,i} \leftarrow \frac{A_{j,i}}{c_i} \text{ if } c_i > 0 \end{cases} \quad \begin{array}{l} _A := \text{submatrix}(A, 1, 5, 1, 5) \\ _A = \begin{pmatrix} 0 & 1 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix} \\ \text{I have pulled page A from my connection} \\ \text{matrix so that I can force its value to 1.} \end{array} \\
 &M
 \end{aligned}$$

Figure 3: Setup for My Iterative Solution of Equation 2.

Figure 4 shows my iteration stage. Note that I was lazy and simply let it iterate a 1001 times. I could have made this much more efficient by simply monitoring the convergence, but that would have taken a bit more time.

$$M := \text{normalize}(_A)$$

$$M = \begin{pmatrix} 0 & 0.333 & 0.333 & 0.25 & 0 \\ 0.333 & 0 & 0.333 & 0.25 & 0 \\ 0.333 & 0.333 & 0 & 0.25 & 0 \\ 0.333 & 0.333 & 0.333 & 0 & 0 \\ 0 & 0 & 0 & 0.25 & 0 \end{pmatrix}$$

$$\text{Iterate}(R) := \begin{array}{l} \text{for } i \in 0..1000 \\ \quad R \leftarrow (1 - d) + d \cdot M \cdot R + d \cdot \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} \\ \text{end for} \\ R \end{array}$$

$$\text{Iterate}(R) = \begin{pmatrix} 2.28 \\ 1.62 \\ 1.62 \\ 1.71 \\ 0.51 \end{pmatrix}$$

Figure 4: Iterative Solution of Equation 2.

Conclusion

I obtained the same results as Ian using two different approaches -- algebraic and iterative. This example was different than most in that a particular web page was forced to a particular PageRank. I hope that I answered my son's question.