

Q .Speculate why these documents have high PageRank values, i.e., why is it that these particular pages are linked to by (possibly) many other pages with (possibly) high PageRank values.

Ans: Top 10 pages have high page rank values because they are linked by pages with high page rank value. Also the out-link counts of those pages are very less, so the share that is spread across the out-links will be less compared to the pages with higher out-link count. So, we can see page ranks as a vote for importance of the page, so greater the page rank of the in-link pages of a particular page, higher will be its page rank.

Q. Are all of these documents ones that users would likely want to see in response to an appropriate query? Give some examples of ones that are and ones that are not. For those that are not "interesting" documents, why might they have high PageRank values?

Ans : No. There are a few documents that user did not want to see in response to an appropriate query, but they appeared as they have high page rank values.

Examples of pages that user did NOT want to see:

WT21-B37-76 – experimental home page of “The Economist”

WT21-B37-75 – copyright notice

WT08-B18-400 - some kind of notice

Examples of pages that user wanted to see:

WT25-B39-116 – security assurance requirement

WT01-B18-225 - Online Library of Drug Policy

The pages which are not “interesting documents” , have high page rank values because of their higher in-link count and lesser out-link counts. There is very high probability that many pages will include the copyright and home page link and many of those pages will have high page rank. Also the out-links on such pages is lesser, so the share spread across the out-links will be relatively smaller than the pages with higher out-link.

Below value will illustrate the answer:

page	in-link count	out-link count
WT21-B37-76	2568	5
WT21-B37-75	1704	1

**Q How do the pages with high PageRank compare to the pages with many in-links?
In short, give an analysis of the PageRank results you obtain.**

Ans: The pages highlighted in green (please refer to below table) are the pages, which are not “interesting documents”. So, we can ignore them for time being. We can see that not all the pages with highest in-link counts have high page rank. This may be because the pages that are linked to such pages may not have high page ranks. For example, the page WT01-B18-225 has in-link count 1137, but PR only 0.0009553.

On the other hand, pages with high page rank may not have many in-links, but they have in-links from important pages i.e. pages with high page rank values.

e.g WT25-B39-116 has PR 0.0014 although its in-link count is 169.

In-a-nut-shell, it is not necessary that a page with very large number of in-link count will have very high page rank. It depends on the page ranks of the pages in-linked to this page.

Top 10 pages by page ranks	Top 10 pages by in-link count
WT21-B37-76: 0.00269447441555	WT21-B37-76: 2568
WT21-B37-75: 0.0015331791	WT21-B37-75: 1704
WT25-B39-116: 0.00146850141633	WT01-B18-225: 1137
WT23-B21-53: 0.00137352493843	WT08-B19-222: 1041
WT24-B26-10: 0.00127615674413	WT08-B18-400: 990
WT24-B40-171: 0.00124529321761	WT21-B40-447: 779
WT23-B39-340: 0.00124291315832	WT27-B34-57: 630
WT23-B37-134: 0.00120542141123	WT27-B32-30: 628
WT08-B18-400: 0.00114477728784	WT25-B15-307: 605