# Analysis of Page Rank Algorithm

* By Nitish Nadkarni

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Why these documents have high PageRank values?**

The top 10 pages with highest PageRank are as follows:

WT21-B37-76 PageRank : 0.0026944737779136755 Number of Inlinks : 2568

WT21-B37-75 PageRank : 0.0015331787464243488 Number of Inlinks : 1704

WT25-B39-116 PageRank : 0.0014685026346494148 Number of Inlinks : 169

WT23-B21-53 PageRank : 0.0013735263702095604 Number of Inlinks : 198

WT24-B26-10 PageRank : 0.0012761659886317569 Number of Inlinks : 291

WT24-B40-171 PageRank : 0.001245220233833983 Number of Inlinks : 270

WT23-B39-340 PageRank : 0.0012428201898918969 Number of Inlinks : 274

WT23-B37-134 PageRank : 0.0012054223989578401 Number of Inlinks : 207

WT08-B18-400 PageRank : 0.0011447771363817557 Number of Inlinks : 990

WT13-B06-284 PageRank : 0.0011365568586814505 Number of Inlinks : 454

As can be seen the page WT21-B37-76 has the highest page rank value. This page also has the largest number of inlinks. Moreover, the pages which are pointing to this page also have large number of inlinks. As a result the relative weight of WT21-B37-76 is highest compared to others.

Now, consider WT21-B37-75 which also has a high page rank. This page has a high page rank as it is linked to the page WT21-B37-76. As a result, the relative weight for WT21-B37-75 is high but not as much as WT21-B37-76. The weight is not as much as WT21-B37-75 has lesser number of inlinks.

(Note: WT21-B37-76 -> Economist homepage

WT21-B37-75 -> Copyright page of Economist)

If you see pages WT25-B39-116 and WT23-B21-53, these pages have lower number of inlinks but they are linked to pages which have higher weight or rank. As a result, they have higher page rank values.

This shows that the page rank is not only dependent on the number of in-links but also on the relative weight of each inlink. A page can have fewer inlinks but higher rank if the ranks of many of the in-links are high.

**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?**

Let’s look at what links are actually present in these document IDs.

|  |  |
| --- | --- |
| Document ID: WT21-B37-76 Value: 0.0026944744155488125 | <http://www.economist.com/index.html> |
| Document ID: WT21-B37-75 Value: 0.0015331790999983418 | http://www.economist.com:80/copyright.html |
| Document ID: WT25-B39-116 Value: 0.0014685014163258261 | [http://utopia.tno.nl:80/instit/fel/refs/cc1.0/fp0\_3.htm 192.87.96.2 - Security assurance lecture notes](http://utopia.tno.nl/instit/fel/refs/cc1.0/fp0_3.htm%20192.87.96.2%20-%20Security%20assurance%20lecture%20notes) |
| Document ID: WT23-B21-53 Value: 0.0013735249384256968 | [http://www.sea.usscouts.org:80/team.html - sea exploring scouts US national](http://www.sea.usscouts.org/team.html%20-%20sea%20exploring%20scouts%20US%20national) |
| Document ID: WT24-B26-10 Value: 0.0012761567441301662 | [http://www.psych.med.umich.edu:80/web/pStar/ - lecture notes about psychiatry star](http://www.psych.med.umich.edu/web/pStar/%20-%20lecture%20notes%20about%20psychiatry%20star) |
| Document ID: WT24-B40-171 Value: 0.0012452932176087732 | http://www.chch.com:80/news/eveningnews/archive/ |
| Document ID: WT23-B39-340 Value: 0.0012429131583240225 | [http://www.streetlink.com:80/REPORTS.HTM - financial reports](http://www.streetlink.com/REPORTS.HTM%20-%20financial%20reports) |
| Document ID: WT23-B37-134 Value: 0.0012054214112259519 | [http://www.public.health.wa.gov.au:80/Copyright.htm](http://www.public.health.wa.gov.au/Copyright.htm) |
| Document ID: WT08-B18-400 Value: 0.0011447772878379282 | [http://www.tdbank.ca:80/tdbank/mail.htm](http://www.tdbank.ca/tdbank/mail.htm) |
| Document ID: WT13-B06-284 Value: 0.0011365580053119455 | http://www.lawlead.com:80/common/credits.html |

(Please note that this is just a rough collection of links and when clicked may not open the actual page. Also the alignment may be skewed a bit)

According to me, the list of interesting/not interesting documents is as follows:

WT21-B37-76 -> interesting -> as it is the homepage of economist and will be used a lot by common users.

WT21-B37-75, WT25-B39-116 -> not interesting -> as these are copyright and requirement pages which may not be needed by regular search users.

WT24-B26-10 -> not interesting -> as this is a link to the department of psychiatry of a university, not needed for a regular user.

WT24-B40-171 and WT23-B39-340 -> these are interesting as these are news and financial reports.

The documents which are not-interesting have high page values because their inlinks have a relatively high page rank. If the page, which has inlink to the not-interesting page, has many other inlinks it will have high rank. As a result of which these documents may also have high page rank.

For Example, The copyright page (which is not interesting) is linked to the homepage of economist. As a result of which it has high page rank

**How do the pages with high PageRank compare to the pages with many in-links?**

SET1 : TOP 10 PAGES SORTED AS PER INLINKS:

WT21-B37-76 number of inlinks : 2568

WT21-B37-75 number of inlinks : 1704

WT01-B18-225 number of inlinks : 1137

WT08-B19-222 number of inlinks : 1041

WT08-B18-400 number of inlinks : 990

WT21-B40-447 number of inlinks : 779

WT27-B34-57 number of inlinks : 630

WT27-B32-30 number of inlinks : 628

WT25-B15-307 number of inlinks : 605

WT27-B28-203 number of inlinks : 589

SET2 : TOP 10 PAGES SORTED AS PER PAGE-RANK:

WT21-B37-76 PageRank : 0.0026944737779136755 Number of Inlinks : 2568

WT21-B37-75 PageRank : 0.0015331787464243488 Number of Inlinks : 1704

WT25-B39-116 PageRank : 0.0014685026346494148 Number of Inlinks : 169

WT23-B21-53 PageRank : 0.0013735263702095604 Number of Inlinks : 198

WT24-B26-10 PageRank : 0.0012761659886317569 Number of Inlinks : 291

WT24-B40-171 PageRank : 0.001245220233833983 Number of Inlinks : 270

WT23-B39-340 PageRank : 0.0012428201898918969 Number of Inlinks : 274

WT23-B37-134 PageRank : 0.0012054223989578401 Number of Inlinks : 207

WT08-B18-400 PageRank : 0.0011447771363817557 Number of Inlinks : 990

WT13-B06-284 PageRank : 0.0011365568586814505 Number of Inlinks : 454

As can be seen above, only 3 pages of Set1 are present in Set2. This shows that inlinks is not the only factor for pagerank.

Example:

WT25-B39-116 PageRank : 0.0014685026346494148 Number of Inlinks : 169

This page has only 169 links but still has a higher page rank than WT08-B18-400 which has higher number of inlinks.