

What is PageRank?

PageRank is a numeric value that represents how important a page is on the web. Google figures that when one page links to another page, it is effectively casting a vote for the other page. The more votes that are cast for a page, the more important the page must be. Also, the importance of the page that is casting the vote determines how important the vote itself is. Google calculates a page's importance from the votes cast for it. How important each vote is is taken into account when a page's PageRank is calculated.

PageRank is Google's way of deciding a page's importance. It matters because it is one of the factors that determines a page's ranking in the search results. It isn't the only factor that Google uses to rank pages, but it is an important one.

From here on in, we'll occasionally refer to PageRank as "PR".

Notes:

Not all links are counted by Google. For instance, they filter out links from known link farms. Some links can cause a site to be penalized by Google. They rightly figure that webmasters cannot control which sites link to their sites, but they can control which sites they link out to. For this reason, links into a site cannot harm the site, but links from a site can be harmful if they link to penalized sites. So be careful which sites you link to. If a site has PR0, it is usually a penalty, and it would be unwise to link to it.

[\[TOP\]](#)

How is PageRank calculated?

To calculate the PageRank for a page, all of its inbound links are taken into account. These are links from within the site and links from outside the site.

$$PR(A) = (1-d) + d(PR(t1)/C(t1) + \dots + PR(tn)/C(tn))$$

That's the equation that calculates a page's PageRank. It's the original one that was published when PageRank was being developed, and it is probable that Google uses a variation of it but they aren't telling us what it is. It doesn't matter though, as this equation is good enough.

In the equation 't1 - tn' are pages linking to page A, 'C' is the number of outbound links that a page has and 'd' is a damping factor, usually set to 0.85.

We can think of it in a simpler way:-

$$\text{a page's PageRank} = 0.15 + 0.85 * (\text{a "share" of the PageRank of every page that links to it})$$

"share" = the linking page's PageRank divided by the number of outbound links on the page.

A page "votes" an amount of PageRank onto each page that it links to. The amount of PageRank that it has to vote with is a little less than its own PageRank value (its own value * 0.85). This value is shared equally between all the pages that it links to.

From this, we could conclude that a link from a page with PR4 and 5 outbound links is worth more than a link from a page with PR8 and 100 outbound links. The PageRank of a page that links to yours is important but the number of links on that page is also important. The more links there are on a page, the less PageRank value your page will receive from it.

If the PageRank value differences between PR1, PR2,.....PR10 were equal then that conclusion would hold up, but many people believe that the values between PR1 and PR10 (the maximum) are set on a logarithmic scale, and there is very good reason for believing it. Nobody outside Google

knows for sure one way or the other, but the chances are high that the scale is logarithmic, or similar. If so, it means that it takes a lot more additional PageRank for a page to move up to the next PageRank level than it did to move up from the previous PageRank level. The result is that it reverses the previous conclusion, so that a link from a PR8 page that has lots of outbound links is worth more than a link from a PR4 page that has only a few outbound links.

Whichever scale Google uses, we can be sure of one thing. A link from another site increases our site's PageRank. Just remember to avoid links from link farms.

Note that when a page votes its PageRank value to other pages, its own PageRank is not reduced by the value that it is voting. The page doing the voting doesn't give away its PageRank and end up with nothing. It isn't a transfer of PageRank. It is simply a vote according to the page's PageRank value. It's like a shareholders meeting where each shareholder votes according to the number of shares held, but the shares themselves aren't given away. Even so, pages do lose some PageRank indirectly, as we'll see later.

Ok so far? Good. Now we'll look at how the calculations are actually done.

For a page's calculation, its existing PageRank (if it has any) is abandoned completely and a fresh calculation is done where the page relies solely on the PageRank "voted" for it by its current inbound links, which may have changed since the last time the page's PageRank was calculated.

The equation shows clearly how a page's PageRank is arrived at. But what isn't immediately obvious is that it can't work if the calculation is done just once. Suppose we have 2 pages, A and B, which link to each other, and neither have any other links of any kind. This is what happens:-

Step 1: Calculate page A's PageRank from the value of its inbound links

Page A now has a new PageRank value. The calculation used the value of the inbound link from page B. But page B has an inbound link (from page A) and its new PageRank value hasn't been worked out yet, so page A's new PageRank value is based on inaccurate data and can't be accurate.

Step 2: Calculate page B's PageRank from the value of its inbound links

Page B now has a new PageRank value, but it can't be accurate because the calculation used the new PageRank value of the inbound link from page A, which is inaccurate.

It's a Catch 22 situation. We can't work out A's PageRank until we know B's PageRank, and we can't work out B's PageRank until we know A's PageRank.

Now that both pages have newly calculated PageRank values, can't we just run the calculations again to arrive at accurate values? No. We can run the calculations again using the new values and the results will be more accurate, but we will always be using inaccurate values for the calculations, so the results will always be inaccurate.

The problem is overcome by repeating the calculations many times. Each time produces slightly more accurate values. In fact, total accuracy can never be achieved because the calculations are always based on inaccurate values. 40 to 50 iterations are sufficient to reach a point where any further iterations wouldn't produce enough of a change to the values to matter. This is precisely what Google does at each update, and it's the reason why the updates take so long.

One thing to bear in mind is that the results we get from the calculations are proportions. The figures must then be set against a scale (known only to Google) to arrive at each page's actual PageRank. Even so, we can use the calculations to channel the PageRank within a site around its pages so that certain pages receive a higher proportion of it than others.

NOTE:

You may come across explanations of PageRank where the same equation is stated but the result of each iteration of the calculation is added to the page's existing PageRank. The new value (result +

existing PageRank) is then used when sharing PageRank with other pages. These explanations are wrong for the following reasons:-

1. They quote the same, published equation - but then change it from $PR(A) = (1-d) + d(\dots)$ to $PR(A) = PR(A) + (1-d) + d(\dots)$

It isn't correct, and it isn't necessary.

2. We will be looking at how to organize links so that certain pages end up with a larger proportion of the PageRank than others. Adding to the page's existing PageRank through the iterations produces different proportions than when the equation is used as published. Since the addition is not a part of the published equation, the results are wrong and the proportioning isn't accurate.

According to the published equation, the page being calculated starts from scratch at each iteration. It relies solely on its inbound links. The 'add to the existing PageRank' idea doesn't do that, so its results are necessarily wrong.

[\[TOP\]](#)

Internal linking

Fact: A website has a maximum amount of PageRank that is distributed between its pages by internal links.

The maximum PageRank in a site equals the number of pages in the site * 1. The maximum is increased by inbound links from other sites and decreased by outbound links to other sites. We are talking about the overall PageRank in the site and not the PageRank of any individual page. You don't have to take my word for it. You can reach the same conclusion by using a pencil and paper and the equation.

Fact: The maximum amount of PageRank in a site increases as the number of pages in the site increases.

The more pages that a site has, the more PageRank it has. Again, by using a pencil and paper and the equation, you can come to the same conclusion. Bear in mind that the only pages that count are the ones that Google knows about.

Fact: By linking poorly, it is possible to fail to reach the site's maximum PageRank, but it is not possible to exceed it.

Poor internal linkages can cause a site to fall short of its maximum but no kind of internal link structure can cause a site to exceed it. The only way to increase the maximum is to add more inbound links and/or increase the number of pages in the site.

Cautions: Whilst I thoroughly recommend creating and adding new pages to increase a site's total PageRank so that it can be channeled to specific pages, there are certain types of pages that should not be added. These are pages that are all identical or very nearly identical and are known as cookie-cutters. Google considers them to be spam and they can trigger an alarm that causes the pages, and possibly the entire site, to be penalized. Pages full of good content are a must.

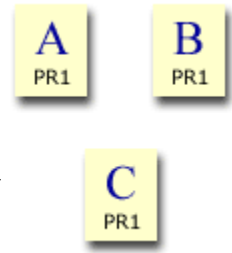
What can we do with this 'overall' PageRank?

We are going to look at some example calculations to see how a site's PageRank can be manipulated, but before doing that, I need to point out that a page will be included in the Google index only if one or more pages on the web link to it. That's according to Google. If a page is not in the Google index, any links from it can't be included in the calculations.

For the examples, we are going to ignore that fact, mainly because other 'Pagerank Explained' type

documents ignore it in the calculations, and it might be confusing when comparing documents. The [calculator](#) operates in two modes:- Simple and Real. In Simple mode, the calculations assume that all pages are in the Google index, whether or not any other pages link to them. In Real mode the calculations disregard unlinked-to pages. These examples show the results as calculated in Simple mode.

Let's consider a 3 page site (pages A, B and C) with no links coming in from the outside. We will allocate each page an initial PageRank of 1, although it makes no difference whether we start each page with 1, 0 or 99. Apart from a few millionths of a PageRank point, after many iterations the end result is always the same. Starting with 1 requires fewer iterations for the PageRanks to converge to a suitable result than when starting with 0 or any other number. You may want to use a pencil and paper to follow this or you can follow it with the [calculator](#).



The site's maximum PageRank is the amount of PageRank in the site. In this case, we have 3 pages so the site's maximum is 3.

At the moment, none of the pages link to any other pages and none link to them. If you make the calculation once for each page, you'll find that each of them ends up with a PageRank of 0.15. No matter how many iterations you run, each page's PageRank remains at 0.15. The total PageRank in the site = 0.45, whereas it could be 3. The site is seriously wasting most of its potential PageRank.

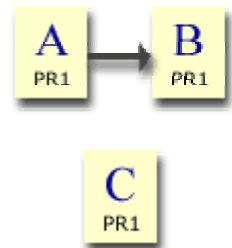
Example 1

Now begin again with each page being allocated PR1. Link page A to page B and run the calculations for each page. We end up with:-

Page A = 0.15

Page B = 1

Page C = 0.15



Page A has "voted" for page B and, as a result, page B's PageRank has increased.

This is looking good for page B, but it's only 1 iteration - we haven't taken account of the Catch 22 situation. Look at what happens to the figures after more iterations:-

After 100 iterations the figures are:-

Page A = 0.15

Page B = 0.2775

Page C = 0.15

It still looks good for page B but nowhere near as good as it did. These figures are more realistic. The total PageRank in the site is now 0.5775 - slightly better but still only a fraction of what it could be.

NOTE:

Technically, these particular results are incorrect because of the special treatment that Google gives to [dangling links](#), but they serve to demonstrate the simple calculation.

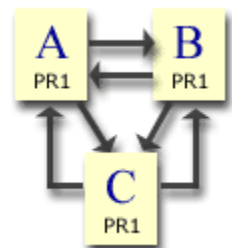
Example 2

Try this linkage. Link all pages to all pages. Each page starts with PR1 again. This produces:-

Page A = 1

Page B = 1

Page C = 1



Now we've achieved the maximum. No matter how many iterations are run, each page always ends up with PR1. The same results occur by linking in a loop. E.g.

A to B, B to C and C to D. [View this in the calculator.](#)

This has demonstrated that, by poor linking, it is quite easy to waste PageRank and by good linking, we can achieve a site's full potential. But we don't particularly want all the site's pages to have an equal share. We want one or more pages to have a larger share at the expense of others. The kinds of pages that we might want to have the larger shares are the index page, hub pages and pages that are optimized for certain search terms. We have only 3 pages, so we'll channel the PageRank to the index page - page A. It will serve to show the idea of channeling.

Example 3

Now try this. Link page A to both B and C. Also link pages B and C to A.

Starting with PR1 all round, after 1 iteration the results are:-

Page A = 1.85

Page B = 0.575

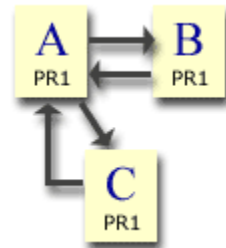
Page C = 0.575

and after 100 iterations, the results are:-

Page A = 1.459459

Page B = 0.7702703

Page C = 0.7702703



In both cases the total PageRank in the site is 3 (the maximum) so none is being wasted. Also in both cases you can see that page A has a much larger proportion of the PageRank than the other 2 pages. This is because pages B and C are passing PageRank to A and not to any other pages. We have channeled a large proportion of the site's PageRank to where we wanted it.

Example 4

Finally, keep the previous links and add a link from page C to page B. Start again with PR1 all round. After 1 iteration:-

Page A = 1.425

Page B = 1

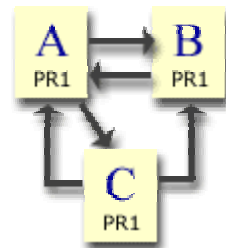
Page C = 0.575

By comparison to the 1 iteration figures in the previous example, page A has lost some PageRank, page B has gained some and page C stayed the same. Page C now shares its "vote" between A and B. Previously A received all of it. That's why page A has lost out and why page B has gained. and after 100 iterations:-

Page A = 1.298245

Page B = 0.9999999

Page C = 0.7017543



When the dust has settled, page C has lost a little PageRank because, having now shared its vote between A and B, instead of giving it all to A, A has less to give to C in the A-->C link. So adding an extra link from a page causes the page to lose PageRank indirectly if any of the pages that it links to return the link. If the pages that it links to don't return the link, then no PageRank loss would have occurred. To make it more complicated, if the link is returned even indirectly (via a page that links to a page that links to a page etc), the page will lose a little PageRank. This isn't really important with internal links, but it does matter when linking to pages outside the site.

Example 5: new pages

Adding new pages to a site is an important way of increasing a site's total PageRank because each new page will add an average of 1 to the total. Once the new pages have been added, their new PageRank can be channeled to the important pages. We'll use the calculator to demonstrate these.

Let's add 3 new pages to Example 3 [\[view\]](#). Three new pages but they don't do anything for us yet. The small increase in the Total, and the new pages' 0.15, are unrealistic as we shall see. So let's link them into the site.

Link each of the new pages to the important page, page A [\[view\]](#). Notice that the Total PageRank has doubled, from 3 (without the new pages) to 6. Notice also that page A's PageRank has almost doubled.

There is one thing wrong with this model. The new pages are orphans. They wouldn't get into Google's index, so they wouldn't add any PageRank to the site and they wouldn't pass any PageRank to page A. They each need to be linked to from at least one other page. If page A is the important page, the best page to put the links on is, surprisingly, page A [\[view\]](#). You can play around with the links but, from page A's point of view, there isn't a better place for them.

It is not a good idea for one page to link to a large number of pages so, if you are adding many new pages, spread the links around. The chances are that there is more than one important page in a site, so it is usually suitable to spread the links to and from the new pages. You can use the calculator to experiment with mini-models of a site to find the best links that produce the best results for its important pages.

Examples summary

You can see that, by organising the internal links, it is possible to channel a site's PageRank to selected pages. Internal links can be arranged to suit a site's PageRank needs, but it is only useful if Google knows about the pages, so do try to ensure that Google spiders them.

Inbound and Outbound links

Examples of these could be given but it is probably clearer to read about them (below) and to 'play' with them in the [calculator](#).
