

PageRank Explained

or ***"Everything you've always wanted to know about PageRank"***

©2001 All Rights Reserved

Written and theorised by Chris Ridings, owner of <http://www.searchenginesystems.net/>

Edited by Jill Whalen, owner of HighRankings.com and co-moderator of the Rank Write Roundtable. <http://www.rankwrite.com/>

You may reproduce this document or theories in whole or in part provided that you credit such reproduction to "Chris Ridings of www.searchenginesystems.net"

VERSION 1.1 Last amended – 9th November 2001

Introduction

This document details my understanding of and theories relating to Google's PageRank. For those who don't know me, I develop custom search engines. I therefore have a programmer's insight into search engine algorithms, how they work and what they can and cannot do. Because of this, I'm able to deduce much about the way PageRank works. I believe that the information in this document is as accurate as it can be. Nobody knows for sure all the specifics of PageRank, except for Google themselves. Feel free to question any logic that you cannot follow; it's through communication that the soundness of my theories can be strengthened. Please contact me at: chris@searchenginesystems.net with your questions or comments.

Enough intro...let's get started with what this document is about - PageRank!

What is PageRank?

PageRank is Google's method of measuring a page's "importance." When all other factors such as Title tag and keywords are taken into account, Google uses PageRank to adjust results so that sites that are more "important" will move up in the results page of a user's search accordingly.

That is, the order of ranking in Google works like this:

- 1) Find all pages matching the keywords of the search.
- 2) Rank accordingly using "on the page factors" such as keywords.
- 3) Calculate in the inbound anchor text.
- 4) Adjust the results by PageRank scores.

How is PageRank determined?

The Google theory goes that if Page A links to Page B, then Page A is saying that Page B is an important page. The actual text of the link is irrelevant to PageRank.

PageRank also factors in the importance of the links to a page. If a page has more important links to it then its links to other pages also become more important.

How significant is PageRank?

The significance of any one factor in search engine algorithms depends on the quality of the information it supplies. So it makes sense to look at this quality of information first.

When Google was just a little Googlet in nappies, it was probably fair to say that a link was an accurate indicator of a recommendation. However, nowadays, that is no longer the case, for two very good reasons:

- 1) The Internet has changed significantly. A link these days is just as likely to be a related site, a licensing requirement or a return of a favour (such as a reciprocal link), rather than a true recommendation.
- 2) As soon as you create a search engine that views links as recommendations, people will begin to try to influence those links. As soon as they influence them - they are no longer recommendations.

Therefore, the reliability of information a link supplies, is not necessarily that good, and is ever decreasing. This accounts for PageRank's now low, and ever decreasing importance in the Google ranking algorithm.

However, PageRank still has one redeeming factor. It's more difficult to influence than any other ranking consideration. Which means it has the potential to give you an advantage over competitors when used in combination with other search engine optimisation techniques. I warn you now though; there are no short cuts. To use PageRank effectively, you'll need to understand it completely; otherwise you'll probably be using your time to no advantage.

A few basic facts about PageRank

To understand the rest of this document, there are a few facts about it you need to know about PageRank.

- 1) PageRank is a number that assesses solely the voting ability of all incoming links to a page, and how much they recommend that page.
- 2) Every unique page of a site that is indexed in Google has a PageRank. People often, mistakenly, think of the PageRank of a site being the PageRank of that site's home page.
- 3) Internal site links *do* count in passing PageRank to other pages of the site.
- 4) PageRank stands on its own; It's not tied in with the anchor text (titling) of links, etc. Sure, they're related, but saying they're the same thing is like saying Title tags are the same as keywords in text.

How can you tell what a page's PageRank is?

You can get a toolbar for Internet Explorer from <http://toolbar.google.com>. Once installed, there will be a bar graph at the top of Internet Explorer showing a version of PageRank for the page you're browsing. When you hold the mouse over the bar, you get a number from zero to ten. (If you don't see the number, you may have an older version of the toolbar installed. Once you completely uninstall it, reboot your computer and reinstall the latest version, you should be able to see the number.)

How accurate is the Google toolbar?

The Google toolbar is not very accurate in telling you the PageRank of a site, but it's the only thing right now that can really give you any idea. As long as you know the toolbar's limitations, then at least you know what you are viewing. There are two limitations to the Google toolbar:

1. **The toolbar sometimes guesses.** If you enter a page, which is not in its index, but where there is a page that is very close to it in Google's index, then it will provide a guesstimate of the PageRank. This guesstimate is worthless for our purposes because it isn't featured in any of the PageRank calculations. The only way to tell if the toolbar is using a guesstimate is to type the URL into the Google search box and see if the page comes up. If it doesn't, then it's guessing!
2. **The toolbar is just a *representation* of actual PageRank.** Whilst PageRank is linear, they've chosen to use a non-linear graph to show it. So on the toolbar, to move from a PageRank of 2 to a PageRank of 3 takes less of an increase than to move from a PageRank of 3 to a PageRank of 4. This is best illustrated by a comparison table, the actual figures are kept secret so we'll just use any figures for demonstration purposes:

If the actual PageRank is between	The Toolbar Shows
0.00000001 and 5	1
6 and 25	2
25 and 125	3
126 and 625	4
626 and 3125	5
3126 and 15625	6
15626 and 78125	7
78126 and 390625	8
390626 and 1953125	9
1953126 and infinity	10

Hopefully you can see from this demonstration how restricted the information is that you get from the toolbar.

From here on in I'm going to use the terms Actual PR to represent the actual PageRank value stored by Google, and Toolbar PR to represent the rather poor representation that the Google Toolbar permits us to see.

The calculation of PageRank

Having explained what PageRank is, i.e., what you're seeing when you get information about it and how important it is...in this section I'll tell you approximately how it's calculated. It's not crucial to know this; however, if you understand this you'll better understand how to properly use it.

When Google was just a research project, they wrote a paper detailing a formula that gives the PageRank for a page. Whilst they may not still be using this exact formula, it seems pretty accurate for today's purposes. Here it is ...¹

$$PR(A) = (1-d) + d (PR(T1)/C(T1) + \dots + PR(Tn)/C(Tn))$$

Where PR(A) is the PageRank of Page A (the one we want to work out).

D is a dampening factor. Nominally this is set to 0.85

PR(T1) is the PageRank of a site pointing to Page A

C(T1) is the number of links off that page

PR(Tn)/C(Tn) means we do that for each page pointing to Page A

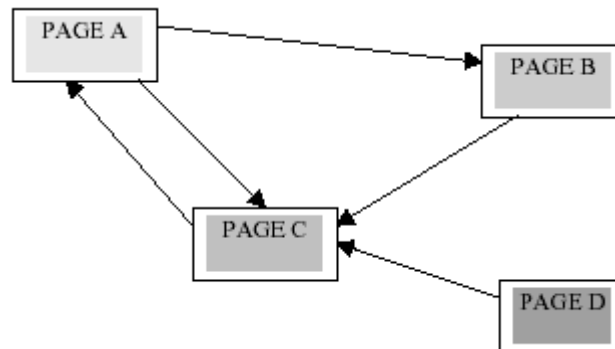
Yikes! So for those of you that aren't mathematicians, here's the low-down on that formula – you can't simply calculate PageRank in one go like that. To calculate the PageRank of Page A you'd need to know the PageRank of all the pages pointing to Page A. Their PageRanks would in part be due to Page A pointing to them or some other site that points to them! What a silly formula. What it does tell us is one very important thing about the PageRank of any page...

The PageRank given to Page A by a Page B pointing to it is decreased with each link to anywhere that exists on Page B. That means a page's PageRank is essentially a measure of its vote; it can split that vote between one link or two links or many more, but it's overall voting power will always be the same.

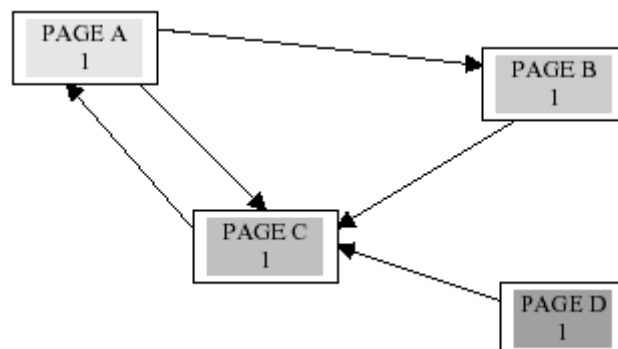
¹ Source: The Anatomy of a Large-Scale Hypertextual Web Search Engine, Sergey Brin and Lawrence Page, <http://www-db.stanford.edu/~backrub/google.html>

Now, put the formula out of your mind for a while because it's far easier to view an example of an implementation, which is very similar to PageRank. This should help us to better understand it. We'll call it MiniRank.

For this example, we have four pages – imaginatively titled Page A, Page B, Page C and Page D. They link to each other as shown in the following diagram.



To begin with we don't know what the MiniRanks are for the pages, so we'll just assign one. For simplicity, we'll choose the number one. So the diagram with the MiniRanks on it becomes...



Easy so far! Now remember the rules about passing rank on. First we apply the dampening factor. (The dampening factor basically says that a page cannot vote another page to be as equally important as it is. This means that pages that are harder to get to in the web are less important.) Then we divide the remaining score by the number of links. We total up the entire ranking that should be added to each and every page before we finally add it on.

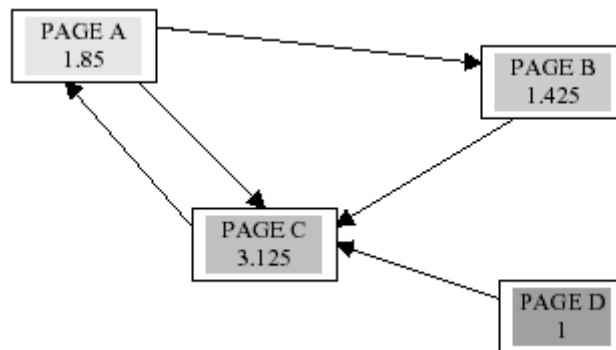
So, looking at Page A first, the amount of MiniRank available to pass on, after dampening it down, is $1 * 0.85 = 0.85$. There's two links out, so at the end of the process we're going to add 0.425 to Page B's MiniRank and 0.425 to Page C's MiniRank. We can't do it until we've calculated all the page's links because it would affect the results.

On to Page B. It has just one link. So it'll pass on $1 * 0.85 = 0.85$ to Page C when we've done all the link calculations.

Page C also only has one link. So it'll pass on $1 * 0.85 = 0.85$ to Page A.

Page D has one link so it passes 0.85 to Page C.

Now we can add all those totals on to all the pages.



The new MiniRank totals show how important Page C is. But we're not done yet. Because they all started out on the same value, we've only really calculated link popularity. The essence of PageRank and MiniRank is that better linked pages should get bigger votes; therefore we have to do the same thing again! This time Page C has more influence, because its current MiniRank is higher.

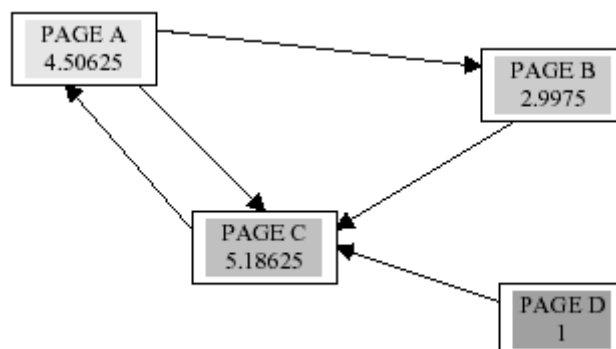
So, let's look at Page A first. Its current MiniRank is 1.85. The amount of MiniRank available to pass on, after dampening it down, is $1.85 * 0.85 = 1.5725$. There's two links out so at the end of the process we're going to add 0.78625 to Page B's MiniRank and 0.78625 to Page C's MiniRank.

On to Page B. It has just one link. So it'll pass on $1.425 * 0.85 = 1.21125$ to Page C when we've done all the link calculations.

Page C also only has one link, but a whopping 3.125 MiniRank. So it'll pass on $3.125 * 0.85 = 2.65625$ to Page A.

Page D has one link so passes 0.85 to Page C.

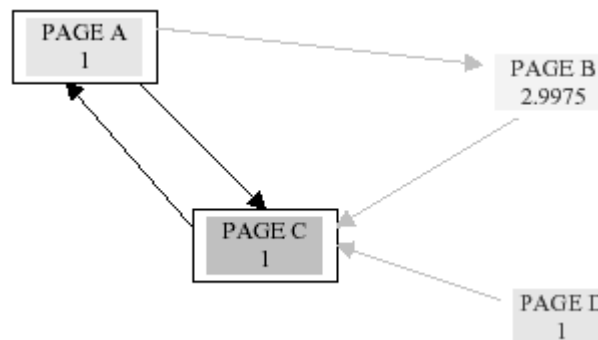
We get...



We can already see what we'd expect; Page C has the highest MiniRank and Page A has the next highest. In practice we'd need to do the equivalent of this 50 to 100 times to ensure that the poor accuracy of the earlier iterations is averaged out and minimalised. Simple!

PageRank Feedback

But hang on a minute! There's something going on between pages A and C, so let's take another look.



In one run of the calculation, Page C is giving Page A a boost in MiniRank (PageRank). In the next run it gets itself an increase in MiniRank that's a proportion of Page A's new improved MiniRank (it's getting a proportion of its MiniRank fed back to it!).

That's PageRank feedback. One would think that Google must discount this kind of link, particularly if Page A and Page C exist on the same site. In fact, I've read several people express the opinion that they must. The truth is, Google cannot do this. Imagine doing a calculation over millions of pages instead of 4...just how do you work out when this feedback is happening and get rid of it. And even if you could, how do you negate the effect that this would have on the rest of the system? PageRank feedback is an essential part of the system! In fact, it's critical to PageRank operating properly, and is part of the way PageRank works.

Influencing the results

Knowing how it works and that Google does influence the PageRank results in some cases, we can decide exactly what Google can do:

Before we start to calculate PageRank let's say a site's links are particularly good. Let's say page B is a page on Yahoo or DMOZ (both of which show this kind of effect), instead of setting their starting value to 1 we can set it to 100 or some higher number. By doing this we're saying that Google tweaks up the PageRanks that are a result of this page by just a fraction.

We can do the reverse, but only to a lesser extent. Let's say Page B is known spam. If we set its starting PageRank to zero then its PageRank will start out having no influence, but will start to gain influence soon as long as it has other sites linking to it.

Keep this in mind: ***With PageRank we can easily tweak up the importance of a page's links by any amount we want, however, the reverse is not true – PageRank severely impedes the ability to tweak down the importance of a page's links.***

This is almost certainly what happens with Yahoo and DMOZ listed sites. Every page within Yahoo or DMOZ seems to be tweaked up so that sites listed in either of these two directories get a nice little increase in PageRank.

Now couldn't they adjust Page B's PageRank after every iteration of the calculation? Well they could, but Google works with millions of pages and would have to adjust each such page every time. It would make things veerrrrryyyy veerrrrryyyy low.

So how about setting a page's PageRank after everything has been calculated and final PageRanks have arrived. Well yes, they can and do do that. This has less to do with links though, than with changing individual results. So let's say the Google home page isn't high enough for Google. They can simply change it. Or if a Google search results page has a PageRank, they can take this off. This is a post-processing step.

Note that there's little point in Google using this to eliminate spammers from the index, however. Please do not assume that if your PageRank has suddenly become zero, then Google must have used this method to set it that way. It's far easier for them to just ban the page entirely. Banning is also more logical, because it removes the influence that your page would otherwise have had in the PageRank calculation process. The zero PageRank is most certainly due to some other factor such as a temporary calculation issue.

What this all means?

PageRank is the hardest factor to manipulate when optimising your pages. Although its effect is not as great as some believe - if you can get it right then you have a fair advantage over your competitors. PageRank is both hard to achieve and harder to catch up with. The information below really takes this to the extremes. In practice you could use all or part depending on how competitive you feel, and how strong your competition is.

There are three fundamental areas to look at and possibly change, when trying to optimise your PageRank:

1. The links you choose to have link to you, i.e., which ones you pick and how much effort you put into getting them.
2. Who you choose to link out to from your site and from which page of your site you place their link.

3. The internal navigational structure and linkage of your pages, in order to create maximum PageRank feedback.

Links to Your Site

When looking for links *to* your site, from a purely PageRank point of view, one might *think* they should simply look for pages that have the highest Toolbar PageRank. (Whilst keeping in mind that every page of a site has its own PageRank, so you must consider the PageRank of the "links" page, or whatever page the actual link will exist on.) **However, this way of thinking is incorrect.** If you've not just jumped to this section then you'll probably have worked out why that is. The PageRank given by a link is far more complicated than this simplification. There may have been a time when that was an okay approximation...but no more. As more and more people try and get links from only high PageRank sites, it becomes less and less of a winning proposition.

The actual PageRank from an individual page is shared out amongst the links on that page (remember the MiniRank calculations?). So, links from pages that have the same PageRank aren't always created equal. It depends on how many other links your link is sharing the links page with. For instance, a link from a page with a PageRank of 4 might be better than a link from a page with a PageRank of 6 if there are less total links on the PR 4 page. It's possible that a page with a PR of 2 might even be better to request a link from, than a page with a PR of 7. Right now there just isn't enough information available to allow us to know to what extent this stretches. However, it's significant enough to make it pointless to just choose high PageRanked sites, as your main linking strategy. There's also another, more matter-of-fact reason why that type of linking strategy might not be the best; sites with high PageRanks are often fussy about which sites they will link out to, making them harder to get linked from, than lower PageRanked sites. However, sites struggling with their own PageRank numbers should be more receptive to exchanging reciprocal links with other like-minded sites.

Now let's factor in feedback. Let's say, for example, that there are two separate pages on other people's sites which both have a PageRank of 4. Both of these have ten links to other pages. But your page that you want them to link to already has a link to the page on the second site. By getting a link from the second site you're generating feedback and getting more PageRank than if you had gotten a link from the first site! That's an over simplification; in fact, feedback loops can get even more complicated. Remember, the number of links on the page linking to you will alter the amount of feedback, etc.

Can you work it all out for a given page's situation? No – neither can I. My advice, therefore, is this – get links from sites that seem appropriate and have good quality, regardless of their current PageRank. If they are relevant to your site, and are high quality sites, they will either help your PageRank now, or will do so in the future. To really get your PageRank humming get yourself a listing in DMOZ and Yahoo to enjoy the artificially enhanced PageRank that they provide.

Links Out From Your Site

To consider the best linking *out* strategy, we first need to consider the links pointing *in* to your site. By which I mean we need to assume you have some links pointing to your site from directories such as DMOZ and Yahoo that are giving it nice bit of PageRank. Using the internal pages of your site, you can control feedback far better than you can with links *to* external pages. This leads to a rule...

Generally, you will want to keep PageRank within your own site.

This means you will only want to link out from a Page on your site that has a low PageRank itself, and which also contains a significant number of internal links (i.e., links pointing to other pages of your site).

Then, when you do link out – you give preference to those pages which either link to your site a page above your links page, or which link to a page that links to a page above your links page (i.e., you will get a better increase in PageRank if the links from external sites do not point to your links page).

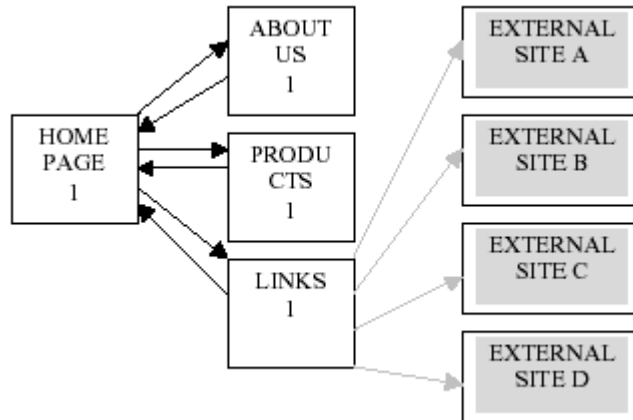
How can we do this? One way would be by writing reviews of the sites we link out to on a separate page of our site, and by providing a link to those reviews along with each hyperlink to the external site. Optionally, it would be okay if these pages open in another window but DO NOT do this in Javascript, because the search engine spiders can't follow javascripted links.

For example, we can do something like this with each link to an external site:

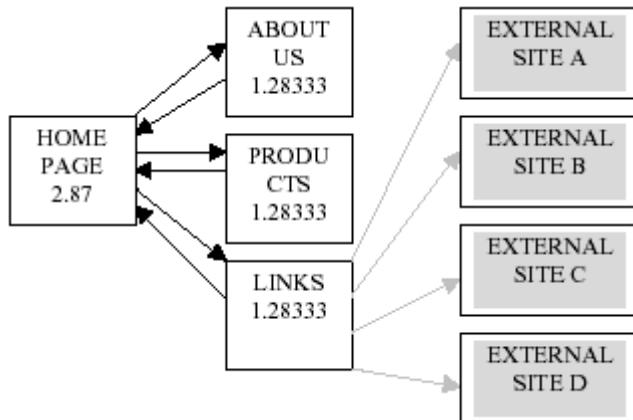
```
<a href=http://www.searchenginesystems.net>Search Engine Systems are the best search engine people in the world</a> <a href="/sesreview.htm">Read my flattering review of them here</a>.
```

Make sure that the review page links back to a page that is high up in your site's structure (It's best if this is your home page, but any important page will do.) By doing this, we've significantly reduced the amount of PageRank you've let out of the your site and ensured that the greater amount of PageRank that's left also gets multiplied by the feedback effect! We've targeted this feedback to the home page to ensure that less is passed back through your links page (which would be a wasted opportunity), and more is put elsewhere in your site. Your links page also needs to link to your home page and the other major pages of your site. However, place no other links on the review page (besides the home page link). It's **very** good if someone links to your review page, so in addition you may let the site know that you have reviewed them – it is quite possible you will get two links from their site (one to your site and one to the review of their site). All very complicated in text form so let's do a simplified example to show the principle and show its effect.

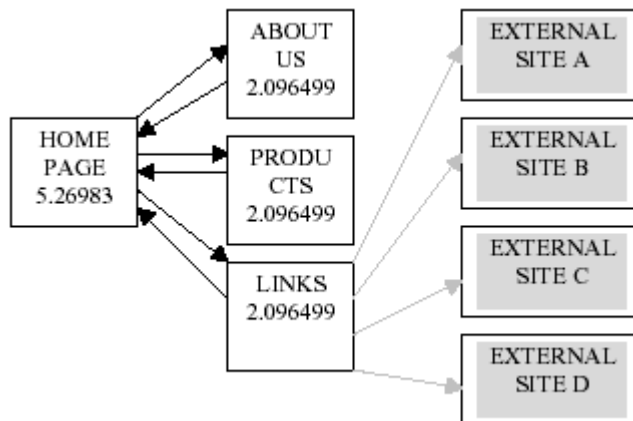
Our simple structure, with MiniRank start values set, is this...



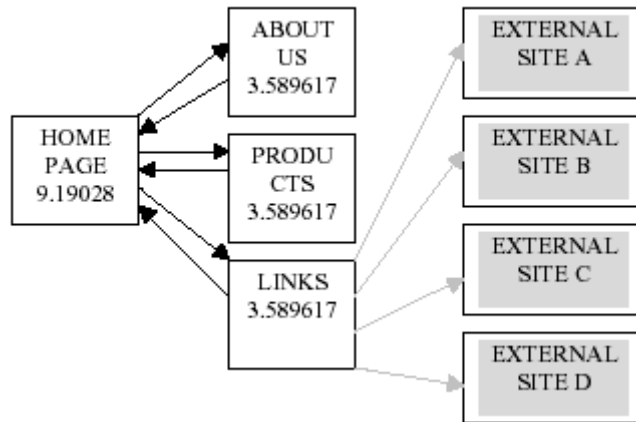
After the first iteration of calculation we get...



At the end of the second calculation we get...

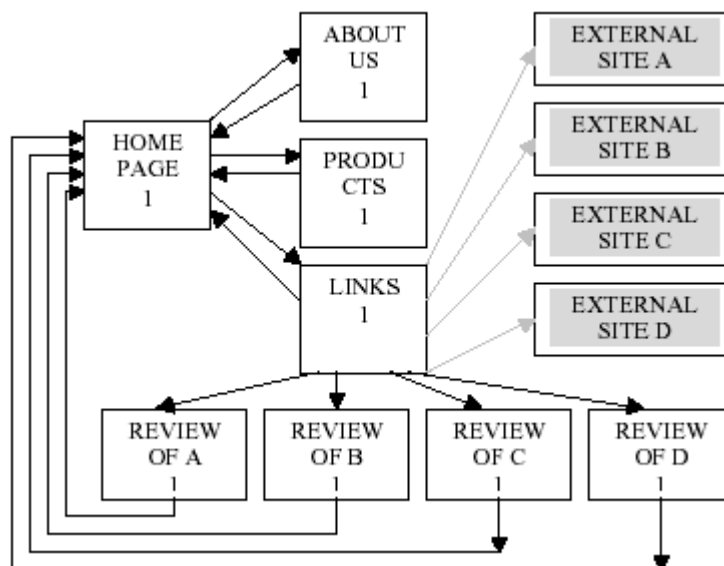


And at the end of the third calculation we get...

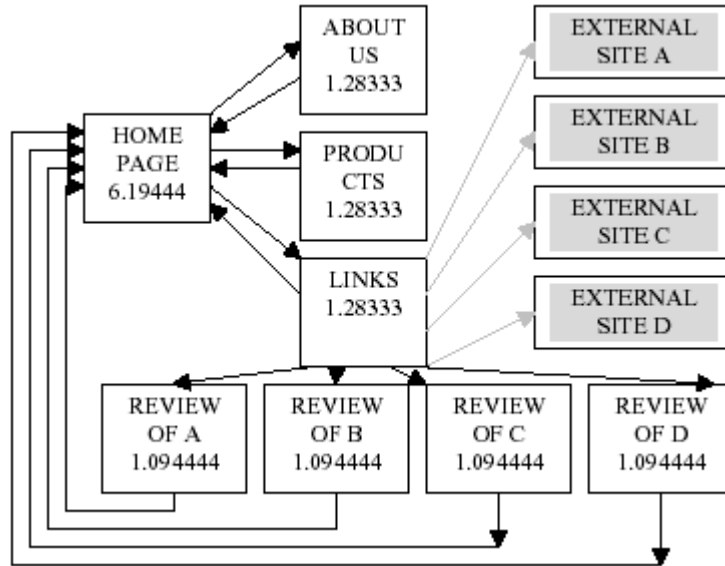


Total MiniRank in the site is: 19.959

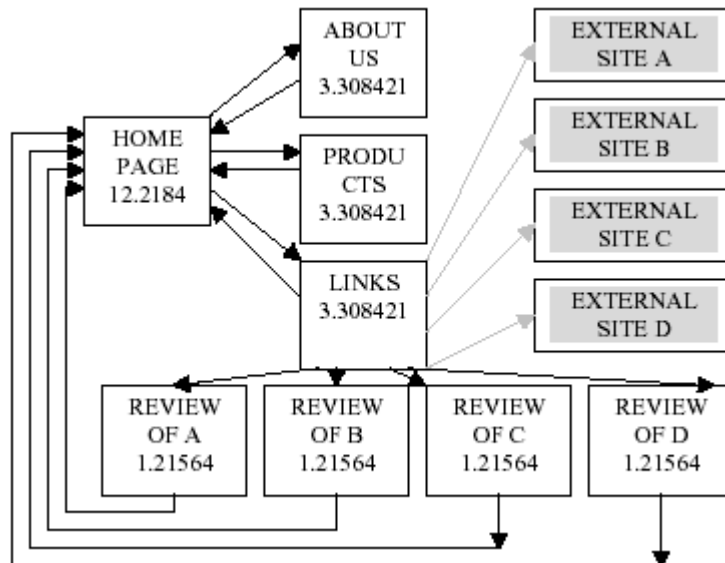
Now if we adjust the links to include the reviews pointed back to the home page we get...



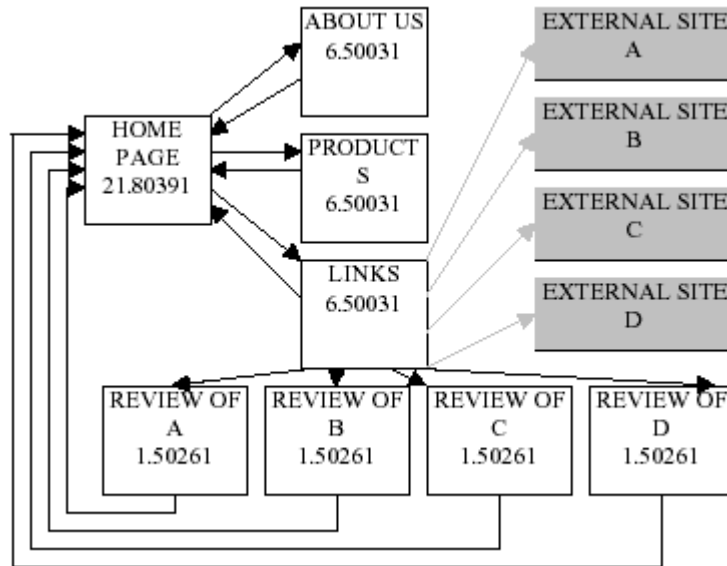
And after the first calculation we get ...



After the second calculation we get...



After the third calculation we get...



Total MiniRank in the site is: 47.31 (but we started with four more!). Some of this is the power of the extra pages and some the power of feedback. But in summary...

First Example	Second Example (with Reviews)
Num of pages = 4	Num of pages = 8
Starting MiniRank = 4	Starting MiniRank = 8
End MiniRank in site = 19.959	End MiniRank in site = 47.31
Homepage is 2.37 times more important using the second method	
Major Pages “About Us”, “Products” and “Links” are 1.8 times more important using the second method	

This nicely demonstrates the power of feedback. We are placing a portion of our links pages votes back into our sites system rather than letting it go to external links. This is why larger sites generally have a better PageRank than smaller sites. So why aren't you doing this already????!!!! Start writing reviews of the sites listed in your links page now! (Note, figures are just for demonstration purposes as a general indicator of the power of this technique – actual numbers will vary).

Internal Structure and Linkages

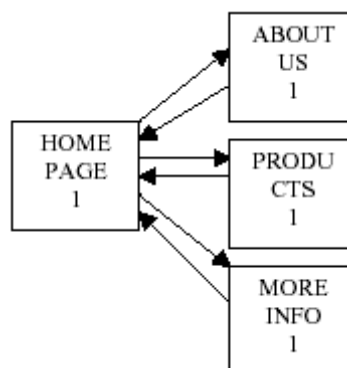
Having talked about linking out, it makes sense to talk about how the internal linking structure of your site also influences its own PageRank. Let's just refresh a couple of facts:

The more pages a particular site has in the Google index, the higher the site's total starting PageRank is, and the more PageRank it has to work with. Because each page is allocated the same starting value before the PageRank calculations are done, more pages can only be better! It should make sense that if we have more to start with then our feedback effect will also be more significant. Ever notice how larger sites tend to have better PageRank? The feedback effect helps explain why. Of course, your pages have to make sense and have good content to get into the index to begin with. (The reviews in the last section would be a good example.)

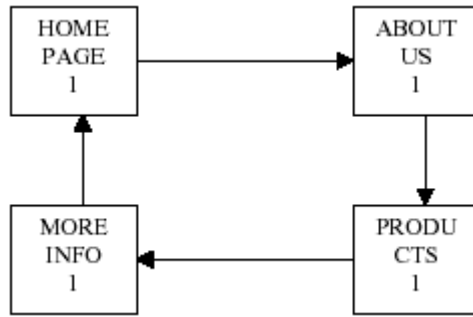
Feedback is a natural effect of the PageRank process. It takes place within internal site links, and is critical to Google's assessment of which pages are important within a site. If the site had no incoming or outgoing links, the structure of the site would provide the same amount of feedback. However when we factor in incoming and outgoing links the internal structure of a site is significant. For example, if the site has outgoing links on a page then we'll want to keep the PageRank of that page minimal.

There are three different ways in which pages can be interlinked within a web site. In practice, web sites might use a combination of these. Using a combination is fine and normal as long as you understand the different sections and how they are affecting your PageRank. For the purposes of this document we'll view the different linkage structures as separate entities. We have:

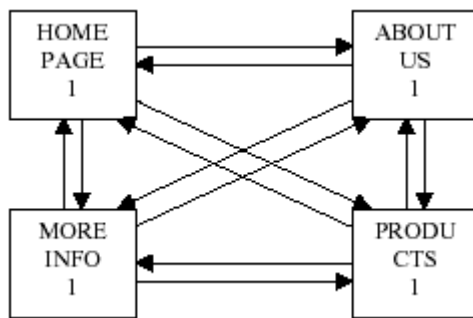
Hierarchical



Looping

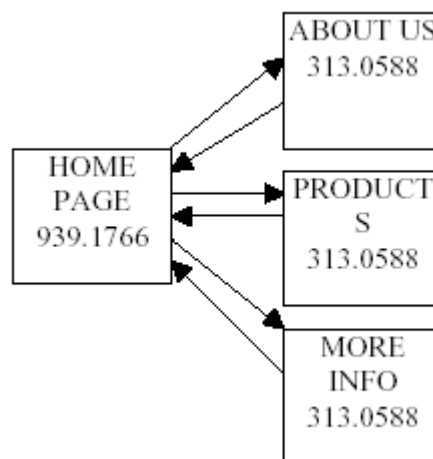


Extensive Interlinking

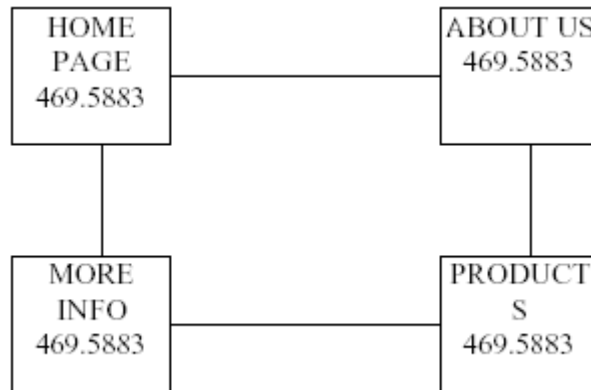


Keep in mind that we do not necessarily want the PageRank to be distributed evenly throughout the site. We want the maximum PageRank Feedback in the system, and we want that to be able to focus on particular pages (i.e., those in which we have optimized with keyword rich text, etc.). Since I've already extensively shown you the MiniRank calculation, I'll just show the results of each form of linkage structure after 10 runs of calculations:

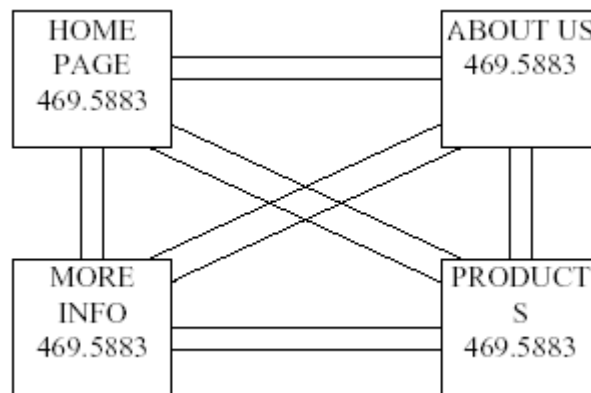
Heirarchical



Looping

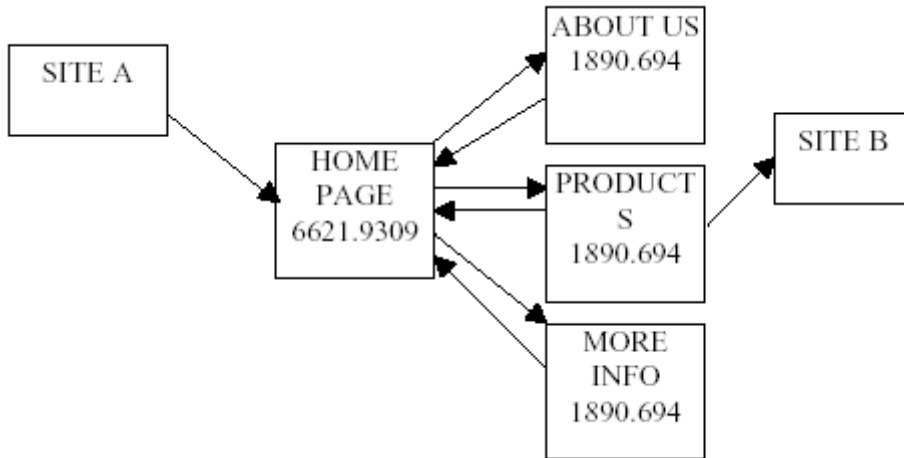


Extensive Interlinking

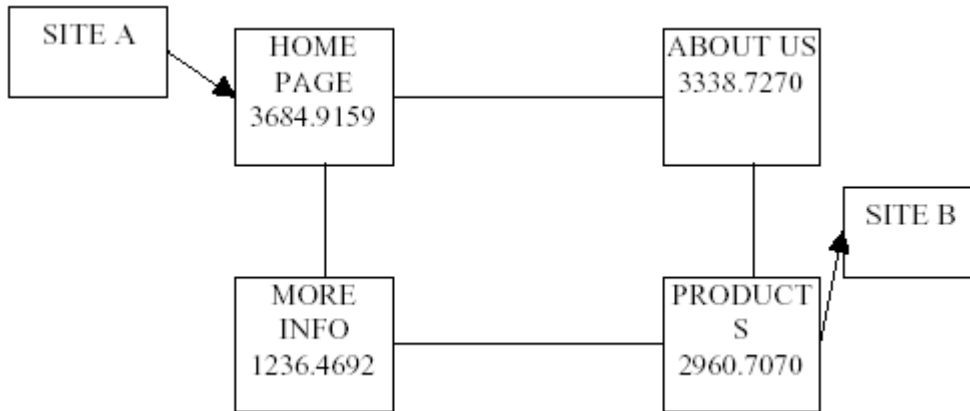


Note how the total amount of Minirank within the site is the same (1878.353). That's because there are not yet any external incoming or outgoing links. What's important is the distribution. The Hierarchical structure pushes more PageRank towards the home page (other sites are more likely to be linking to the home page, and this page is less likely to have outbound links). There's no apparent difference between the Looping structure and the Extensive Interlinking structure. Let's see what happens when we complicate the structure by adding external incoming and outgoing links...

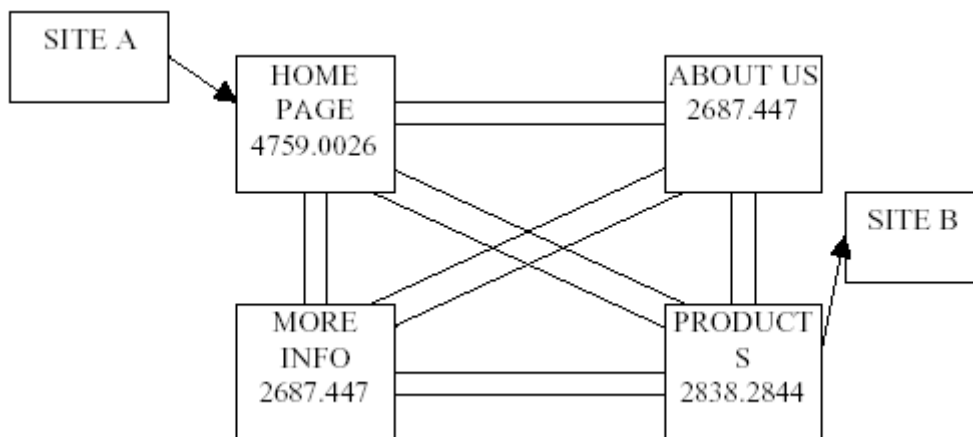
Hierarchical



Looping



Extensive Interlinking



Although these examples have only a few iterations of the formula calculated, they are already beginning to show the rules of interlinking within a site:

Extensive Interlinking provides a marginally better PageRank feedback than does Hierarchical, and both provide a marginally better PageRank feedback than Looping.

With a lot of Hierarchical interlinking there is a far greater level of PageRank assigned to pages higher up in the structure. This means we are giving away less PageRank on our outbound links.

What this means in practice is that you should combine these methods of linking. The rules are...

- 1. Where a group of pages may contain outward links - use a hierarchical structure.**
- 2. Where a group of pages do not contain outward links - use the extensive linking structure, but expand on this by including a link back to the home page.**
- 3. If a particular page is highly important - place it higher up a hierarchical structure.**

How to Use Your Site Map for PageRank Purposes

Many people believe that Site Maps help search engine spiders to crawl pages. I'm not convinced of this, but due to their popularity and the fact that they involve some nice anchor text links, let's examine how to best implement them from a PageRank perspective.

First, link to your SiteMap from your home page as you normally would.

Keep in mind that the presence of the SiteMap is effectively pulling down the PageRank of your other pages (particularly if it's linked from the highest PageRanked page of your site). So we want to make sure the Site Map page does two things:

1. Maximises your starting total (by adding more pages).
2. Feeds back as much PageRank as possible

The first is slightly controversial. To make the sitemap maximise your starting total, we need to break it down into several pages. You might be thinking, "Well, that makes it harder for the spider to crawl, right?" The answer is yes and no. Sometimes it's actually harder for spiders to pick hundreds of links off of one page. However,

any spider worth its salt nowadays indexes deep enough to cope with a sitemap spread over several pages.

Now let's deal with maximising your PageRank feedback. Every site map page should have a link back to your home page and all other significant pages on your site. If your sitemap page contains links to pages, which contain links to external sites, then you should be especially sure to minimise the amount of PageRank these pages will let trickle out of your site.

So, here's what you can do...

Break your sitemap down into categories, and give each category a different page of its own. Your sitemap page now becomes an index of these category pages (instead of a map of the entire site). Into each category you can place up to about 30 links. You should also give a description of what each page is about, along with the link. When you are placing a page that has external links into your sitemap categories, choose a category that has a large number of links on it.

Next, you must interlink each of your category pages together, along with the category index (your old sitemap page). To do this, simply place a navigational menu, which links directly to the category index page, or any of the other category pages on each of them. Be sure to also include a link to your home page and other major pages on each of these.

This maximises the feedback and keeps the PageRank of the sitemap pages low. The idea is to make the entire sitemap less of a PageRank drain to the site as a whole.

Final Word

PageRank is a highly complicated topic that is often misunderstood. I think it's worth re-stating a few things about this document and PageRank. This document is a work in progress and will probably remain so for a long period of time. There is, at this point in time, not enough information for us to be 100% certain about anything. I am merely presenting theories, based upon the best information available, which seem to largely hold true. When Google chooses to let us see PageRank information, they do so via the Google toolbar. When you look at the Google toolbar I hope you'll remember at least one line from earlier on – *“The Google toolbar is not very accurate in telling you the PageRank of a site, but it's the only thing right now that can really give you any idea.”*

PageRank has its place in the ranking process. That place is not as big as many might imagine. Its significance in the ranking algorithm is less than many other factors such as Title tags and anchor text. Optimising a site for PageRank alone will not get you good rankings. At PageRank's core, is the fact that it is difficult to manipulate. Therefore, if you do get a good PageRank, your competitors will find it hard to equal. Whether it's worth your time to focus heavily on PageRank is a personal decision that will depend on the level of your competition. I do believe that at a minimum, it's

always worth understanding how PageRank works and keeping it in mind whenever changes are made, or new sites are built, just as other factors such as anchor text and keywords are always in our thoughts.

This document has created a fair amount of queries and discussion among those that have read it so far. It's possible that at some point in the future, I'll create a list of frequently asked questions. Until then, however, those seeking further information can email me at chris@searchenginesystems.net or you might wish to take a look at a discussion resulting from an early release of this document at <http://www.ihelpyouservices.com/forums/t916/s.html>.