

“PageRank Uncovered”

Formerly PageRank Explained

©2002 All Rights Reserved

Written and theorised by Chris Ridings and Mike Shishigin

Edited by Jill Whalen

Technical Editing by Yuri Baranov

Profiles:

Chris Ridings is the original author of PageRank Explained. A former programmer of custom search engines, search engine expert, creator of the search engine optimization support forums (<http://www.supportforums.org>) and author of marketing software tools (<http://www.iebar.com/>). Chris uses his programming experience and knowledge of algorithms to get an insight into how search engines work “under the cover.”

Mike Shishigin, PhD, is a brilliant search engine expert, whose philosophy is to always back up all his words with hard facts derived from deep knowledge of applied mathematics. As Director of Technology at Radiocom Ltd. Mike uses methods of sampling analysis to get into the nature of things, providing users of WebSite CEO with an outstanding Web site promotion suite (<http://www.websiteceo.com>); thus providing them with the best possible results in the search engines. Mike controls all R&D and search engine studies within WebSite CEO. His latest "hobby" is the PageRank technology because it is similar to graph theory, his favourite university course. His strongest dislike is search engine spam, which he considers the deadly sin of Internet age. Mike mercilessly fights it by enlightening minds of WebSite CEO users and making regular raids into spammers' camps.

Edited by Jill Whalen, owner of High Rankings and moderator of the High Rankings Advisor free weekly email newsletter (<http://www.highrankings.com/advisor.htm>) . Jill has been practicing successful search engine optimisation since 1995 for a wide variety of clients. Rather than worry about PageRank, Jill prefers to help sites be the best they can be, which in turn garners them lots of inbound links (and thus high PageRank!).

Technical Editing by Yuri Baranov: Yuri is Director of Marketing at Radiocom Ltd. He concentrates on general Web site promotion and search engine optimization. Yuri has a groundbreaking insight into search engine marketing, and does all in his power to make it more consistent and comprehensible to ordinary people. He currently coordinates all search engine studies, sets goals, and evaluates results. He is also interested in human engineering and usability issues. Yuri agreed to become the technical editor of this paper because he is head over hills in love with Google, which he considers "the best search engine ever."

Reproduction of this document in whole or in part is permitted solely with the prior consent of the authors. “Fair Use” reproduction is allowed, however, you should think carefully if you need to reproduce more than a few paragraphs. Educational Institutions are allowed to copy this document freely.

VERSION 3.0 Last amended – September 2002

Introduction	2
What is PageRank?.....	2
How is PageRank determined?.....	3
How can you tell what a page's PageRank is?.....	3
How accurate is the Google toolbar?.....	3
How significant is PageRank?.....	5
Is PageRank a good determination of the quality of a page?	6
Google's First 1000 Results	7
The difference between PageRank and other factors	8
Non-PageRank Factor Threshold	8
Using the threshold to derive the worth of two ranking strategies	9
Really heavy competition	11
How PageRank is calculated	12
A word on convergence.....	16
Derived specifics of how Google calculates PageRank	18
PageRank Feedback, how linking sometimes helps	19
Influencing the results	22
Controlling PageRank.....	22
Links to Your Site.....	23
Links out of your site.....	24
Internal Structures and Linkages	28
What Google says.....	34
Final word to non mathematicians.....	35
Advanced PageRank topics	36
Speeding Up PageRank	36
How PageRank could be more than it seems	42
Favouring pages in directories.....	44
Examples	45
Mathematical aspects of PageRank for advanced readers.....	45
Further Reading and Resources.....	54
Technical Proof Readers credits	555

Introduction

The original PageRank Explained document has existed for a long time. It was the first of its kind to present many of the ideas in a way that the average person can understand. You may be wondering why there is a need for a new version of this document. The reasoning is simple; everything in the original document was theoretical and subject to change. These changes have occurred through the authors' further learning on how Google reacts and how we think about the subject. To this end you will note that an additional author and technical editor has been added to the paper. Mike Shishigin's extensive mathematical ability enables us to explore some new areas of PageRank not mentioned before, as well as provide a section at the end that contains more advanced information. We hope to present PageRank in a way that gets across everything you could possibly need to know -- without baffling you. However, if you want to know the extremely technical stuff, then we'll give you that at the end.

We're pleased that the principles in the original PageRank Explained document seem to have been accepted universally within the search engine optimization field. The one document that appears to criticise it, later explains why each of the points it makes are valid. We're also pleased that since its publication, a vast amount of additional articles have been written about PageRank.

Despite all this, things **have changed**, and there are still a lot of misunderstandings regarding the importance (or unimportance) of PageRank. We've added new sections to address these, and new sections to address some new things. We've also tried to make the existing sections more understandable. This is a long document; it's written so that a broad range of people can understand it. As it's also used as a major reference in several university courses, we will do our best detail as much as possible. If you are "Joe Average," then you can simply limit your concern to understanding the principles until you're ready to dig deeper.

Finally, we thought it was time to dedicate some space to Google's thoughts about PageRank. Obviously Google is not going to comment on exact ranking issues but, as it is *their* technology, what they say about it gives us lots of insight. We asked Barry Schnitt, spokesperson for Google, a few questions about various PageRank related topics and he was kind enough to give us some answers; you'll find those answers later in the document.

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 deemed more "important" will move up in the results page of a user's search accordingly.

A basic overview of how Google ranks pages in their search engine results pages (SERPS) follows:

- 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.

In reality, it's slightly more complex and we'll discuss this in more depth later, but for now the above description serves our purposes. It's worth noting that PageRank is a multiplier and is not just simply added to the score. Thus, if your page had a PageRank of zero, it would rank at the very end of the SERPS.

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. PageRank also factors in the importance of the links pointing to a page. If a page has important links pointing to it, then its links to other pages also become important. The actual text of the link is irrelevant when discussing PageRank.

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

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

How accurate is the Google toolbar?

The Google toolbar is not very accurate in showing you the actual PageRank of a site, but it's the only thing right now that can 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 a guesstimate is to type the URL into the Google search box and see if the page shows up in the SERPS. If it doesn't, then the toolbar is guessing!

2. **The toolbar is just a *representation* of actual PageRank.** Whilst PageRank is linear, Google has chosen to use a non-linear graph to portray 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. A comparison table best illustrates this phenomenon. 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

The PageRank shown in the Google directory (<http://directory.google.com>) suffers from the same problems. The PageRank shown in the directory is also on a different scale. There have been attempts to cross-reference these two scales but because they are non-linear, the results really do not tell you anything more than you already know.

Also of note is that a programmer managed to generate a tool to look up PageRank without using Internet Explorer. This tool has since been withdrawn, but whilst originally the numbers given by this software and Google's toolbar matched - presently querying with such software sometimes produces different numbers than querying with the toolbar. This is Google's right to protect their data, but is the strongest indication that:

Sometimes what you see on the Google toolbar *may* not be related to actual PageRank at all. (Google can and does, assign whatever toolbar PageRank value they want to assign to a page.)

How significant is PageRank?

The significance of any one factor in search engine algorithms depends on the quality of the information it supplies. A factor's importance is known as its weight. To demonstrate how weighting is arrived at, it's easiest to move away from PageRank for a second and look at Meta tags. Originally, when the Meta keyword tag was new, you could write something like this in your document:

```
<meta name="keywords" content="pagerank, pagerank uncovered, algorithm, algorithms">
```

In theory, the Meta keyword tag was a very good indicator of what the page was about. However, as most are well aware – the weighting for the keywords tag is fast approaching nothing. Two things have contributed to this:

1. The *ease* at which Webmasters can manipulate it.
2. The *level* of manipulation by Webmasters.

These two things are separate factors, but with human nature being what it is, the easier something is to influence – the more it is manipulated. The combination of these factors determines the “weighting” – i.e., how much we trust the information provided by that factor.

So it makes sense to look at these factors in relation to PageRank first.

PageRank is, without doubt, one of the hardest things for a Webmaster to manipulate ethically. However, it is possible to generate links to your site from other sites fairly simply through the use of link farms and guestbooks. Google frowns upon this kind of abuse, and many sites that have tried this have had their PageRank influence blocked. But it must be said that the abuse is still rampant, and that it can have an influence on PageRank. So, whilst not easy to do, PageRank is still subject to manipulation.

The extent to which PageRank is manipulated has also changed. Most people no longer believe Google's old line of people not being able to influence PageRank and the results based on it. However, there is more information about PageRank available than ever, and people are more aware of manipulation techniques.

So whilst PageRank is valuable, you should be careful not to over-estimate its usage and capabilities. Your final ranking in Google is due to a mix of factors, of

which PageRank is only one. We'll get into more details later by discussing how PageRank is different than the other ranking factors, and thus, when it applies and when it doesn't. Ironically enough, PageRank's weighting factor is undeniably declining. Since the original version of this paper gave out detailed information about PageRank, in all likelihood it may also have contributed in some small way to the decline in weighting of the very subject it talks about!

Is PageRank a good determination of the quality of a page?

To examine the worth of PageRank, we need to first look at its premise, and how accurate it is. Basically PageRank says:

1. *If a page links to another page, it is casting a vote, which indicates that the other page is good.*
2. *If lots of pages link to a page, then it has more votes and its worth should be higher.*

The basic implication here is: *People only link to pages they think are good.*

It shouldn't be hard to convince you that this premise is wrong. A few of the reasons people link to pages *other than ones they think are good* are:

- 1) Reciprocal links – “Link to me and I'll link to you.”
- 2) Link Requirements – “Using our script requires you to put a link to our page.” or “We'll give you an award solely because you link to our page.”
- 3) Friends and Family – “This is my friend Pete's site.” or “My mum's site is here, my dad's site is here. My dog's site is here.”
- 4) Free Page Add-ons – “This counter was provided by www.linktountersite.com.”

Furthermore, anybody who has a top-ranking site will tell you that it tends to get links from new sites. This is not necessarily because it's good (although they generally are). Assume a Webmaster is setting up a new site and they are looking for some outbound links. Nowadays, one of the first things they do is a Google search for similar sites. The links they end up with may not necessarily be the best sites, but merely the easiest ones to find. If PageRank influences rankings, and if they subsequently link to those pages – the new Webmaster will be adding to the inaccuracies in the judging of the quality of a page. The same is true when these new Webmasters use the Google Toolbar PageRank indicator to choose whom to link to.

To put this another way:

PageRank is determined by the links pointing to a page. But if PageRank itself has an influence on the number of links to a page, it is influencing (in a circular way) the quality of that page. The links are no longer based solely on human judgement. If a Webmaster picks their outbound links by searching on Google or by looking at the Google toolbar (even if this is only part of their judgement), then there is a corresponding increase in a page's PageRank. This increase is not solely because it is a good page, but because its PageRank is already high.

Google's First 1000 Results

Remember, PageRank alone cannot get you high rankings. We've mentioned before that PageRank is a multiplier; so if your score for all other factors is 0 and your PageRank is twenty billion, then you still score 0 (last in the results). This is not to say PageRank is worthless, but there is some confusion over when PageRank is useful and when it is not. This leads to many misinterpretations of its worth. The only way to clear up these misinterpretations is to point out when PageRank is *not* worthwhile.

If you perform any broad search on Google, it will appear as if you've found several thousand results. However, you can only view the first 1000 of them. Understanding why this is so, explains why you should always concentrate on "on the page" factors and anchor text first, and PageRank last.

Assume that you perform a search on Google and it returns 200,000 results. If we were to calculate every factor for each 200,000 pages – do you think it would really take just 0.34 seconds to search? The answer to speeding up the search is to get a subset of documents that are most likely to be **related** to the query. This subset of documents needs to be larger than the number of search results. For example, let's say that number is 2000. What the search engine does is query the whole database using 2 or 3 factors, finding the 2000 documents that rank highest for them. (Remember, there were 200,000 possible documents, and that's the number that actually gets shown). Then the engine applies all the factors to those 2000 and ranks them accordingly. Because there's a drop in the quality of the results (not the pages) at the bottom of this subset, the engine just shows the first 1000. PageRank is almost certainly not one of those factors. Notice how before, we highlighted the word "related," in creating the subset of 2000 pages. The search engine is looking for pages that are on-topic. If we included PageRank in that list we'd get a lot of high PageRank pages with topics that are only slightly related (because of the second factor), but that's not what we want.

Why this is **critical**:

You **must** do enough "on the page" work and/or anchor text work to get into that subset of 2000 pages for your chosen key phrase, otherwise your high PageRank will be completely in vain. PageRank means nothing if you do not have enough ranking from other factors to make it into the first subset.

The difference between PageRank and other factors

To assess when PageRank is important and when it is not, we need to understand how PageRank is different from all other ranking factors. To do this, here's a quick table that lists a few other factors, and how they add to the ranking score:

Title Tag	Can only be listed once.
Keywords in Body Text	Each successive repetition is less important. Proximity is important.
Anchor Text	Highly weighted, but like keywords in body text, there is a cut off point where further anchor text is no longer worthwhile.
PageRank	Potentially infinite. You are always capable of increasing your PageRank significantly, but it takes work.

All other ranking factors have cut off points beyond which they will no longer add to your ranking score, or will not add significantly enough for it to be worthwhile. PageRank has no cut off point.

Non-PageRank Factor Threshold

Having written about the difference between PageRank and other factors, and how PageRank is harder to get, it should now be clear that whilst we could use many methods to get good rankings, there is a threshold which defines when high PageRank is worth striving for and when it is not.

With ranking factors other than PageRank, there is a score beyond which the slow down in the rate that any factor adds to this score is so insignificant that it is

not worthwhile. This is the Non-PageRank Factor Threshold. To illustrate this, let's put an example figure on this of 1000.

If we have a query where the results are Page A and Page B, then Page A and B have scores for that query which are the total scores for all ranking factors (including PageRank). Let's say Page A's score is 900 and page B's score is 500. Obviously Page A will be listed first. These are both below our hypothetical Non-PageRank Factor Threshold, thus without any change in PageRank, it is possible for page B to improve their optimization to beat Page A for this particular query. There are lots of queries like this on Google; they're more commonly thought of as less competitive queries!

Now assume Page A raises its score to 1100. Suddenly page B cannot compete in the SERPs (search engine results pages) without increasing its PageRank. In all probability, page B must also improve for all the other ranking factors, but an increase in PageRank is almost certainly necessary. There are also lots of queries like this on Google, which are more commonly thought of as more competitive queries!

Generally, when querying Google, the group of pages in the SERPs will contain some pages that have a score above the Non-PageRank Factor Threshold, and some that do not.

There is an important point to be made here:

To be competitive you **must** raise your page's search engine ranking score beyond the Non-PageRank Factor Threshold. To fail to do so means that you can **easily** be beaten in the search results for your query terms. The quickest way to approach the Non-PageRank Factor Threshold is through "on the page factors," however you cannot move above the Non-PageRank Factor Threshold without PageRank.

The obvious question is what's the numerical value of the Non-PageRank Factor Threshold, and how much work do you need to do to get past it. The answer is that it has no value; it is a hypothetical line. Google could put a value on it, but that would not help us unless we know what the page's individual scores are. We need only be aware that the threshold exists, and that it gives us information about principles.

Using the threshold to derive the worth of two ranking strategies

The threshold explains principles and the different ways that search engine marketers work. It also demonstrates why some of the misunderstandings about PageRank occur. Let's consider the strategies of two people, Person A considers PageRank to be unimportant, and Person B considers PageRank to be very important.

Person A says “PageRank” is unimportant. They have optimised pages for years and know how to use “on the page” factors very successfully. They understand the basics of anchor text but they couldn’t care at all about PageRank.

What’s happening: person A is reaching the Non-PageRank Factor Threshold very quickly because they are maximising the “on the page” factors. Through carefully choosing keywords they jump-start themselves up the SERPs. As long as their content is good, high-ranking sites (over time) tend to get linked to. Whilst they didn’t directly ask for it, a slow trickle of sites will begin to link to them and give them PageRank, which helps consolidate their position.

Person B says “PageRank” is important. We’ve all seen those pages in the results that have no content, but great rankings (With big brands, this can often occur naturally even when they have no idea what PageRank is. This would be Person C who is not relevant to the discussion at hand.) Person B understands lots about PageRank and concentrates heavily on it.

What’s happening: person B is doing the reverse of person A. Whilst person A concentrated on the Non-PageRank factors and found herself getting PageRank anyway, person B concentrates on the PageRank Factor and finds himself getting Non-PageRank factors. The reason for this is that increasing PageRank requires links, and links have anchor text. Thus, through carefully choosing the anchor text linking to his page, person B automatically increases his Non-PageRank factor scores whilst obtaining his high PageRank score.

Obviously, these are two extremes, but we can use these to extrapolate the advantages and disadvantages of each approach:

	Advantages	Disadvantages
Person A	<ul style="list-style-type: none"> • Quick entry to the results pages. • Self-generating links limit the amount of work needed. 	<ul style="list-style-type: none"> • Securing the results to make it more difficult for competitors to compete takes more work. • Slower to react to new competitors.
Person B	<ul style="list-style-type: none"> • Solid position. Can easily modify on the page factors for a quick boost if required. • Probably higher traffic from non-search engine sources. 	<ul style="list-style-type: none"> • Slower entry to results pages. • Difficult to do well. • Increased likelihood of tripping spam filters.

It is clear that both strategies can and do work. Both strategies are using PageRank as part the mix of factors that will ultimately improve their ranking in the SERPS. Because there is such a mix, we can use them to different degrees, depending on the strategy that best suits your style. My personal strategy is to use a combination, but to save some of the “on the page” factors for later, in case I need a quick boost if the competition heats up at a later time.

Really heavy competition

No explanation of PageRank strategies would be complete without a final statement regarding heavy keyword competition. There are some queries where competition is so intense that you must do everything possible to maximize your ranking score (for example “web hosting”). In such situations it is impossible to rank highly through Non-PageRank factors alone (as you will not initially be listed high enough to be noticed and linked to). That is not to say that Non-PageRank factors aren't important. Consider what your final rank score is:

Final Rank Score = (score for all Non-PageRank factors) x (actual PageRank score).

Improving either side of the equation can have a positive effect. However, because the Non-PageRank factors have a restricted maximum benefit, the actual PageRank score must be improved in order to compete successfully. Under really heavy competition – it holds true that you cannot rank well unless your actual PageRank score is above a certain level. In other words:

There exists a query specific “Minimum PageRank level.” For queries that do not have heavy competition, this level is easy to achieve without even trying. However, where heavy competition exists, Non-PageRank factors are just as important (and easier to get) until they reach the Non-PageRank factor threshold. This is why careful keyword choice can help you avoid the extensive work associated with highly competitive search phrases.

How PageRank is calculated

On a simple level, we can tell quite a lot about how PageRank is calculated. This is because when Google was just a university research project, the creators of PageRank published a paper that detailed a formula for calculating it. This formula is now more than a handful of years old, and we suspect that it has changed somewhat since then. However, for detailing the over-riding principles of how PageRank works, it is as accurate today as the day it was written.

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

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

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

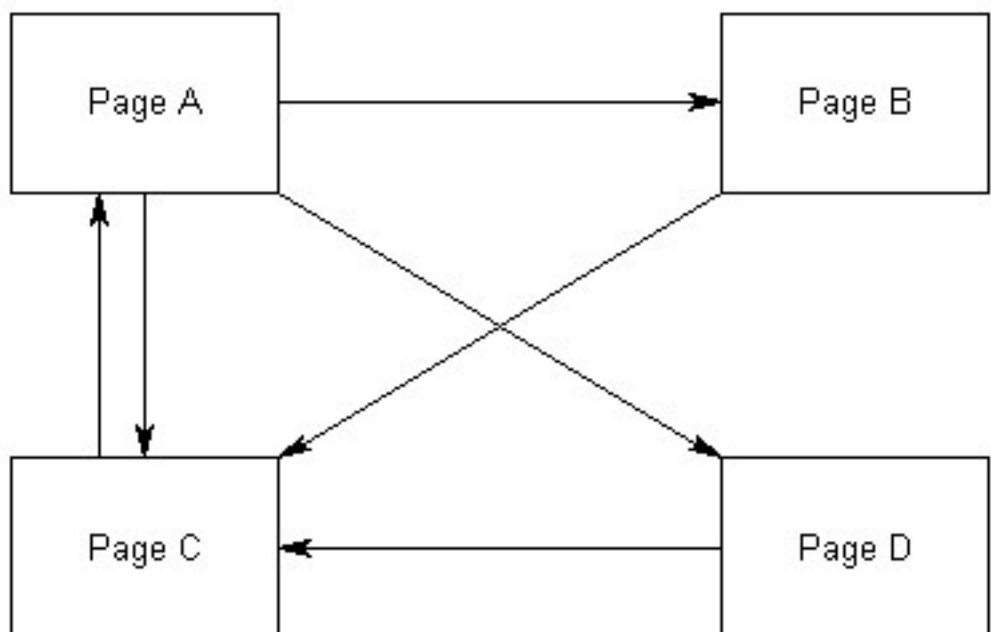
Couldn't be simpler right? Depending on your mathematical competence, this is either remarkably easy to understand, or remarkably complex. So here's the deal – this isn't a one-time calculation. It looks pretty above, but you can't just do one simple calculation and get an answer with the PageRank of a page! If you look at it, to calculate the PageRank of Page A, we need to know the PageRank of all the pages pointing to it. To calculate the PageRank of those pages we need to know the PageRank of all the pages pointing to them (one of which could, and is quite likely to be Page A!). If this seems like it could go around in circles forever,

that's because it nearly does. We have to do a whole lot of little calculations many times over before we actually get the answer. What this formula tells us is that however you slice it, and however the formula may have changed since it was written:

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. This means that a page's PageRank is essentially a measure of its vote; it can split that vote between one link or two or many more, but its overall voting power will always remain the same.

This really is important. So to be totally clear, let's put some numbers to it (these numbers are made up for demonstration purposes only, and have no relevancy to any particular page). Say Page B has a PageRank value of 5 and has a single link on it pointing to Page A. Page A's PageRank is improved by a proportion of Page B's value of 5 (Page B doesn't lose anything, but Page A gains). If Page B has two links, that PageRank improvement would be split, and Page A would only gain half the PageRank that it did before.

Now put the formula out of your mind for a moment, as it's easier to understand how it works using a diagram. Let's say we have a hypothetical set of pages imaginatively titled Page A, Page B, Page C and Page D. They link to each other as shown below:



To begin with, in our example at least, we don't know what the page's starting PageRanks are. There's nothing special about the number that we pick to start with (in fact, if you read forward to the section on convergence – you can see that

we can start at any number we want). Since in the last version of this paper we performed the calculations by setting these values to 1 – we’re going to set them to zero this time around in order to prove that it doesn’t matter what the starting values are.

Next, we perform the necessary calculation to obtain the PageRank for each page. The rules are:

1. We take a $0.85 * \text{a page's PageRank}$, and divide it by the number of links on the page.
2. We add that amount on to a new total for each page it's being passed to.
3. We add 0.15 to each of those totals.

The first calculation is easy. Because we’ve started at zero – $0 * 0.85$ is always 0. So each page gets just $0.15 + 0$. Meaning each page now has a PageRank of 0.15. Clearly we’re not done – we want to show the importance of each page based upon links, and they’re all the same; so we need to run the calculation again.

Page A links to pages B, C and D. Page A’s PageRank is 0.15 so it will add $0.85 * 0.15 = 0.1275$ to the new PageRank scores of the pages it links to. There are three of them so they each get 0.0425.

Page B links to page C. Page B’s PageRank is 0.15 so it will add $0.85 * 0.15 = 0.1275$ to the new PageRank score of the pages it links to. Since it only links to page C, page C will get it all.

Page C links to Page A, all 0.1275 passes to page A.

Page D links to Page C. Again all 0.1275 passes to page C.

The new totals for each page them become:

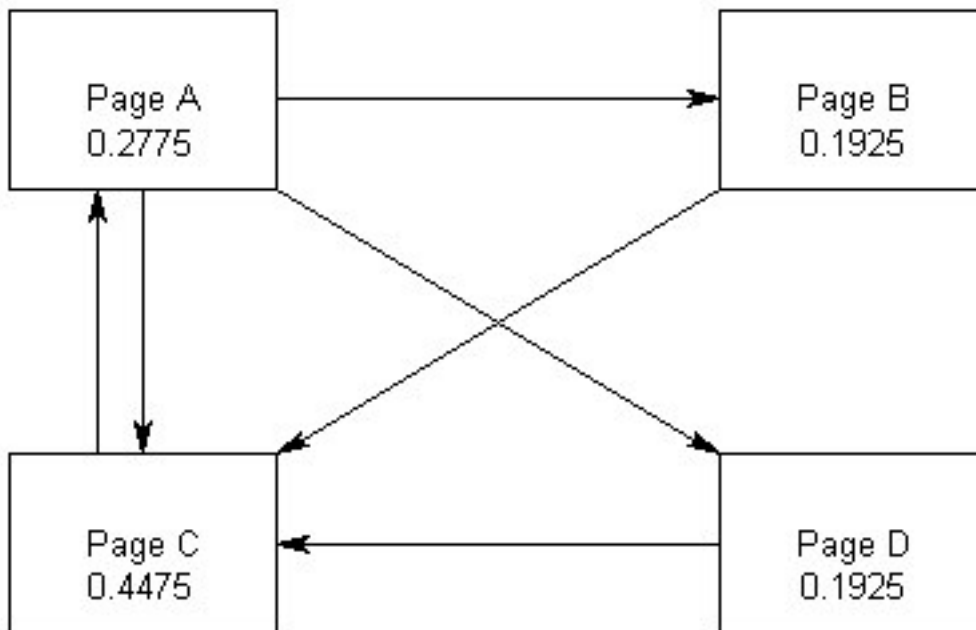
Page A: $0.15 \text{ (base)} + 0.1275 \text{ (from Page C)} = 0.2775$

Page B: $0.15 \text{ (base)} + 0.0425 \text{ (from Page A)} = 0.1925$

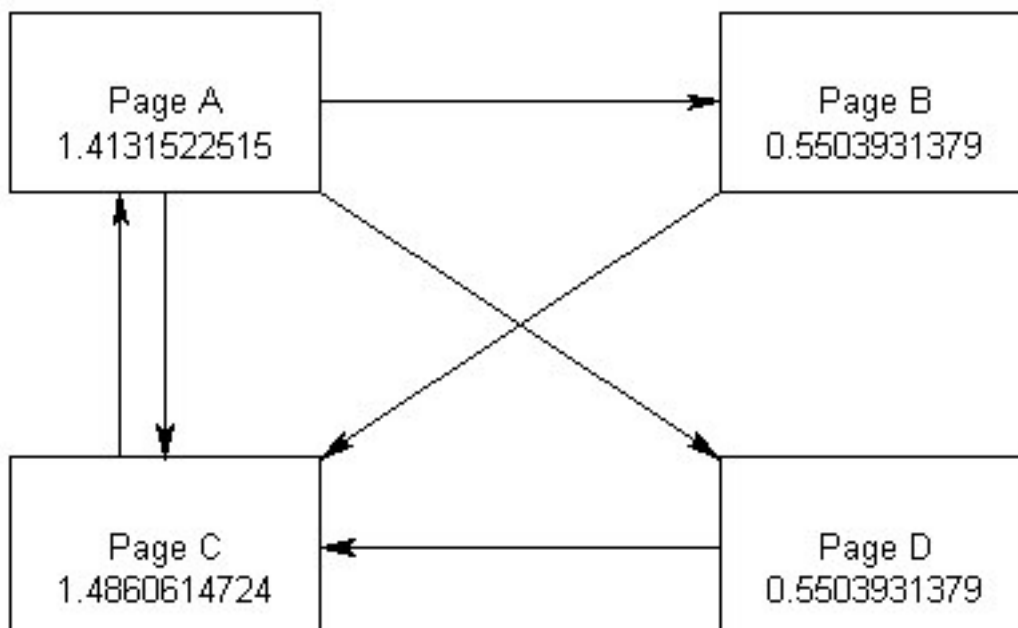
Page C: $0.15 \text{ (base)} + 0.0425 \text{ (from Page A)} + 0.1275 \text{ (from Page B)} + 0.1275 \text{ (from Page D)} = 0.4475$

Page D: $0.15 \text{ (base)} + 0.0425 \text{ (from Page A)} = 0.1925$

So we’ve got:



Pretty neat huh? Already we're beginning to see that Page C is probably the most important page in the system (but we can't be sure yet – it could well change). We carry on doing these calculations until the value for each page no longer changes (this is called convergence – there's more about this in the next section). In practice, Google probably doesn't wait for this convergence, but instead run a number of iterations of the calculation which is likely to give them fairly accurate values (more on this later as well). If we carried out all the calculations for the example given, it would take us 143 calculations [**Excel Example 1**] and we'd reach final values of:



As suspected, Page C is the most important. If we take a quick look at these raw values we can see something about the number of links pointing out from a page. Look at Page A, which has a link from a high PageRank page (Page C), which has only one outbound link. Then look at Page B and D; both share links from a high PageRank page (Page A), with three outbound links. The number of links significantly alters the way PageRank is distributed.

A word on convergence

Convergence is an important mathematical aspect of PageRank, which allows Google to provide unprecedented search quality at comparably low costs. This is a semi-complex topic but it is important to your understanding of how and why PageRank works. We've tried to make it as simple as possible but, unless you're Sergey Brin or Larry Page, you'll still need to concentrate!

Whilst it will take some concentration, it isn't that hard to understand.

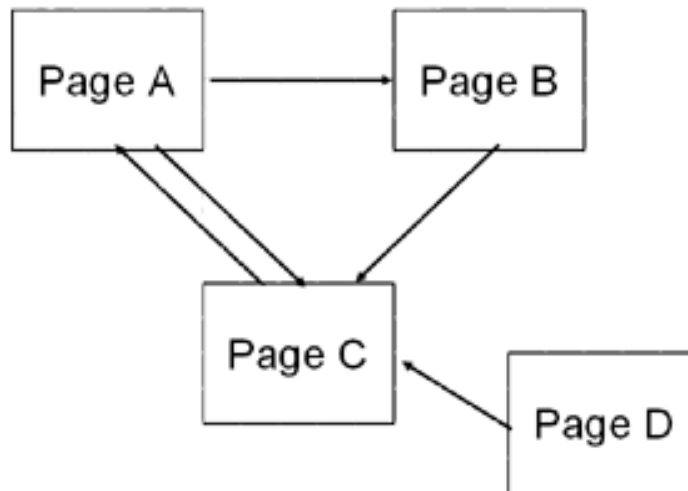
We've shown that the outlet values (*final values*) of one stage of the calculation become the inlet values (*starting values*) of the next stage, and that we keep on doing this (it's known as a recursive procedure). But the really big question is how and when *does the recursive procedure stop?*

The answer is "convergence". Provided the dampening factor (d in our equation) is less than one, then convergence will occur. Nominally we set it to 0.85 (because that's the value mentioned in the Stanford papers).

This convergence basically means that whatever values we start at, after running the calculation a number of times we will end up with the same final values and that these values will no longer change if we do further iterations of the calculation. These final values are known as *limiting values*.

Once the limiting values have been reached, Google no longer needs to expend processing power on calculating the PageRank. They can finish there!

This is easier to understand with an example. Let's take a look at the following structure:



PageRank for pages A, B, C, D at various stages of iteration

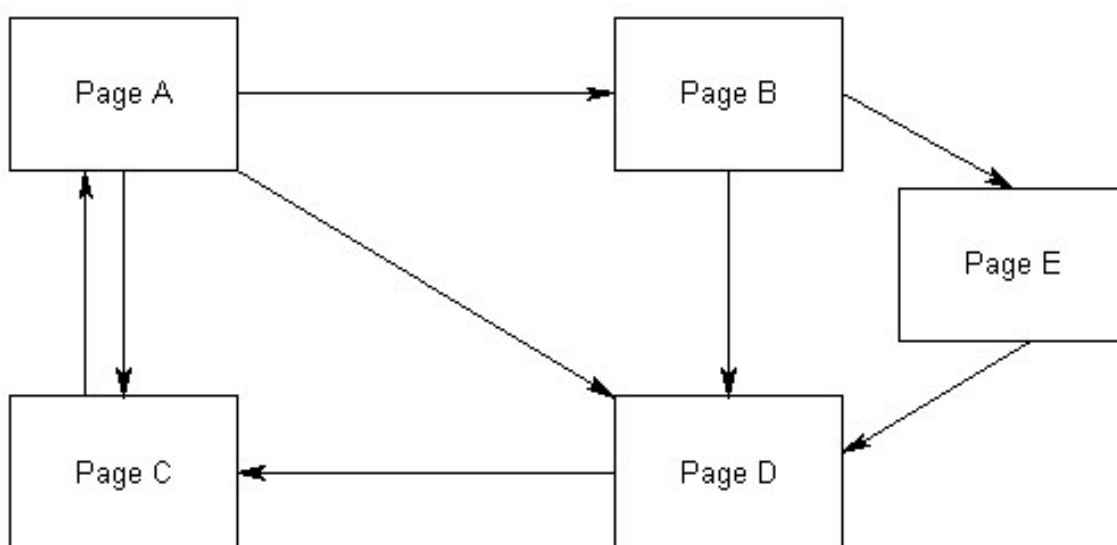
	A	B	C	D
Iteration	1	1	1	1
1	1,000000000	0,575000000	2,275000000	0,150000000
2	2,083750000	0,575000000	1,191250000	0,150000000
3	1,162562500	1,035593750	1,651843750	0,150000000
*****	*****	*****	*****	*****
19	1,4900124031	0,7833688246	1,5766187723	0,150000000
20	1,4901259564	0,7832552713	1,5766187723	0,150000000
21	1,4901259564	0,7833035315	1,5765705121	0,150000000
46	1,4901074052	0,7832956473	1,5765969475	0,150000000
47	1,4901074054	0,7832956472	1,5765969474	0,150000000
48	1,4901074053	0,7832956473	1,5765969474	0,150000000
49	1,4901074053	0,7832956473	1,5765969474	0,150000000
50	1,4901074053	0,7832956473	1,5765969474	0,150000000
51	1,4901074053	0,7832956473	1,5765969474	0,150000000
52	1,4901074053	0,7832956473	1,5765969474	0,150000000
53	1,4901074053	0,7832956473	1,5765969474	0,150000000
54	1,4901074053	0,7832956473	1,5765969474	0,150000000
*****	*****	*****	*****	*****

No matter how much we try to calculate the value past the 48th iteration, the values do not change – they have converged at the limiting values.

In practice there's no need to wait until they do not change at all before we stop calculations. We simply wait until they do not change very much. In our example above, we have done that to 10 decimal places. I.e. we have calculated until they do not change by 0.0000000001 or more (or mathematically $|PR_{48}(A) - PR_{50}(A)| < 10^{-10}$)

Derived specifics of how Google calculates PageRank

We started our values from zero, and now see that the calculations are fairly complicated and resource intensive. So two questions arise: When Google calculates PageRank themselves, do they start each month with an arbitrary value? And secondly do they calculate PageRank across several machines? We can answer these questions by a couple of simple experiments. Assume that having calculated the PageRanks in the structure shown earlier, we would like to add a page and change a link to make the structure look like this:



If we start each page from an arbitrary starting value of one, then how many iterations of the calculation does it take to achieve convergence? [**Excel Example 2**]

75

Now, if we start each page from the values we ended up with when we calculated the PageRank of the original structure (and set Page E's starting value to 1) then how many iterations does it take to achieve convergence? [**Excel Example 3**]

78

These numbers are really very close. Therefore, it would appear logical that Google does not reset the values with each calculation as they may simply carry on from where they left off.

So do they calculate across several machines or one? Obviously there is too much work for one machine when you get to billions of pages, so several

machines must be used. Logically, how can they do this? Assume Pages A and C are on one machine, and Pages B, C and D are on another. If we ignore the links between pages across machines (i.e., A->D and D->C), then for Pages A and C it takes:

1 calculation to converge at A=1, C=1

Then for Pages B, D and E it takes:

3 calculations to converge at B=0.15, D=0.3954375000, E=0.3954375000

If we then do calculations for all pages using these as starting values it takes:

A huge 146 iterations to converge [**Excel Example 4**]

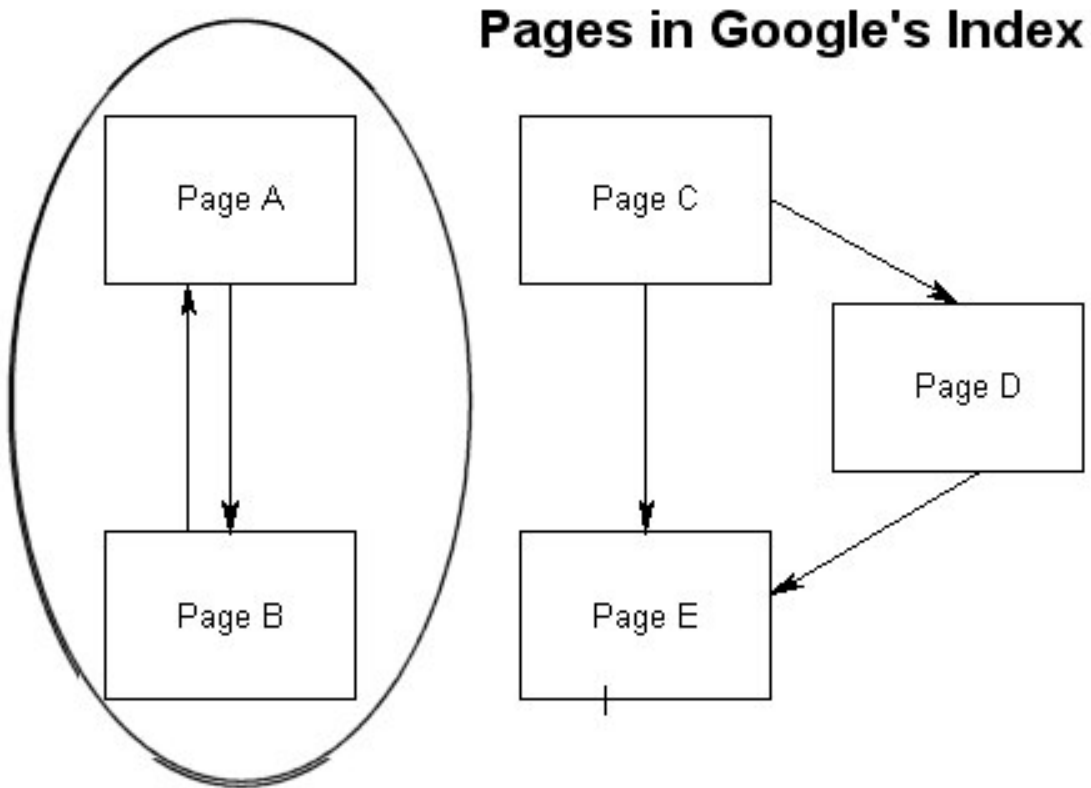
The conclusion is that splitting PageRank calculations up into chunks for different machines is not a viable proposition. Therefore, they must also track PageRank transfers across links between machines. This means that each machine must notify the other machines as PageRank is given to a page not on that machine. Indeed, each machine must also run the iterations in sync with each other. Even if they don't do it exactly this way, the calculation of PageRank for so many pages is very, very complicated indeed.

PageRank Feedback, how linking sometimes helps

When the new term PageRank Feedback was presented in the original version of this paper, it was largely taken up and used – although not always correctly. PageRank Feedback is difficult to understand because it is a concept that implies creation of PageRank. However, it is really an alteration of distribution of PageRank on a large scale. PageRank Feedback is the principle that explains how, sometimes, it is good to link to outside pages. Consider we have a Page -> Page A. Its PageRank, by definition, will be 0.15. Now assume it links out to Page B, which in return links back to it. Page A's PageRank now becomes 1. Now assume it links out to Page C and Page C links back to it. Page A's PageRank now becomes 1.4594594595.

The very fact that Page A links out to a page that links back to it causes its own PageRank to rise. (This could also occur if Page A links to a page that links to a page that links to Page A, etc....) This has occurred, not because we are generating PageRank, but because it is taking it from the system as a whole. But if we view this very small system "Page A" (which is a subsystem of all the pages in the index), then we can firmly say that linking out from Page A has generated PageRank in this system – that is, the very act of linking out has fed PageRank back in.

Let's put this diagrammatically using a slightly more complicated system.

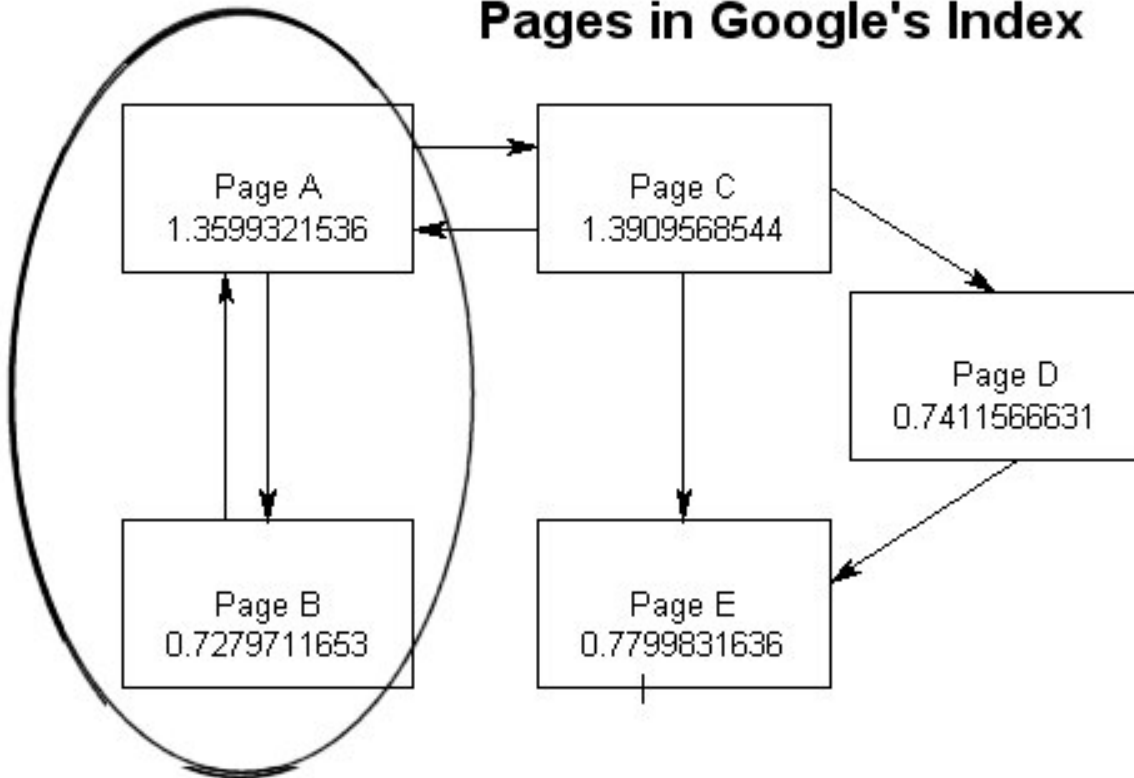


Pages A to E are all of the pages in Google's index. Pages A and B are your site. When we do the calculations for this set up, the final PageRanks for each page are all 1 [**Excel Example 5**]. As far as your site is concerned:

Page A's PageRank is 1
Page B's PageRank is 1
The sum of the PageRank of all the pages in your site is 2

Changing the structure to link A to C and C to A gives the following final PageRanks [**Excel Example 6**].

Pages in Google's Index



Page A's PageRank is 1.3599321536
Page B's PageRank is 0.7279711653
The sum of the PageRank of all the pages in your site is 2.0879033189

This increase is small; the amount will of course vary depending on circumstances. But the point was easily proved – linking out can *sometimes* improve your PageRank through a mechanism known as PageRank Feedback.

If we added an extra page into the loop between C -> D and E, then the total increase in PageRank in your site would be 2.1462030505 [**Excel Example 7**]. This suggests that it is wiser to reciprocal link with larger sites that employ a link structure where a higher amount of focus is given to the page on which your link appears (or will appear). The number of pages, and the structure you are linking to has significant influence on the amount of PageRank feedback that occurs.

Influencing the results

The question that arises from all this is how and when can or will Google influence the results of the PageRank calculation. In this respect, the situation has changed significantly since the original PageRank paper. Google has shown that they can, and will modify the data on which PageRank is based. The primary example of this is what has become known as PageRank Zero (PR0). Basically speculation says that when Google wants to penalize a page, it is assigned a PageRank of zero. As PageRank is a multiplier, this will obviously always list PR0 pages as the very last entry in the SERPS:

Your score from other factors multiplied by your PageRank score of zero, will always be zero (the lowest score possible).

How they do this is important. Let's say they set the PageRank of the page to zero before doing the calculations – would this work? From what we know so far, we would think not. **[Excel Example 8]** And true enough, it doesn't. Why is this significant? Because the PageRank must be set to zero at the end of the process. Having a PageRank of zero in itself does not disqualify a page from "voting" exactly as it would have done in the first place.

To stop its voting power, the second penalty must also be applied. This is the same penalty that is applied to link farms. Google has shown that they are capable of ignoring links they believe have been artificially created. This is simple enough; they merely remove the link entry from the matrix. This second penalty would also need to be applied to a PR0 penalty to stop the links from that page from counting in the PageRank process. However, there doesn't appear to be any way to test this. I suspect that both penalties would be applied.

Controlling PageRank

As Webmasters, contrary to the suggestions of some, we have a fair amount of control over PageRank. It is, however, the most difficult of ranking factors to control, and you might well be able to achieve the desired effects by using other techniques. However, a well-worked PageRank added to your optimization mix can make it difficult for competitors to catch up.

There are three fundamental areas to look at when trying to optimize the PageRank for your site:

1. The links you choose to have link to you, i.e., which ones you choose, and how much effort you put in to getting them.
2. Who you choose to link out to from your site, and from which page of your site you place their link. (Maximising PageRank Feedback and minimising PageRank leakage).

3. The internal navigational structure and linkage of your pages, in order to best distribute PageRank within your site.

Links to Your Site

When looking for links **to** your site, from a purely PageRank point of view, one might *think* you 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 to 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 PageRank 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. 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 the page on your site 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 and enjoy the artificially enhanced PageRank that they provide.

Links out of your site

When considering links out of your site there is one golden rule:

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

This does not mean that you will lose PageRank from your site by linking out, but that the total PageRank within it may be lower than it could have been had you not linked out. From this, we can derive another rule, which gives away as little PageRank as possible (meaning that we are giving our own pages as much PageRank as possible):

The best outbound link PageRank scenario occurs when the outbound link comes from a page that has both:

a) A low PageRank

and

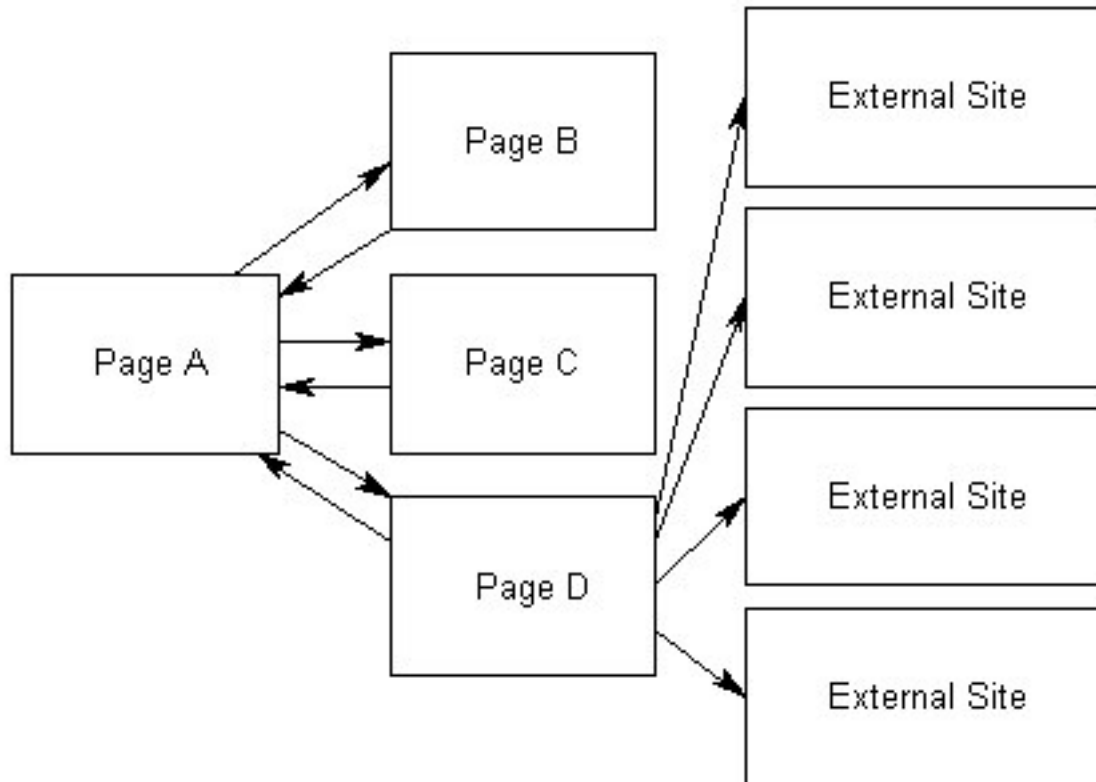
b) A lot of links to pages on your site.

How can this best be achieved? 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** open these pages using javascript because the spiders cannot yet follow this.

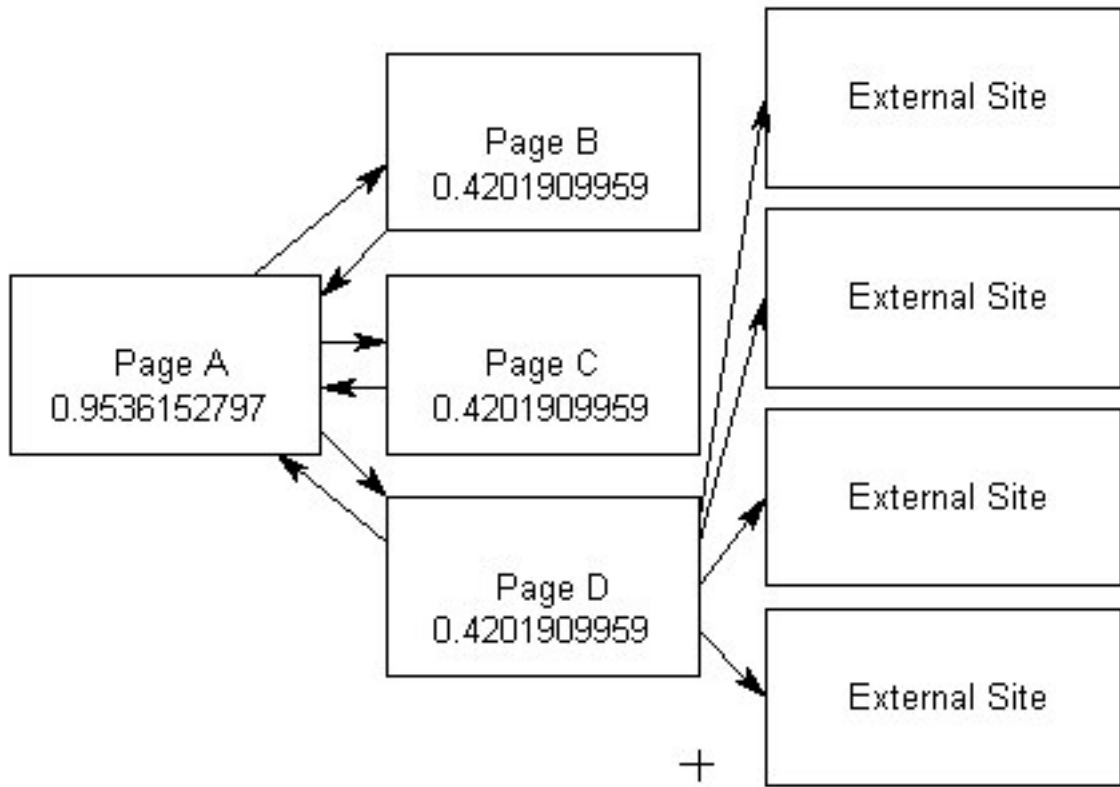
For example, we can do something like this with each link to an external site:
<a href=<http://www.supportforums.org>>The best search engine resource and forum site in the world Read my flattering review of them here.

Make sure that the review page also links back to a page in your own site that is high up its structure. (It's best if this is your home page, but any important page will do.) By doing this, you've significantly reduced the amount of PageRank you've let out of your site! We've targeted the distribution of PageRank 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).

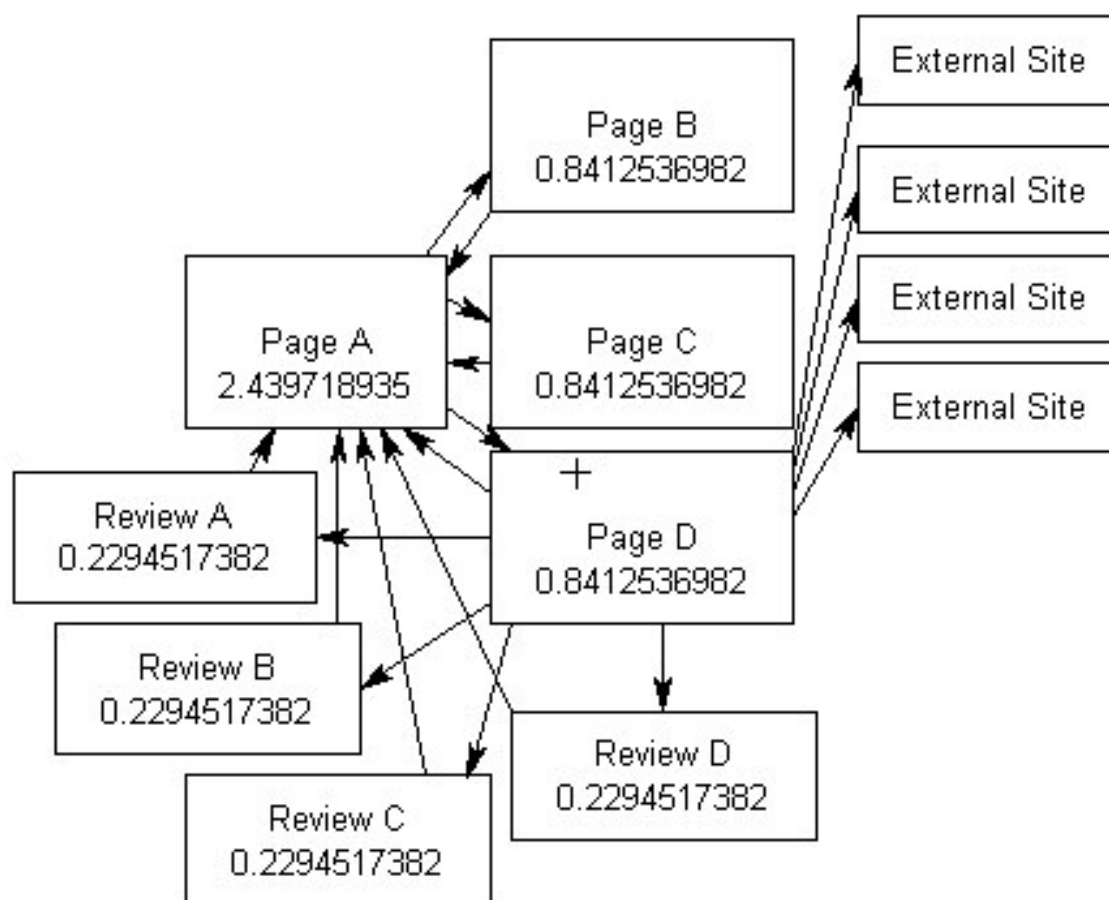
It's **very** good if someone links to your review page, so in addition you may let the site you've linked to know that you have reviewed them – it is quite possible you will get two links from their site (one to your site and one to the review of their site). This is all very complicated in text form, so here's a simplified example to show the principle and its effect.



Doing all the calculations for PageRanks we converge at the following values [**Excel Example 9**]:



If we perform the same calculations but include the review pages, here's what we get:



Just to be very clear here (and because in earlier versions of this paper some questioned the validity of this technique), if we look at only pages A B C and D then:

Without the Review Pages	With the Review Pages
The HomePage PR is: 0.9536152797	The HomePage PR is: 2.439718935
B,C,D PR is: 0.4201909959	B,C,D PR is: 0.8412536982
Total is: 2.2141882674	Total is: 4.9634800296

Some of this is the power of extra pages, some of it is feedback and some of it is drawing PageRank away from your outbound links pages. Fundamentally, however, adding extra internal links to your outbound pages is one of the most (if not the most) important internal factors to improve your PageRank. This boost may not be as much as from getting a new outside link to your site, but it's also much easier and more beneficial for your site's readers.

Internal Structures and Linkages

Having spoken about “Linking out” to external sites, the next logical progression is to discuss internal links within your site. If we recognise that PageRank is really about pages having a vote, then straight away we can derive one important thing about internal link structure and PageRank:

A page that is in the Google index has a vote, however small. Thus, the more pages you have in the index – the more overall vote you are likely to have. Or, simply put, bigger sites tend to hold a greater total amount of PageRank within their site (as they have more pages to work with).

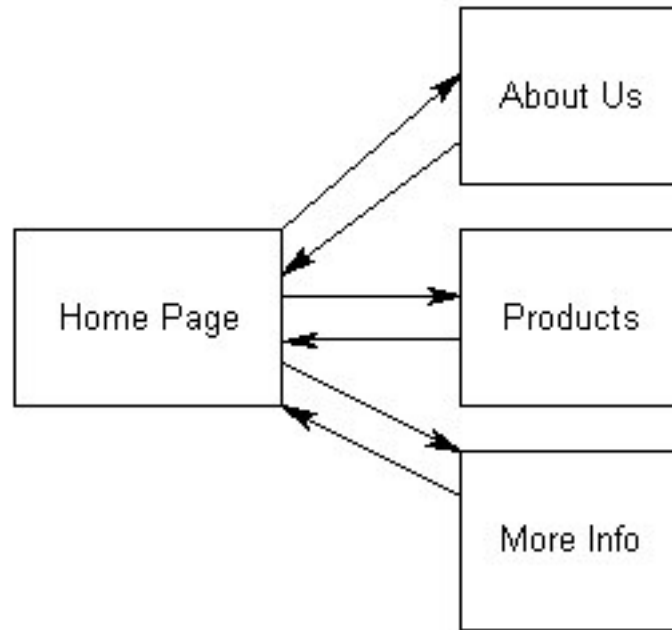
We need to clarify this a bit more. To get a high PageRank, it is not enough to have tens of thousands of pages. Those pages must also be in Google’s index. To achieve this they must contain enough content for Google’s algorithms to consider them worthy of being added to the index. As you develop content for your site, you are also creating more PageRank for your site. It’s hard work, and you’re creating it slowly – but if you’re also creating pages that people will want to link to, then you’re doing yourself two favours at once: PageRank from both directions. Or, to put it in very basic terms:

The best “internal” thing you can do to build PageRank is to write lots of good content. Ensure that pages aren’t overly short or overly long and break the content into several pages where necessary.

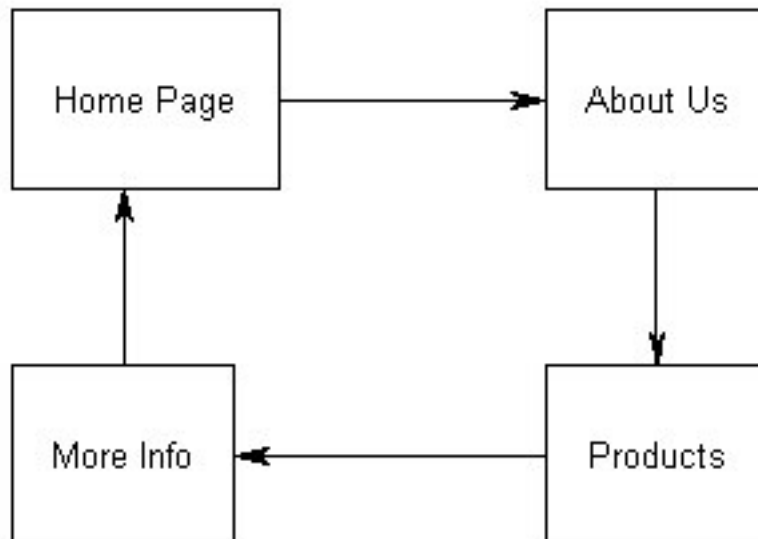
There are three different ways in which pages can be interlinked within a site. **In practice, sites might use a combination of these. Using a combination is fine and normal as long as you understand the different sections and how they are affecting your PageRank.**

For the purposes of this document we’ll view the different linkage structures as separate entities. We have:

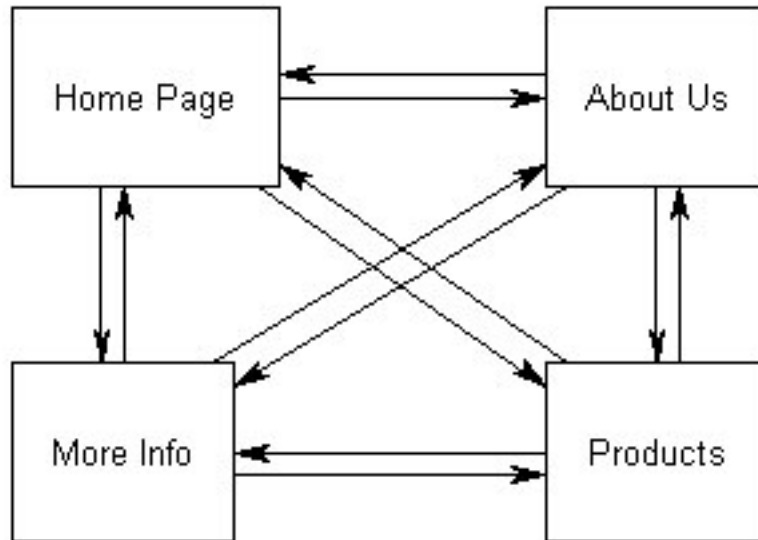
Hierarchical



Looping

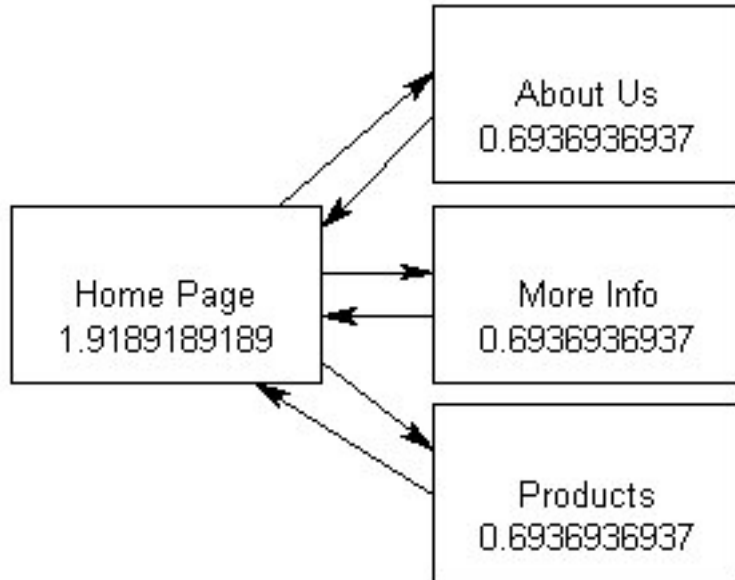


Extensive Interlinking



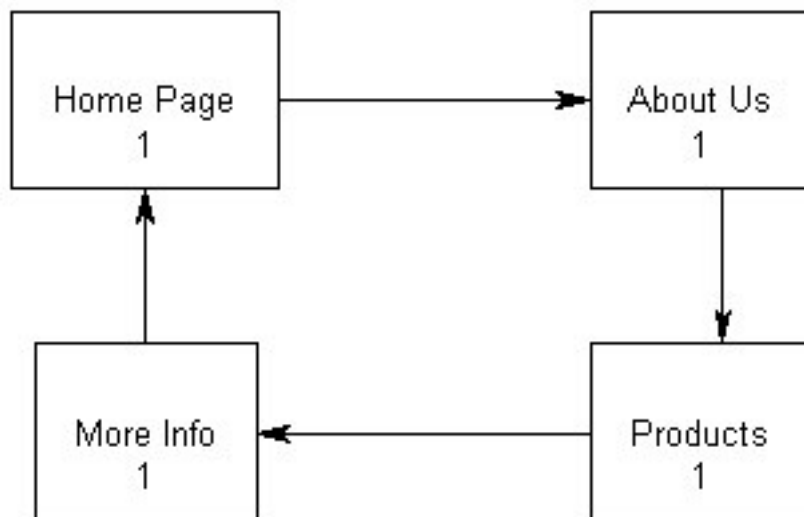
By performing the PageRank calculations for these structures, we can see how internal link structure can distribute PageRank around your site. This is critical because whilst at this stage of the document these are closed systems, when inbound and outbound links come into play you will want to be able to control your PageRank so that it is in the right places to maximise PageRank feedback. Since we have already extensively shown you the PageRank calculations, we'll just dive right in with the final value for each structure.

Hierarchical
[Excel Example 11]



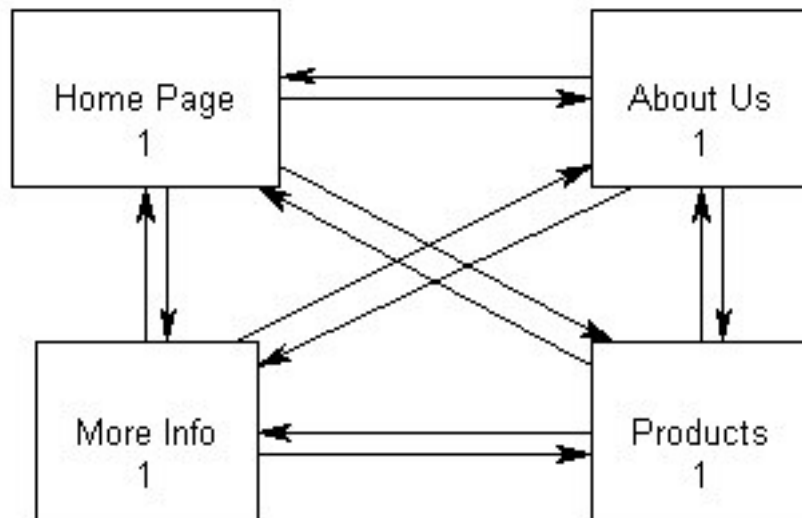
Total PageRank in this site is 4

Looping
[Excel Example 12]



Total PageRank in this site is 4

Extensive Interlinking *[Excel Example 13]*



Total PageRank in this site is 4

Notice something that is critically important. Because there are no inbound or outbound links, we are restricted to having the same total PageRank across all pages in our example site (4). However, introducing a hierarchical-like structure can alter the distribution of the PageRank. (Note, it does not have to be a strict hierarchy, but must contain more of the properties of the hierarchical structure than the looping or extensive interlinking.) This ability to shift PageRank to the pages that we want it to is the Webmaster's easiest way to manipulate PageRank.

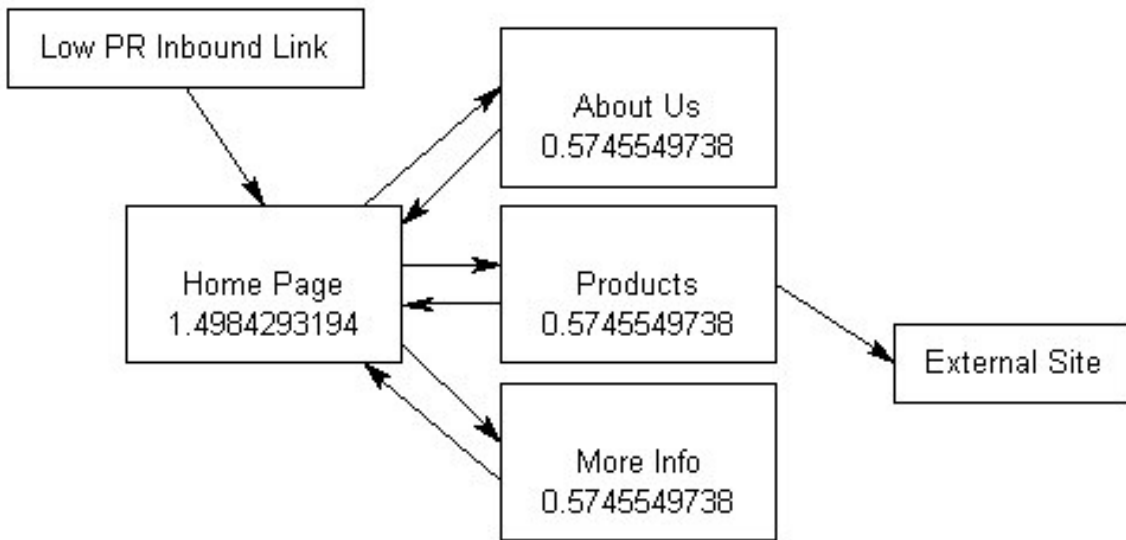
In general, we can state this as the following rule:

It is rare that a site will want to have an entirely linking or an entirely extensively interlinked structure. By using elements of a hierarchical structure, a Webmaster is able to shift PageRank to the pages of his or her choice. If PageRank was spread out evenly across all pages of the site, then the Webmaster is very close to obtaining maximum benefit.

The maximum benefit from PageRank is derived when this methodology is applied towards pages that want to rank high for highly competitive keyword phrases, or towards pages that must compete on a large number of keyword phrases. To do this the webmaster must move it away from pages that they do not need to rank as highly, or from pages that can rank well on the strength of keyword phrases alone.

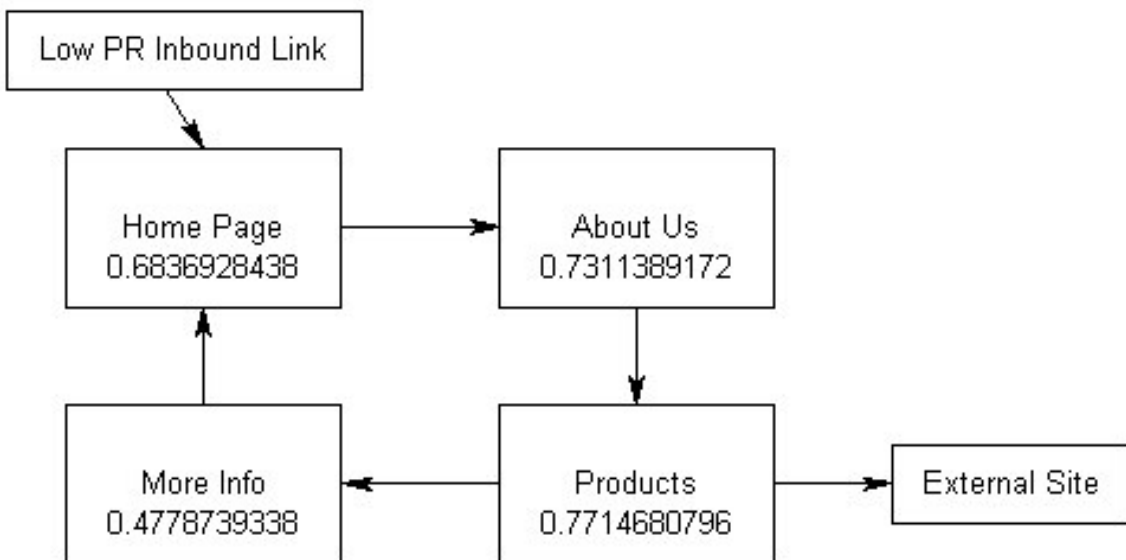
Closed systems (with no inbound and outbound links) are one thing, but let's take a look what happens when we add an inbound and outbound link:

Hierarchical
[Excel Example 14]



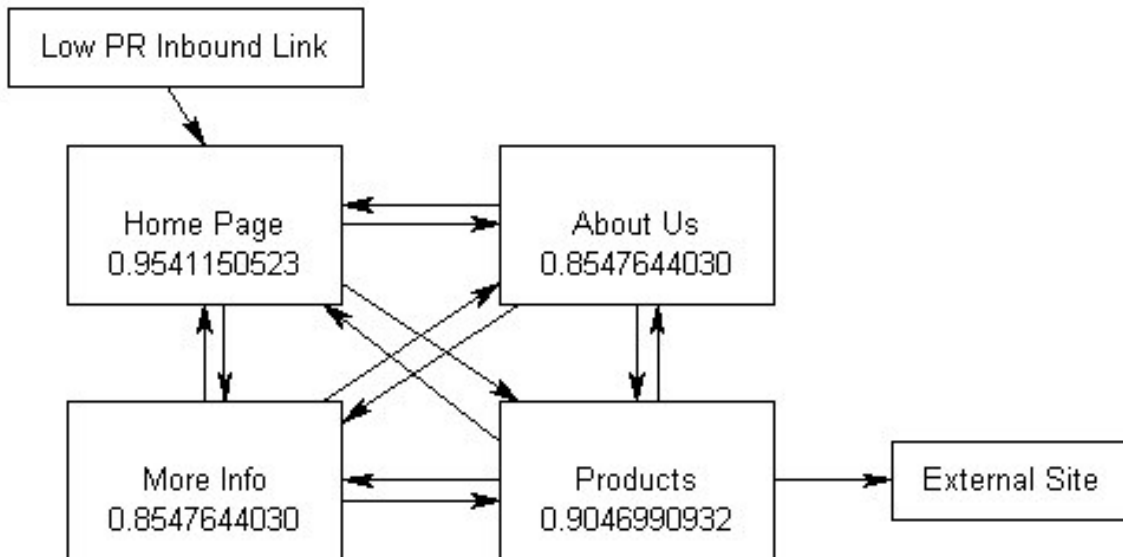
Total PageRank in this site: 3.2220942408

Looping
[Excel Example 15]



Total PageRank in this site: 2.6641737744

Extensive Interlinking *[Excel Example 16]*



Total PageRank in this site: 3.5683429515

When a system is non-closed, as most are, then we can say that the following holds true:

The Extensive Interlinking strategy retains the most PageRank within the site. Following this is the Hierarchical strategy and lastly the Looping strategy.

Please keep in mind that this is just a principle. In practice, you will be unlikely to extensively interlink 10,000 pages. You must also choose the appropriate structure for sub-sections of your site. As we add more pages into the hierarchical structure, it becomes more successful. This is purely because more links are being added to the page that contains the outbound link (it is divided amongst more pages, most of which are yours). However, as we add more pages we also dilute the PageRank of the home page (which could be important to you).

What Google says

The best information about PageRank is, of course, straight from the horse's mouth. We therefore felt it would be helpful to ask Google themselves about PageRank. Obviously, we realize Google won't provide us with the exact specifics about PageRank or comment on the validity of what we've written. However, we did want to ask them some general questions to provide us with an

overview of how they see PageRank - now and in the future. We hope this this short interview will help debunk a few myths and answer some questions

Chris: Can you answer the age old question for us? Is PageRank named after the fact it is "page" based, or is it named after one of the creators?

Google: PageRank is named for Larry Page, Google co-founder and president, Products.

Chris: Does Google view PageRank as being a significant distinguishing feature from the other search engines?

Google: PageRank is a technology which contributes to the speed and relevance that differentiate Google from other search engines. In addition, Google employs 100 other algorithms to its ranking formula.

Chris: Does Google believe that PageRank significantly benefits the quality of its results pages and does it expect this to continue in the future?

Google: Because of the scalable approach PageRank uses to analyze links, it will continue to be a significant factor in Google's search results.

Chris: Several people have been known to try to data mine the PageRank information given by the toolbar, how does Google feel about this?

Google: Mining PageRank data from the Google Toolbar is against Google's terms of service.

Chris: How useful do you think the PageRank information is on the toolbar for a) normal users b) webmasters c) search engine optimization professionals?

Google: The PageRank information on the Google Toolbar is an estimate and is intended only for informational purposes. Many users have found it interesting and webmasters have been known to use it as a measure of their performance. However, because it is a rough estimate, the value to search engine optimization professionals is limited.

Chris: One method that people have been known to try to boost PageRank is by using link farms and guestbooks. How is Google likely to react to a site doing that?

Google: Google's engineers are constantly working to update Google's ranking algorithm to prevent manipulation of Google's rankings.

Final word to non mathematicians

It feels a little strange having a final word when it isn't the end of the document. But if you read the table of contents, you'll see that we've added an advanced section for the techies, mathematicians and Phd students. If your goal was just to find out about PageRank and what you can do about it, then you're done!

In summary, PageRank is probably the most misunderstood search engine concepts in existence today. There are those who profess to know everything about PageRank, and there are those who claim to not need to know anything. There are those who claim PageRank is important, and there are those who claim it is unimportant. There are those who claim Google oversells PageRank and there are those who do not believe this to be the case..

The truth of the matter is that nobody except Google can tell you the exact specifics. The authors of this document do not in any way profess to tell you everything. What is presented here are theories based on the evidence available. All we can do is to derive the principles, not exact figures.

To quantify the worth of PageRank is an impossible task. Under differing circumstances and differing Webmaster styles, it will have a different value. What PageRank is worth to you is probably an individual thing and you must decide how it fits into your philosophy and situation. You can do well by concentrating heavily on PageRank, or by not. However, you should always keep PageRank and the principles discussed here, in the back of your mind when designing sites. Just like you remember to create good Titles and anchor text, you should remember to consider PageRank.

The process of improving your PageRank should always be in symbiosis with your visitor experience. The highest PageRank in the world isn't worth much if your visitors can't navigate your site. Gaining PageRank should only be done whilst adding value to your site – whether that is through better targeting of visitors or improving content.

Advanced PageRank topics

As we delve into the advanced topics, we move more into the realm of speculation. Whilst we can be relatively sure that the principles derived earlier are sound, for the advanced PageRank discussion, we need to place a greater emphasis on assumption.

It should be said, however, that Google has surely thought of everything we're discussing here. It is therefore perfectly possible that they have done some of what we suggest below. If they have not already then, there's a good possibility they will in the future.

Speeding Up PageRank

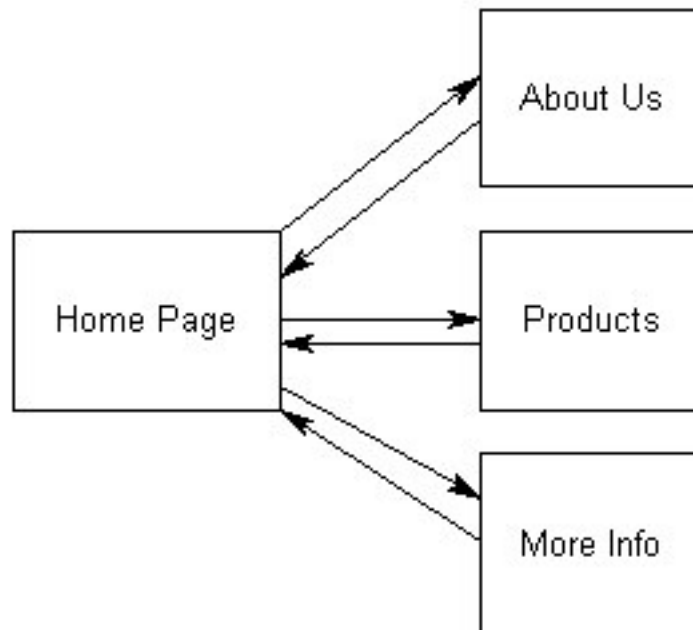
It is apparent that Google does not need a 100% accurate value for PageRank; so the question arises – “Can Google calculate an approximation of the final values without using as many iterations?” The answer appears to be yes. Let us explain. In previous calculations we have set the normalisation co-efficient at a

fixed amount. We can vary this co-efficient for each page depending upon the final value of the previous computation using the formula:

$$(1-d) * PR_i(A)$$

That is, we are replacing the (1-d) of the previous formula with the above. $PR_i(A)$ simply means with the PR deduced by the previous iteration

Let's do this by example. Assume we have the hierarchical structure mentioned earlier:



With this very simple structure we arrive at the final values like this:

	A	B	C	D	Sum
Start	1.000000000	1.000000000	1.000000000	1.000000000	4.000000000
1	2.700000000	0.433333333	0.433333333	0.433333333	4.000000000
2	1.255000000	0.915000000	0.915000000	0.915000000	4.000000000
3	2.483250000	0.505583333	0.505583333	0.505583333	4.000000000
4	1.439237500	0.853587500	0.853587500	0.853587500	4.000000000
5	2.326648125	0.557783958	0.557783958	0.557783958	4.000000000
6	1.572349093	0.809216968	0.809216968	0.809216968	4.000000000
7	2.213503270	0.595498909	0.595498909	0.595498909	4.000000000
8	1.668522220	0.777159259	0.777159259	0.777159259	4.000000000
9	2.131756112	0.622747962	0.622747962	0.622747962	4.000000000
10	1.738007304	0.753997565	0.753997565	0.753997565	4.000000000
11	2.072693795	0.642435402	0.642435402	0.642435402	4.000000000
12	1.788210272	0.737263240	0.737263240	0.737263240	4.000000000
13	2.030021264	0.656659578	0.656659578	0.656659578	4.000000000
14	1.824481925	0.725172691	0.725172691	0.725172691	4.000000000

15	1.9991903635	0.6669365455	0.6669365455	0.6669365455	4.0000000000
16	1.8506881910	0.7164372697	0.7164372697	0.7164372697	4.0000000000
17	1.9769150376	0.6743616541	0.6743616541	0.6743616541	4.0000000000
18	1.8696222180	0.7101259273	0.7101259273	0.7101259273	4.0000000000
19	1.9608211147	0.6797262951	0.6797262951	0.6797262951	4.0000000000
20	1.8833020525	0.7055659825	0.7055659825	0.7055659825	4.0000000000
21	1.9491932554	0.6836022482	0.6836022482	0.6836022482	4.0000000000
22	1.8931857329	0.7022714224	0.7022714224	0.7022714224	4.0000000000
23	1.9407921270	0.6864026243	0.6864026243	0.6864026243	4.0000000000
24	1.9003266921	0.6998911026	0.6998911026	0.6998911026	4.0000000000
25	1.9347223118	0.6884258961	0.6884258961	0.6884258961	4.0000000000
26	1.9054860350	0.6981713217	0.6981713217	0.6981713217	4.0000000000
27	1.9303368702	0.6898877099	0.6898877099	0.6898877099	4.0000000000
28	1.9092136603	0.6969287799	0.6969287799	0.6969287799	4.0000000000
29	1.9271683888	0.6909438704	0.6909438704	0.6909438704	4.0000000000
30	1.9119068696	0.6960310435	0.6960310435	0.6960310435	4.0000000000
31	1.9248791609	0.6917069464	0.6917069464	0.6917069464	4.0000000000
32	1.9138527133	0.6953824289	0.6953824289	0.6953824289	4.0000000000
33	1.9232251937	0.6922582688	0.6922582688	0.6922582688	4.0000000000
34	1.9152585853	0.6949138049	0.6949138049	0.6949138049	4.0000000000
35	1.9220302025	0.6926565992	0.6926565992	0.6926565992	4.0000000000
36	1.9162743279	0.6945752240	0.6945752240	0.6945752240	4.0000000000
37	1.9211668213	0.6929443929	0.6929443929	0.6929443929	4.0000000000
38	1.9170082019	0.6943305994	0.6943305994	0.6943305994	4.0000000000
39	1.9205430284	0.6931523239	0.6931523239	0.6931523239	4.0000000000
40	1.9175384259	0.6941538580	0.6941538580	0.6941538580	4.0000000000
41	1.9200923380	0.6933025540	0.6933025540	0.6933025540	4.0000000000
42	1.9179215127	0.6940261624	0.6940261624	0.6940261624	4.0000000000
43	1.9197667142	0.6934110953	0.6934110953	0.6934110953	4.0000000000
44	1.9181982929	0.6939339024	0.6939339024	0.6939339024	4.0000000000
45	1.9195314510	0.6934895163	0.6934895163	0.6934895163	4.0000000000
46	1.9183982666	0.6938672445	0.6938672445	0.6938672445	4.0000000000
47	1.9193614734	0.6935461755	0.6935461755	0.6935461755	4.0000000000
48	1.9185427476	0.6938190841	0.6938190841	0.6938190841	4.0000000000
49	1.9192386645	0.6935871118	0.6935871118	0.6935871118	4.0000000000
50	1.9186471352	0.6937842883	0.6937842883	0.6937842883	4.0000000000
51	1.9191499351	0.6936166883	0.6936166883	0.6936166883	4.0000000000
52	1.9187225552	0.6937591483	0.6937591483	0.6937591483	4.0000000000
53	1.9190858281	0.6936380573	0.6936380573	0.6936380573	4.0000000000
54	1.9187770461	0.6937409846	0.6937409846	0.6937409846	4.0000000000
55	1.9190395108	0.6936534964	0.6936534964	0.6936534964	4.0000000000
56	1.9188164158	0.6937278614	0.6937278614	0.6937278614	4.0000000000
57	1.9190060466	0.6936646511	0.6936646511	0.6936646511	4.0000000000
58	1.9188448604	0.6937183799	0.6937183799	0.6937183799	4.0000000000
59	1.9189818686	0.6936727105	0.6936727105	0.6936727105	4.0000000000
60	1.9188654117	0.6937115294	0.6937115294	0.6937115294	4.0000000000
61	1.9189644001	0.6936785333	0.6936785333	0.6936785333	4.0000000000
62	1.9188802599	0.6937065800	0.6937065800	0.6937065800	4.0000000000
63	1.9189517791	0.6936827403	0.6936827403	0.6936827403	4.0000000000
64	1.9188909878	0.6937030041	0.6937030041	0.6937030041	4.0000000000
65	1.9189426604	0.6936857799	0.6936857799	0.6936857799	4.0000000000

66	1.9188987387	0.6937004204	0.6937004204	0.6937004204	4.0000000000
67	1.9189360721	0.6936879760	0.6936879760	0.6936879760	4.0000000000
68	1.9189043387	0.6936985538	0.6936985538	0.6936985538	4.0000000000
69	1.9189313121	0.6936895626	0.6936895626	0.6936895626	4.0000000000
70	1.9189083847	0.6936972051	0.6936972051	0.6936972051	4.0000000000
71	1.9189278730	0.6936907090	0.6936907090	0.6936907090	4.0000000000
72	1.9189113080	0.6936962307	0.6936962307	0.6936962307	4.0000000000
73	1.9189253882	0.6936915373	0.6936915373	0.6936915373	4.0000000000
74	1.9189134200	0.6936955267	0.6936955267	0.6936955267	4.0000000000
75	1.9189235930	0.6936921357	0.6936921357	0.6936921357	4.0000000000
76	1.9189149459	0.6936950180	0.6936950180	0.6936950180	4.0000000000
77	1.9189222959	0.6936925680	0.6936925680	0.6936925680	4.0000000000
78	1.9189160484	0.6936946505	0.6936946505	0.6936946505	4.0000000000
79	1.9189213588	0.6936928804	0.6936928804	0.6936928804	4.0000000000
80	1.9189168450	0.6936943850	0.6936943850	0.6936943850	4.0000000000
81	1.9189206817	0.6936931061	0.6936931061	0.6936931061	4.0000000000
82	1.9189174205	0.6936941932	0.6936941932	0.6936941932	4.0000000000
83	1.9189201926	0.6936932691	0.6936932691	0.6936932691	4.0000000000
84	1.9189178363	0.6936940546	0.6936940546	0.6936940546	4.0000000000
85	1.9189198391	0.6936933870	0.6936933870	0.6936933870	4.0000000000
86	1.9189181367	0.6936939544	0.6936939544	0.6936939544	4.0000000000
87	1.9189195838	0.6936934721	0.6936934721	0.6936934721	4.0000000000
88	1.9189183538	0.6936938821	0.6936938821	0.6936938821	4.0000000000
89	1.9189193993	0.6936935336	0.6936935336	0.6936935336	4.0000000000
90	1.9189185106	0.6936938298	0.6936938298	0.6936938298	4.0000000000
91	1.9189192660	0.6936935780	0.6936935780	0.6936935780	4.0000000000
92	1.9189186239	0.6936937920	0.6936937920	0.6936937920	4.0000000000
93	1.9189191697	0.6936936101	0.6936936101	0.6936936101	4.0000000000
94	1.9189187058	0.6936937647	0.6936937647	0.6936937647	4.0000000000
95	1.9189191001	0.6936936333	0.6936936333	0.6936936333	4.0000000000
96	1.9189187649	0.6936937450	0.6936937450	0.6936937450	4.0000000000
97	1.9189190498	0.6936936501	0.6936936501	0.6936936501	4.0000000000
98	1.9189188077	0.6936937308	0.6936937308	0.6936937308	4.0000000000
99	1.9189190135	0.6936936622	0.6936936622	0.6936936622	4.0000000000
100	1.9189188385	0.6936937205	0.6936937205	0.6936937205	4.0000000000
101	1.9189189872	0.6936936709	0.6936936709	0.6936936709	4.0000000000
102	1.9189188608	0.6936937131	0.6936937131	0.6936937131	4.0000000000
103	1.9189189683	0.6936936772	0.6936936772	0.6936936772	4.0000000000
104	1.9189188770	0.6936937077	0.6936937077	0.6936937077	4.0000000000
105	1.9189189546	0.6936936818	0.6936936818	0.6936936818	4.0000000000
106	1.9189188886	0.6936937038	0.6936937038	0.6936937038	4.0000000000
107	1.9189189447	0.6936936851	0.6936936851	0.6936936851	4.0000000000
108	1.9189188970	0.6936937010	0.6936937010	0.6936937010	4.0000000000
109	1.9189189375	0.6936936875	0.6936936875	0.6936936875	4.0000000000
110	1.9189189031	0.6936936990	0.6936936990	0.6936936990	4.0000000000
111	1.9189189324	0.6936936892	0.6936936892	0.6936936892	4.0000000000
112	1.9189189075	0.6936936975	0.6936936975	0.6936936975	4.0000000000
113	1.9189189286	0.6936936905	0.6936936905	0.6936936905	4.0000000000
114	1.9189189107	0.6936936964	0.6936936964	0.6936936964	4.0000000000
115	1.9189189259	0.6936936914	0.6936936914	0.6936936914	4.0000000000
116	1.9189189130	0.6936936957	0.6936936957	0.6936936957	4.0000000000

117	1.9189189240	0.6936936920	0.6936936920	0.6936936920	4.0000000000
118	1.9189189146	0.6936936951	0.6936936951	0.6936936951	4.0000000000
119	1.9189189226	0.6936936925	0.6936936925	0.6936936925	4.0000000000
120	1.9189189158	0.6936936947	0.6936936947	0.6936936947	4.0000000000
121	1.9189189216	0.6936936928	0.6936936928	0.6936936928	4.0000000000
122	1.9189189167	0.6936936944	0.6936936944	0.6936936944	4.0000000000
123	1.9189189208	0.6936936931	0.6936936931	0.6936936931	4.0000000000
124	1.9189189173	0.6936936942	0.6936936942	0.6936936942	4.0000000000
125	1.9189189203	0.6936936932	0.6936936932	0.6936936932	4.0000000000
126	1.9189189177	0.6936936941	0.6936936941	0.6936936941	4.0000000000
127	1.9189189199	0.6936936934	0.6936936934	0.6936936934	4.0000000000
128	1.9189189181	0.6936936940	0.6936936940	0.6936936940	4.0000000000
129	1.9189189196	0.6936936935	0.6936936935	0.6936936935	4.0000000000
130	1.9189189183	0.6936936939	0.6936936939	0.6936936939	4.0000000000
131	1.9189189194	0.6936936935	0.6936936935	0.6936936935	4.0000000000
132	1.9189189185	0.6936936938	0.6936936938	0.6936936938	4.0000000000
133	1.9189189193	0.6936936936	0.6936936936	0.6936936936	4.0000000000
134	1.9189189186	0.6936936938	0.6936936938	0.6936936938	4.0000000000
135	1.9189189192	0.6936936936	0.6936936936	0.6936936936	4.0000000000
136	1.9189189187	0.6936936938	0.6936936938	0.6936936938	4.0000000000
137	1.9189189191	0.6936936936	0.6936936936	0.6936936936	4.0000000000
138	1.9189189188	0.6936936937	0.6936936937	0.6936936937	4.0000000000
139	1.9189189191	0.6936936936	0.6936936936	0.6936936936	4.0000000000
140	1.9189189188	0.6936936937	0.6936936937	0.6936936937	4.0000000000
141	1.9189189190	0.6936936937	0.6936936937	0.6936936937	4.0000000000
142	1.9189189188	0.6936936937	0.6936936937	0.6936936937	4.0000000000
143	1.9189189190	0.6936936937	0.6936936937	0.6936936937	4.0000000000
144	1.9189189189	0.6936936937	0.6936936937	0.6936936937	4.0000000000
145	1.9189189190	0.6936936937	0.6936936937	0.6936936937	4.0000000000
146	1.9189189189	0.6936936937	0.6936936937	0.6936936937	4.0000000000
147	1.9189189190	0.6936936937	0.6936936937	0.6936936937	4.0000000000
148	1.9189189189	0.6936936937	0.6936936937	0.6936936937	4.0000000000

[Excel Example 17]

As you can see, it takes 148 iterations. Now, if we recalculate the normalization co-efficient at each iteration, we get the following:

	A	B	C	D	Sum
Start	1.0000000000	1.0000000000	1.0000000000	1.0000000000	4.0000000000
1	2.7000000000	0.4333333333	0.4333333333	0.4333333333	4.0000000000
2	1.5100000000	0.8300000000	0.8300000000	0.8300000000	4.0000000000
3	2.3430000000	0.5523333333	0.5523333333	0.5523333333	4.0000000000
4	1.7599000000	0.7467000000	0.7467000000	0.7467000000	4.0000000000
5	2.1680700000	0.6106433333	0.6106433333	0.6106433333	4.0000000000
6	1.8823510000	0.7058830000	0.7058830000	0.7058830000	4.0000000000
7	2.0823543000	0.6392152333	0.6392152333	0.6392152333	4.0000000000
8	1.9423519900	0.6858826700	0.6858826700	0.6858826700	4.0000000000
9	2.0403536070	0.6532154643	0.6532154643	0.6532154643	4.0000000000
10	1.9717524751	0.6760825083	0.6760825083	0.6760825083	4.0000000000
11	2.0197732674	0.6600755775	0.6600755775	0.6600755775	4.0000000000
12	1.9861587128	0.6712804291	0.6712804291	0.6712804291	4.0000000000

13	2.0096889010	0.6634370330	0.6634370330	0.6634370330	4.0000000000
14	1.9932177693	0.6689274102	0.6689274102	0.6689274102	4.0000000000
15	2.0047475615	0.6650841462	0.6650841462	0.6650841462	4.0000000000
16	1.9966767069	0.6677744310	0.6677744310	0.6677744310	4.0000000000
17	2.0023263051	0.6658912316	0.6658912316	0.6658912316	4.0000000000
18	1.9983715864	0.6672094712	0.6672094712	0.6672094712	4.0000000000
19	2.0011398895	0.6662867035	0.6662867035	0.6662867035	4.0000000000
20	1.9992020773	0.6669326409	0.6669326409	0.6669326409	4.0000000000
21	2.0005585459	0.6664804847	0.6664804847	0.6664804847	4.0000000000
22	1.9996090179	0.6667969940	0.6667969940	0.6667969940	4.0000000000
23	2.0002736875	0.6665754375	0.6665754375	0.6665754375	4.0000000000
24	1.9998084188	0.6667305271	0.6667305271	0.6667305271	4.0000000000
25	2.0001341069	0.6666219644	0.6666219644	0.6666219644	4.0000000000
26	1.9999061252	0.6666979583	0.6666979583	0.6666979583	4.0000000000
27	2.0000657124	0.6666447625	0.6666447625	0.6666447625	4.0000000000
28	1.9999540013	0.6666819996	0.6666819996	0.6666819996	4.0000000000
29	2.0000321991	0.6666559336	0.6666559336	0.6666559336	4.0000000000
30	1.9999774607	0.6666741798	0.6666741798	0.6666741798	4.0000000000
31	2.0000157775	0.6666614075	0.6666614075	0.6666614075	4.0000000000
32	1.9999889557	0.6666703481	0.6666703481	0.6666703481	4.0000000000
33	2.0000077310	0.6666640897	0.6666640897	0.6666640897	4.0000000000
34	1.9999945883	0.6666684706	0.6666684706	0.6666684706	4.0000000000
35	2.0000037882	0.6666654039	0.6666654039	0.6666654039	4.0000000000
36	1.9999973483	0.6666675506	0.6666675506	0.6666675506	4.0000000000
37	2.0000018562	0.6666660479	0.6666660479	0.6666660479	4.0000000000
38	1.9999987007	0.6666670998	0.6666670998	0.6666670998	4.0000000000
39	2.0000009095	0.6666663635	0.6666663635	0.6666663635	4.0000000000
40	1.9999993633	0.6666668789	0.6666668789	0.6666668789	4.0000000000
41	2.0000004457	0.6666665181	0.6666665181	0.6666665181	4.0000000000
42	1.9999996880	0.6666667707	0.6666667707	0.6666667707	4.0000000000
43	2.0000002184	0.6666665939	0.6666665939	0.6666665939	4.0000000000
44	1.9999998471	0.6666667176	0.6666667176	0.6666667176	4.0000000000
45	2.0000001070	0.6666666310	0.6666666310	0.6666666310	4.0000000000
46	1.9999999251	0.6666666916	0.6666666916	0.6666666916	4.0000000000
47	2.0000000524	0.6666666492	0.6666666492	0.6666666492	4.0000000000
48	1.9999999633	0.6666666789	0.6666666789	0.6666666789	4.0000000000
49	2.0000000257	0.6666666581	0.6666666581	0.6666666581	4.0000000000
50	1.9999999820	0.6666666727	0.6666666727	0.6666666727	4.0000000000
51	2.0000000126	0.6666666625	0.6666666625	0.6666666625	4.0000000000
52	1.9999999912	0.6666666696	0.6666666696	0.6666666696	4.0000000000
53	2.0000000062	0.6666666646	0.6666666646	0.6666666646	4.0000000000
54	1.9999999957	0.6666666681	0.6666666681	0.6666666681	4.0000000000
55	2.0000000030	0.6666666657	0.6666666657	0.6666666657	4.0000000000
56	1.9999999979	0.6666666674	0.6666666674	0.6666666674	4.0000000000
57	2.0000000015	0.6666666662	0.6666666662	0.6666666662	4.0000000000
58	1.9999999990	0.6666666670	0.6666666670	0.6666666670	4.0000000000
59	2.0000000007	0.6666666664	0.6666666664	0.6666666664	4.0000000000
60	1.9999999995	0.6666666668	0.6666666668	0.6666666668	4.0000000000
61	2.0000000004	0.6666666665	0.6666666665	0.6666666665	4.0000000000
62	1.9999999998	0.6666666667	0.6666666667	0.6666666667	4.0000000000
63	2.0000000002	0.6666666666	0.6666666666	0.6666666666	4.0000000000

64	1.9999999999	0.6666666667	0.6666666667	0.6666666667	4.0000000000
65	2.0000000001	0.6666666666	0.6666666666	0.6666666666	4.0000000000
66	1.9999999999	0.6666666667	0.6666666667	0.6666666667	4.0000000000
67	2.0000000000	0.6666666667	0.6666666667	0.6666666667	4.0000000000

[Excel Example 18]

A very good approximation of the values with only 67 calculations. This is known as the “accelerated technology of the computation of the Page Rank estimating values.” So by being a little bit clever, Google can considerably reduce the computation time needed to calculate PageRank!

How PageRank could be more than it seems

There’s a possibility, just a possibility, that PageRank could well be more than it seems. If we think about the factors that could affect a site’s ranking in Google we come up with the obvious things such as PageRank, keywords, anchor text and so on. But where do the less obvious factors like “quantity of text” fit? What if, for example, some pages were preferred because they had the correct quantity of text? This doesn’t seem to fit into the traditional factors, but it can be worked into PageRank. Let’s explain:

How we distribute the normalization co-efficients is pretty much up to us, as long as the total value stays the same. So with our previous example, we may set the “About Us” page (Page B) higher, and lower the rest of the pages because page B meets some criteria we define (e.g., it has the correct amount of text). So we could set page B to:

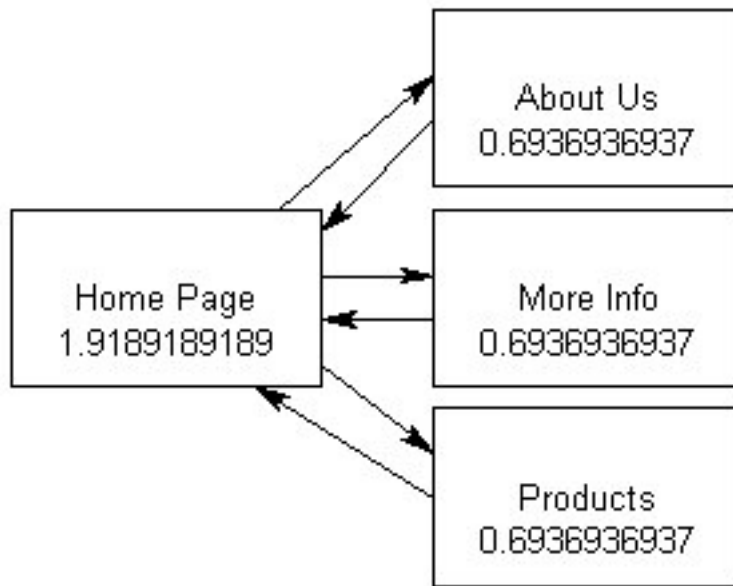
$$(0.4*n + 0.6)*0.15$$

as long as we set the other pages to:

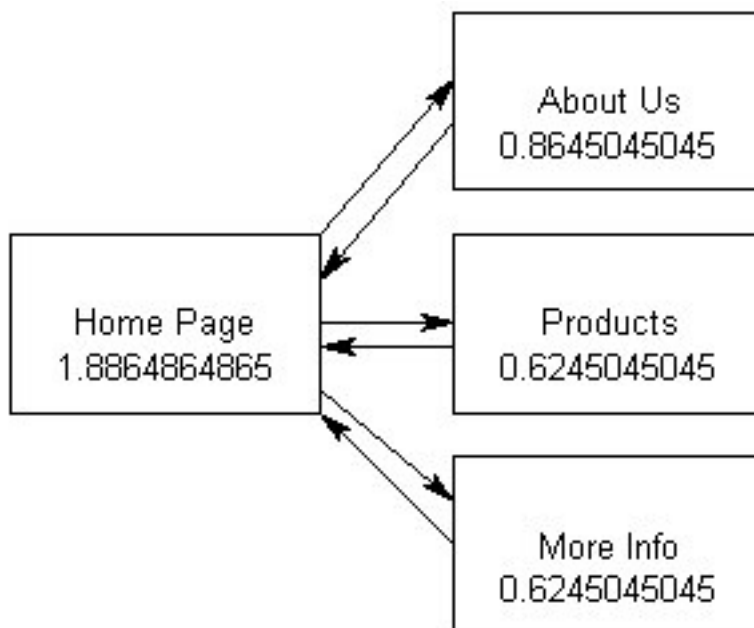
$$0.6*0.15$$

i.e. $[0.1*n + 0.9+0.9* (n-1)=n$, where n is the number of Pages in the Site.]

This strongly favours Page B. Let’s give an example. Normally, the final values for the previous example site structure would be:



Using our new values for the co-efficients in an attempt to show favour to Page B (that meets our arbitrary criteria), we get:



[Excel Example 19]

As you can see, we've boosted the PageRank of the About Us page at the expense of all other pages – simply because we've introduced a new arbitrary factor called "amount of text."

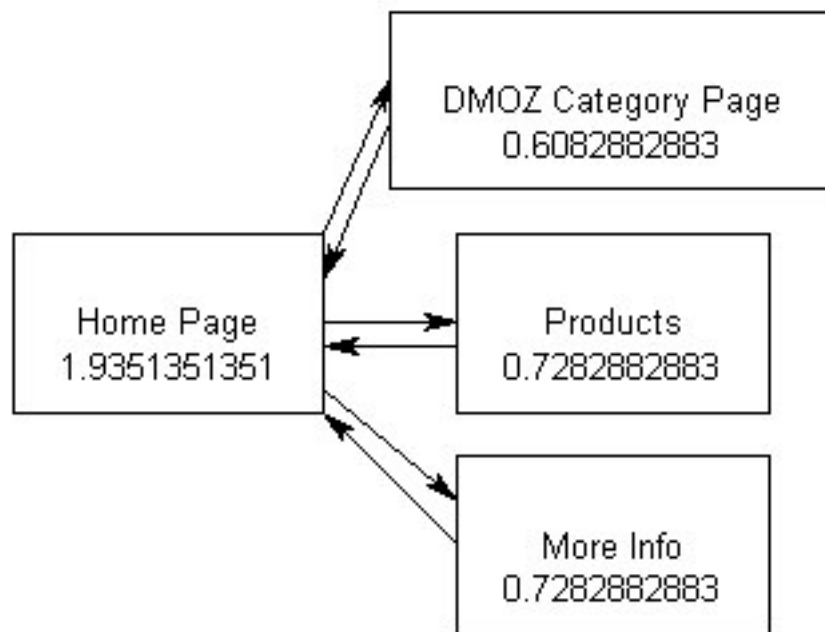
PageRank not only considers links, but it may also consider certain other factors.

There is of course, no reason why we should stop at boosting a single page's PageRank – it is just as easy to boost an entire site's PageRank based on any factor we like. Examples of possible factors are "it's popular" or "we like it"!

Favouring pages in directories

It stands to reason that if we can cause an increase in the PageRank of a page or pages at the expense of those it links to and from, then we can also cause a decrease in its PageRank, to the benefit of those it links to and from. Why would we want to do this? Because the page represents a list of links that points to high-quality pages. Let's pretend that Page B is such a list, say a category page with links in ODP.

We'll set Page B's normalization co-efficient to 0.06 and the rest to 0.18. We end up with:



[Excel Example 20]

And sure enough, it does exactly what we want it to do.

This has several advantages. The page itself, which is actually just a list of links, will be pushed down in the rankings. Whilst, the pages it links to will get a boost! Thus, the searcher is less likely to see pages that are just links and more likely to see good quality pages. In addition, those with the Google Toolbar installed won't be racing each other to get links on such pages, because their PageRank will appear deceptively low!

Examples

Throughout this document you've seen the phrase [Excel Example x]. This refers to the actual excel spreadsheet that contains the calculations used to produce this example (so you can look at it in more depth if you want to).

You can get the examples from: <http://www.supportforums.org/pagerank/>

Mathematical aspects of PageRank for advanced readers.

The theoretical concept of PageRank interprets Web sites as directed multigraphs with pages on their tops. Googlebot (Google's spider), explores a site, traces links between pages, and forms the matrix of links, which contains all information about the topology of that site.

Let's write down the elements of the matrix of links $\|a_{km}\|$, ($k=1,\dots,n$), ($m=1,\dots,n$).

What does that mean? Let us explain the designations:

T_1, T_2, \dots, T_n correspond to Web site pages (n is the number of pages)

Components a_{km} , which are on the intersection of the k -line and m -column in the matrix of links determine the number of links off T_k to T_m .

$C_r(T_k)$ is the total number of links off T_k .

DIAGRAM OF THE MATRIX OF LINKS (1)

	T_1	T_2			T_n	
T_1	a_{11}	a_{12}	a_{1n}	$C_r(T_1)$
T_2	a_{21}	a_{22}	a_{2n}	$C_r(T_2)$
.	.	.			.	
.	.	.			.	
.	.	.			.	
.	.	.			.	
T_n	a_{n1}	a_{n2}	a_{nn}	$C_r(T_n)$

Number of links off
the page

Let's write down the equation for the computation scheme.

$$PR_{l+1}(T_m) = (1-d) + d \sum_{k=1}^n a_{km} PR_l(T_k) / C_r(T_k),$$

Where

$PR_{l+1}(T_m)$ is Page Rank of the Page T_m on $(l+1)$ iteration (the estimating value of the page T_m ($m=1, \dots, n$)).

$PR_l(T_k)$ is the Page Rank of the Page T_k on the l iteration (the estimating value of the Page T_k ($k=1, \dots, n$)).

$C_r(T_k)$ is the total number of links off the page T_k ,

$C_m(T_k) = a_{km}$ is the number of the individual links off the page T_k to the page T_m ($k=1, \dots, n$) ($m=1, \dots, n$).

d is a dampening factor (usually 0,85).

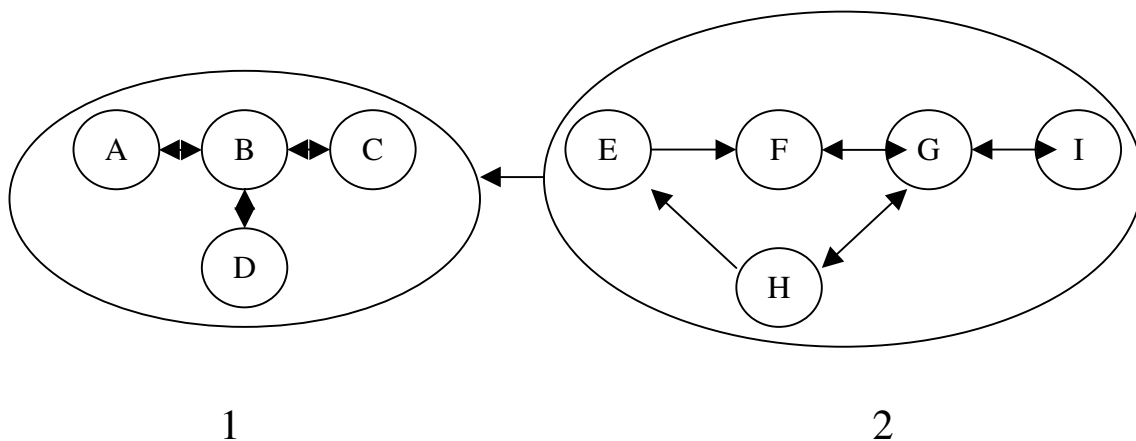
$(1-d)$ is a normalization coefficient,

n is the number of the Pages on the Site.

This equation of a recursive computational scheme allows taking into consideration, and then computing, the estimating values of a Web site paying attention to the probable complexity of links between pages.

Take a look at the following example. Generally, it is a multigraph of a small Web site. Let's apply PageRank methods to this site.

MULTIGRAPH OF A SIMPLE WEB SITE



You may have already noticed that every page of the second subgraph links to every page of the first subgraph.

Let's construct a matrix of links of this site. The values A, B, C, D, E, F, G, H, I refer to the corresponding pages.

DIAGRAM OF THE MATRIX OF LINKS (2)

		A	B	C	D	E	F	G	H	I	Number of links off the page	
Home.index	A	0	1	0	0	0	0	0	0	0		$C_r(A)$
	B	1	0	1	1	0	0	0	0	0	$C_r(B)$	3
	C	0	1	0	0	0	0	0	0	0	$C_r(C)$	1
	D	0	1	0	0	0	0	0	0	0	$C_r(D)$	1
	E	1	1	1	1	0	1	0	0	0	$C_r(E)$	5
	F	1	1	1	1	0	0	1	0	0	$C_r(F)$	5
	G	1	1	1	1	0	1	0	1	1	$C_r(G)$	7
	H	1	1	1	1	1	0	1	0	0	$C_r(H)$	6
	I	1	1	1	1	0	0	1	0	0	$C_r(I)$	5

Note the wonderful ability of PageRank to distinguish the most important pages and the relationship of external pages to them. Later we will use the term “base pages” to describe those important pages. The main feature of the base pages is that they are traversed by a directed cycle.

Let's apply the so-called accelerated technology of computation of the estimating values of the site pages.

Suppose, $PR_i(T)$ is the estimating value for the page T (T=A, B, C, D, E, F, G, H, I) page on i-iteration.

The normalization coefficient for the $PR_{i+1}(T)$ computation (on (i + 1)-iteration) can be calculated by formula: $(1-d) \cdot PR_i(T)$.

Applying the accelerated technology, the computation of the estimating values of the site pages the limiting values of the estimating values of external pages in relation to the base sites are equal to zero.

This approach permits us to distinguish the external pages in relation to the base pages. Thus, we can introduce the following definition:

The external pages in relation to the base pages are ones that have zero limiting values on application of the accelerated technology of the computation of the estimating values of the site pages.

ACCELERATED COMPUTATION OF PAGERANK for recognition of external links

Iteration	A	B	C	D	E	F	G	H	I	Amount
	1,000000	1,000000	1,000000	1,000000	1,000000	1,000000	1,000000	1,000000	1,000000	9,000
1	1,206429	3,473095	1,206429	1,206429	0,291667	0,441429	0,631667	0,271429	0,271429	9,000
2	1,450932	3,883281	1,450932	1,450932	0,082202	0,192500	0,254388	0,117417	0,117417	9,000
3	1,432087	4,396552	1,432087	1,432087	0,028964	0,073739	0,107478	0,048502	0,048502	9,000
4	1,506130	4,356932	1,506130	1,506130	0,011216	0,029036	0,043774	0,020326	0,020326	9,000
5	1,478877	4,512665	1,478877	1,478877	0,004562	0,011577	0,017837	0,008364	0,008364	9,000
6	1,507936	4,455552	1,507936	1,507936	0,001869	0,004678	0,007251	0,003421	0,003421	9,000
7	1,491656	4,516630	1,491656	1,491656	0,000765	0,001900	0,002949	0,001394	0,001394	9,000
8	1,504706	4,482464	1,504706	1,504706	0,000312	0,000773	0,001200	0,000567	0,000567	9,000
9	1,496244	4,509876	1,496244	1,496244	0,000127	0,000315	0,000488	0,000231	0,000231	9,000
10	1,502441	4,492110	1,502441	1,502441	0,000052	0,000128	0,000199	0,000094	0,000094	9,000
11	1,498215	4,505125	1,498215	1,498215	0,000021	0,000052	0,000081	0,000038	0,000038	9,000
12	1,501219	4,496251	1,501219	1,501219	0,000009	0,000021	0,000033	0,000016	0,000016	9,000
13	1,499134	4,502559	1,499134	1,499134	0,000003	0,000009	0,000013	0,000006	0,000006	9,000
14	1,500601	4,498182	1,500601	1,500601	0,000001	0,000004	0,000005	0,000003	0,000003	9,000
15	1,499577	4,501262	1,499577	1,499577	0,000001	0,000001	0,000002	0,000001	0,000001	9,000
16	1,500295	4,499112	1,500295	1,500295	0,000000	0,000001	0,000001	0,000000	0,000000	9,000
17	1,499793	4,500620	1,499793	1,499793	0,000000	0,000000	0,000000	0,000000	0,000000	9,000
18	1,500145	4,499566	1,500145	1,500145	0,000000	0,000000	0,000000	0,000000	0,000000	9,000
19	1,499899	4,500304	1,499899	1,499899	0,000000	0,000000	0,000000	0,000000	0,000000	9,000
20	1,500071	4,499787	1,500071	1,500071	0,000000	0,000000	0,000000	0,000000	0,000000	9,000
21	1,499950	4,500149	1,499950	1,499950	0,000000	0,000000	0,000000	0,000000	0,000000	9,000
22	1,500035	4,499896	1,500035	1,500035	0,000000	0,000000	0,000000	0,000000	0,000000	9,000
23	1,499976	4,500073	1,499976	1,499976	0,000000	0,000000	0,000000	0,000000	0,000000	9,000
24	1,500017	4,499949	1,500017	1,500017	0,000000	0,000000	0,000000	0,000000	0,000000	9,000
25	1,499988	4,500036	1,499988	1,499988	0,000000	0,000000	0,000000	0,000000	0,000000	9,000
26	1,500008	4,499975	1,500008	1,500008	0,000000	0,000000	0,000000	0,000000	0,000000	9,000

27	1,499994	4,500018	1,499994	1,499994	0,000000	0,000000	0,000000	0,000000	0,000000	9,000
28	1,500004	4,499988	1,500004	1,500004	0,000000	0,000000	0,000000	0,000000	0,000000	9,000
29	1,499997	4,500009	1,499997	1,499997	0,000000	0,000000	0,000000	0,000000	0,000000	9,000
30	1,500002	4,499994	1,500002	1,500002	0,000000	0,000000	0,000000	0,000000	0,000000	9,000
31	1,499999	4,500004	1,499999	1,499999	0,000000	0,000000	0,000000	0,000000	0,000000	9,000
32	1,500001	4,499997	1,500001	1,500001	0,000000	0,000000	0,000000	0,000000	0,000000	9,000
33	1,499999	4,500002	1,499999	1,499999	0,000000	0,000000	0,000000	0,000000	0,000000	9,000
34	1,500000	4,499999	1,500000	1,500000	0,000000	0,000000	0,000000	0,000000	0,000000	9,000
35	1,500000	4,500001	1,500000	1,500000	0,000000	0,000000	0,000000	0,000000	0,000000	9,000
36	1,500000	4,499999	1,500000	1,500000	0,000000	0,000000	0,000000	0,000000	0,000000	9,000
37	1,500000	4,500000	1,500000	1,500000	0,000000	0,000000	0,000000	0,000000	0,000000	9,000
38	1,500000	4,500000	1,500000	1,500000	0,000000	0,000000	0,000000	0,000000	0,000000	9,000
39	1,500000	4,500000	1,500000	1,500000	0,000000	0,000000	0,000000	0,000000	0,000000	9,000

For the Web site that we have just examined the pages E, F, G, I, H are external in relation to the base pages A, B, C, D.

Let's calculate the estimating values for the pages of the Web site examined with the constant normalization coefficient $(1-d) = 0.15$, $d = 0.85$

CLASSIC COMPUTATION OF PAGERANK

Iteration	A	B	C	D	E	F	G	H	I	Amount
	1,000000	1,000000	1,000000	1,000000	1,000000	1,000000	1,000000	1,000000	1,000000	9,000
1	1,206429	3,473095	1,206429	1,206429	0,291667	0,441429	0,631667	0,271429	0,271429	9,000
2	1,419967	3,512317	1,419967	1,419967	0,188452	0,276286	0,309638	0,226702	0,226702	9,000
3	1,332416	3,958177	1,332416	1,332416	0,182116	0,219636	0,267624	0,187599	0,187599	9,000
4	1,430747	3,706925	1,430747	1,430747	0,176577	0,213457	0,245806	0,182497	0,182497	9,000
5	1,353327	3,951436	1,353327	1,353327	0,175854	0,209866	0,243166	0,179848	0,179848	9,000
6	1,420726	3,752137	1,420726	1,420726	0,175478	0,209422	0,241730	0,179527	0,179527	9,000
7	1,363844	3,923590	1,363844	1,363844	0,175433	0,209184	0,241554	0,179353	0,179353	9,000
8	1,412299	3,778418	1,412299	1,412299	0,175408	0,209155	0,241460	0,179332	0,179332	9,000
9	1,371139	3,901949	1,371139	1,371139	0,175405	0,209140	0,241448	0,179320	0,179320	9,000
10	1,406132	3,796985	1,406132	1,406132	0,175404	0,209138	0,241442	0,179319	0,179319	9,000
11	1,376390	3,886213	1,376390	1,376390	0,175403	0,209137	0,241441	0,179318	0,179318	9,000
12	1,401671	3,810371	1,401671	1,401671	0,175403	0,209136	0,241441	0,179318	0,179318	9,000
13	1,380182	3,874838	1,380182	1,380182	0,175403	0,209136	0,241441	0,179318	0,179318	9,000
14	1,398448	3,820041	1,398448	1,398448	0,175403	0,209136	0,241441	0,179318	0,179318	9,000
15	1,382922	3,866618	1,382922	1,382922	0,175403	0,209136	0,241441	0,179318	0,179318	9,000
16	1,396119	3,827028	1,396119	1,396119	0,175403	0,209136	0,241441	0,179318	0,179318	9,000
17	1,384901	3,860680	1,384901	1,384901	0,175403	0,209136	0,241441	0,179318	0,179318	9,000
18	1,394436	3,832076	1,394436	1,394436	0,175403	0,209136	0,241441	0,179318	0,179318	9,000
19	1,386332	3,856389	1,386332	1,386332	0,175403	0,209136	0,241441	0,179318	0,179318	9,000
20	1,393220	3,835723	1,393220	1,393220	0,175403	0,209136	0,241441	0,179318	0,179318	9,000

21	1,387365	3,853289	1,387365	1,387365	0,175403	0,209136	0,241441	0,179318	0,179318	9,000
22	1,392342	3,838358	1,392342	1,392342	0,175403	0,209136	0,241441	0,179318	0,179318	9,000
23	1,388112	3,851049	1,388112	1,388112	0,175403	0,209136	0,241441	0,179318	0,179318	9,000
24	1,391708	3,840261	1,391708	1,391708	0,175403	0,209136	0,241441	0,179318	0,179318	9,000
25	1,388651	3,849431	1,388651	1,388651	0,175403	0,209136	0,241441	0,179318	0,179318	9,000
26	1,391249	3,841637	1,391249	1,391249	0,175403	0,209136	0,241441	0,179318	0,179318	9,000
27	1,389041	3,848262	1,389041	1,389041	0,175403	0,209136	0,241441	0,179318	0,179318	9,000
28	1,390918	3,842631	1,390918	1,390918	0,175403	0,209136	0,241441	0,179318	0,179318	9,000
29	1,389322	3,847417	1,389322	1,389322	0,175403	0,209136	0,241441	0,179318	0,179318	9,000
30	1,390678	3,843349	1,390678	1,390678	0,175403	0,209136	0,241441	0,179318	0,179318	9,000
31	1,389526	3,846807	1,389526	1,389526	0,175403	0,209136	0,241441	0,179318	0,179318	9,000
32	1,390506	3,843867	1,390506	1,390506	0,175403	0,209136	0,241441	0,179318	0,179318	9,000
33	1,389673	3,846366	1,389673	1,389673	0,175403	0,209136	0,241441	0,179318	0,179318	9,000
34	1,390381	3,844242	1,390381	1,390381	0,175403	0,209136	0,241441	0,179318	0,179318	9,000
35	1,389779	3,846048	1,389779	1,389779	0,175403	0,209136	0,241441	0,179318	0,179318	9,000
36	1,390290	3,844513	1,390290	1,390290	0,175403	0,209136	0,241441	0,179318	0,179318	9,000
37	1,389856	3,845817	1,389856	1,389856	0,175403	0,209136	0,241441	0,179318	0,179318	9,000
38	1,390225	3,844709	1,390225	1,390225	0,175403	0,209136	0,241441	0,179318	0,179318	9,000
39	1,389911	3,845651	1,389911	1,389911	0,175403	0,209136	0,241441	0,179318	0,179318	9,000
40	1,390178	3,844850	1,390178	1,390178	0,175403	0,209136	0,241441	0,179318	0,179318	9,000
41	1,389951	3,845531	1,389951	1,389951	0,175403	0,209136	0,241441	0,179318	0,179318	9,000
42	1,390144	3,844952	1,390144	1,390144	0,175403	0,209136	0,241441	0,179318	0,179318	9,000
43	1,389980	3,845444	1,389980	1,389980	0,175403	0,209136	0,241441	0,179318	0,179318	9,000
44	1,390119	3,845026	1,390119	1,390119	0,175403	0,209136	0,241441	0,179318	0,179318	9,000
45	1,390001	3,845381	1,390001	1,390001	0,175403	0,209136	0,241441	0,179318	0,179318	9,000
46	1,390102	3,845079	1,390102	1,390102	0,175403	0,209136	0,241441	0,179318	0,179318	9,000
47	1,390016	3,845336	1,390016	1,390016	0,175403	0,209136	0,241441	0,179318	0,179318	9,000
48	1,390089	3,845118	1,390089	1,390089	0,175403	0,209136	0,241441	0,179318	0,179318	9,000
49	1,390027	3,845303	1,390027	1,390027	0,175403	0,209136	0,241441	0,179318	0,179318	9,000
50	1,390080	3,845146	1,390080	1,390080	0,175403	0,209136	0,241441	0,179318	0,179318	9,000
51	1,390035	3,845280	1,390035	1,390035	0,175403	0,209136	0,241441	0,179318	0,179318	9,000
52	1,390073	3,845166	1,390073	1,390073	0,175403	0,209136	0,241441	0,179318	0,179318	9,000
53	1,390041	3,845263	1,390041	1,390041	0,175403	0,209136	0,241441	0,179318	0,179318	9,000
54	1,390068	3,845180	1,390068	1,390068	0,175403	0,209136	0,241441	0,179318	0,179318	9,000
55	1,390045	3,845250	1,390045	1,390045	0,175403	0,209136	0,241441	0,179318	0,179318	9,000
56	1,390064	3,845191	1,390064	1,390064	0,175403	0,209136	0,241441	0,179318	0,179318	9,000
57	1,390048	3,845241	1,390048	1,390048	0,175403	0,209136	0,241441	0,179318	0,179318	9,000
58	1,390062	3,845198	1,390062	1,390062	0,175403	0,209136	0,241441	0,179318	0,179318	9,000
59	1,390050	3,845235	1,390050	1,390050	0,175403	0,209136	0,241441	0,179318	0,179318	9,000
60	1,390060	3,845204	1,390060	1,390060	0,175403	0,209136	0,241441	0,179318	0,179318	9,000
61	1,390051	3,845230	1,390051	1,390051	0,175403	0,209136	0,241441	0,179318	0,179318	9,000
62	1,390059	3,845208	1,390059	1,390059	0,175403	0,209136	0,241441	0,179318	0,179318	9,000
63	1,390052	3,845227	1,390052	1,390052	0,175403	0,209136	0,241441	0,179318	0,179318	9,000
64	1,390058	3,845211	1,390058	1,390058	0,175403	0,209136	0,241441	0,179318	0,179318	9,000
65	1,390053	3,845224	1,390053	1,390053	0,175403	0,209136	0,241441	0,179318	0,179318	9,000
66	1,390057	3,845213	1,390057	1,390057	0,175403	0,209136	0,241441	0,179318	0,179318	9,000
67	1,390054	3,845223	1,390054	1,390054	0,175403	0,209136	0,241441	0,179318	0,179318	9,000
68	1,390057	3,845214	1,390057	1,390057	0,175403	0,209136	0,241441	0,179318	0,179318	9,000
69	1,390054	3,845221	1,390054	1,390054	0,175403	0,209136	0,241441	0,179318	0,179318	9,000
70	1,390056	3,845215	1,390056	1,390056	0,175403	0,209136	0,241441	0,179318	0,179318	9,000

71	1,390055	3,845220	1,390055	1,390055	0,175403	0,209136	0,241441	0,179318	0,179318	9,000
72	1,390056	3,845216	1,390056	1,390056	0,175403	0,209136	0,241441	0,179318	0,179318	9,000
73	1,390055	3,845220	1,390055	1,390055	0,175403	0,209136	0,241441	0,179318	0,179318	9,000
74	1,390056	3,845217	1,390056	1,390056	0,175403	0,209136	0,241441	0,179318	0,179318	9,000
75	1,390055	3,845219	1,390055	1,390055	0,175403	0,209136	0,241441	0,179318	0,179318	9,000
76	1,390056	3,845217	1,390056	1,390056	0,175403	0,209136	0,241441	0,179318	0,179318	9,000
77	1,390055	3,845219	1,390055	1,390055	0,175403	0,209136	0,241441	0,179318	0,179318	9,000
78	1,390056	3,845217	1,390056	1,390056	0,175403	0,209136	0,241441	0,179318	0,179318	9,000
79	1,390055	3,845219	1,390055	1,390055	0,175403	0,209136	0,241441	0,179318	0,179318	9,000
80	1,390056	3,845218	1,390056	1,390056	0,175403	0,209136	0,241441	0,179318	0,179318	9,000
81	1,390055	3,845219	1,390055	1,390055	0,175403	0,209136	0,241441	0,179318	0,179318	9,000
82	1,390055	3,845218	1,390055	1,390055	0,175403	0,209136	0,241441	0,179318	0,179318	9,000
83	1,390055	3,845218	1,390055	1,390055	0,175403	0,209136	0,241441	0,179318	0,179318	9,000
84	1,390055	3,845218	1,390055	1,390055	0,175403	0,209136	0,241441	0,179318	0,179318	9,000
85	1,390055	3,845218	1,390055	1,390055	0,175403	0,209136	0,241441	0,179318	0,179318	9,000

The limiting estimating values for the site pages are darkened.

Having divided the estimating values to their total amount, which is unchangeable upon execution of the iteration process, we'll receive the set of probability values (the amount of all probability values is equal to one).

										Amount
P	0,154451	0,427246	0,154451	0,154451	0,019489	0,023237	0,026827	0,019924	0,019924	1,000

Running a parallel demonstration, we'd like to use the famous classical Shannon's formula, which was normally used for calculation of the volume of the information, which an experiment contains, and which has n outcomes with the set probabilities.

Experiment outcome	A_1	A_2	...	A_n
Probabilities	p_1	p_2	...	p_n

$$H(p_1, p_2, \dots, p_n) = - p(A_1) \cdot \log p(A_1) - p(A_2) \cdot \log p(A_2) - \dots - p(A_n) \cdot \log p(A_n).$$

Let's calculate the values for a set of probability values, which correspond to the site pages, and we'll receive the following set of the values.

This set of values of Page Rank for the pages of the site examined. (Logarithms are to be calculated on base 2).

										Amount
PR	0,416211	0,524171	0,416211	0,416211	0,110722	0,126119	0,140040	0,112558	0,112558	2,375

The Toolbar Shows	1	1	1	1	1	1	1	1	1	1
----------------------------------	---	---	---	---	---	---	---	---	---	---

The limiting estimating values for the site pages are darkened.

Having divided the estimating values to their total amount, which is unchangeable upon execution of the iteration process, we'll receive the set of probability values (the amount of all probability values is equal to one).

											Amount
P	0,154451	0,427246	0,154451	0,154451	0,019489	0,023237	0,026827	0,019924	0,019924	1,000	

Running a parallel demonstration, we'd like to use the famous classical Shannon's formula, which was normally used for calculation of the volume of the information, which an experiment contains, and which has **n** outcomes with the set probabilities.

Experiment outcome	A ₁	A ₂	...	A _n
Probabilities	p ₁	p ₂	...	p _n

$$H(p_1, p_2, \dots, p_n) = - p(A_1) \cdot \log p(A_1) - p(A_2) \cdot \log p(A_2) - \dots - p(A_n) \cdot \log p(A_n).$$

Let's calculate the values for a set of probability values, which correspond to the site pages, and we'll receive the following set of the values.

This set of values of PageRank for the pages of the site examined. (Logarithms are to be calculated on base 2).

											Amount
PR	0,416211	0,524171	0,416211	0,416211	0,110722	0,126119	0,140040	0,112558	0,112558	2,375	

The Toolbar Shows	1	1	1	1	1	1	1	1	1	1
------------------------------	---	---	---	---	---	---	---	---	---	---

PageRank is a successful representation of the directed multigraph. It is widely known that multigraphs are used to describe and build self-learning systems, which model the behavior of animals.

Let's assume that every time an incentive is engaged an animal must make a choice, e.g. to turn to left or right. So every time an animal behaves as we want it to, we can encourage it by applying operator Z_+ :

$$P_{n+1} = Z_+(P_n) = d P_n + (1 - d); 0 < d < 1. (P - \text{the estimating value})$$

The theory for this subject can be found at the classic work of Bush and Mosteller -- Stochastic Models for Learning, Wiley, New York, 1955.

We suppose that it was a variation of Shannon's formula that led to birth of PageRank.

Further Reading and Resources

For those interested in doing further PageRank research, we would recommend visiting the following resources:

PageRank Question and Answer Area:

<http://www.supportforums.org/pagerank/q&a/>

Search Engine Marketing – the essential best practice guide (an in depth view of the other aspects of how search engine work)

<http://www.supportforums.org/pagerank/sem>

PageRank: Bringing order to the web

<http://hci.stanford.edu/~page/papers/pagerank/ppframe.htm>

The anatomy of a Large-Scale HyperTextual Web Search Engine

<http://www-db.stanford.edu/~backrub/google.html>

The first PageRank calculator (with the exception of Google's!) by Mark Horrell

<http://www.markhorrell.com/seo/pagerank.asp>

Technical Proof Readers credits

The authors of this document welcome comments from experts in the fields of search engines, mathematics, and search engine optimization. We feel this feedback is important and provide the following comments and links as a thanks for the feedback of the people mentioned. If you'd like to make a comment on this document then please email: chris@supportforums.org

In the field of information retrieval on the web, PageRank has emerged as the primary (and most widely discussed) hyperlink analysis algorithm. But how it works is still largely a mystery to many in the SEO online community. PageRank Uncovered by Chris Ridings and Mike Shishigin, is an illuminating document which unveils the mystery and provides the most detailed, in-depth, yet easy to follow explanation of the power behind Google. No matter what level you're at in the field of SEO, this is not something you should read: it's something you must read.

Mike Grehan,

Author:

Search Engine Marketing: The essential best practice guide.