

PageRank

PageRank (PR) is an algorithm used by Google Search to rank web pages in their search engine results. It is named after both the term "web page" and co-founder Larry Page. PageRank is a way of measuring the importance of website pages. According to Google:

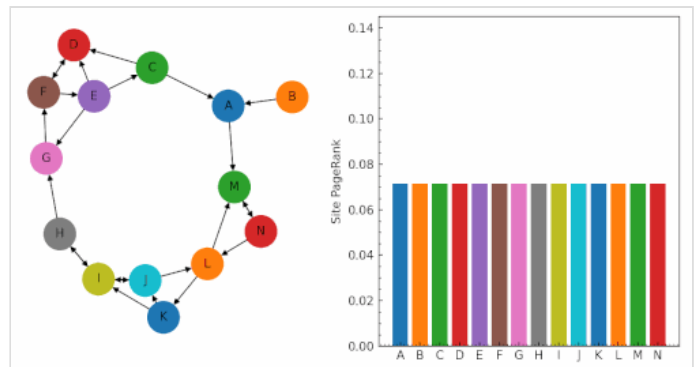
PageRank works by counting the number and quality of links to a page to determine a rough estimate of how important the website is. The underlying assumption is that more important websites are likely to receive more links from other websites.^[1]

Currently, PageRank is not the only algorithm used by Google to order search results, but it is the first algorithm that was used by the company, and it is the best known.^{[2][3]} As of September 24, 2019, all patents associated with PageRank have expired.^[4]

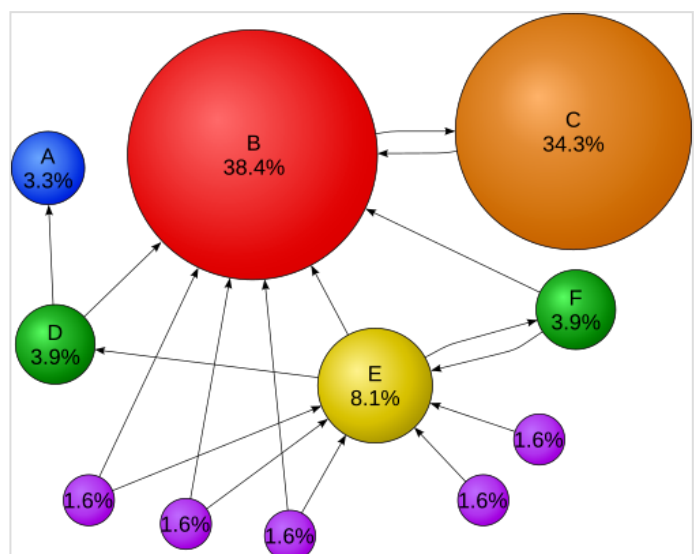
Description

PageRank is a link analysis algorithm and it assigns a numerical weighting to each element of a hyperlinked set of documents, such as the World Wide Web, with the purpose of "measuring" its relative importance within the set. The algorithm may be applied to any collection of entities with reciprocal quotations and references. The numerical weight that it assigns to any given element *E* is referred to as the *PageRank of E* and denoted by ***PR*(*E*)**.

A PageRank results from a mathematical algorithm based on the Webgraph, created by all World Wide Web pages as nodes and hyperlinks as edges, taking into consideration authority hubs such as cnn.com or mayoclinic.org. The rank value indicates an importance of a particular page. A hyperlink to a page counts



An animation of the PageRank algorithm running on a small network of pages. The size of the nodes represents the perceived importance of the page, and arrows represent hyperlinks.



A simple illustration of the PAGERANK algorithm. The percentage shows the perceived importance, and the arrows represent hyperlinks.

as a vote of support. The PageRank of a page is defined recursively and depends on the number and PageRank metric of all pages that link to it ("incoming links"). A page that is linked to by many pages with high PageRank receives a high rank itself.^[5]

Numerous academic papers concerning PageRank have been published since Page and Brin's original paper.^[6] In practice, the PageRank concept may be vulnerable to manipulation. Research has been conducted into identifying falsely influenced PageRank rankings. The goal is to find an effective means of ignoring links from documents with falsely influenced PageRank.^[7]

Other link-based ranking algorithms for Web pages include the HITS algorithm invented by Jon Kleinberg (used by Teoma and now Ask.com), the IBM CLEVER project, the TrustRank algorithm, the Hummingbird algorithm,^[8] and the SALSA algorithm.^[9]

History

The eigenvalue problem behind PageRank's algorithm was independently rediscovered and reused in many scoring problems. In 1895, Edmund Landau suggested using it for determining the winner of a chess tournament.^{[10][11]} The eigenvalue problem was also suggested in 1976 by Gabriel Pinski and Francis Narin, who worked on scientometrics ranking scientific journals,^[12] in 1977 by Thomas Saaty in his concept of Analytic Hierarchy Process which weighted alternative choices,^[13] and in 1995 by Bradley Love and Steven Sloman as a cognitive model for concepts, the centrality algorithm.^{[14][15]}

A search engine called "RankDex" from IDD Information Services, designed by Robin Li in 1996, developed a strategy for site-scoring and page-ranking.^[16] Li referred to his search mechanism as "link analysis," which involved ranking the popularity of a web site based on how many other sites had linked to it.^[17] RankDex, the first search engine with page-ranking and site-scoring algorithms, was launched in 1996.^[18] Li filed a patent for the technology in RankDex in 1997; it was granted in 1999.^[19] He later used it when he founded Baidu in China in 2000.^{[20][21]} Google founder Larry Page referenced Li's work as a citation in some of his U.S. patents for PageRank.^{[22][18][23]}

Larry Page and Sergey Brin developed PageRank at Stanford University in 1996 as part of a research project about a new kind of search engine. An interview with Héctor García-Molina, Stanford Computer Science professor and advisor to Sergey,^[24] provides background into the development of the page-rank algorithm.^[25] Sergey Brin had the idea that information on the web could be ordered in a hierarchy by "link popularity": a page ranks higher as there are more links to it.^[26] The system was developed with the help of Scott Hassan and Alan Steremberg, both of whom were cited by Page and Brin as being critical to the development of Google.^[6] Rajeev Motwani and Terry Winograd co-authored with Page and Brin the first paper about the project, describing PageRank and the initial prototype of the Google search engine, published in 1998.^[6] Shortly after, Page and Brin founded Google Inc., the company behind the Google search engine. While just one of many factors that determine the ranking of Google search results, PageRank continues to provide the basis for all of Google's web-search tools.^[27]

The name "PageRank" plays on the name of developer Larry Page, as well as of the concept of a web page.^{[28][29]} The word is a trademark of Google, and the PageRank process has been patented.^[30] However, the patent is assigned to Stanford University and not to Google. Google has exclusive license rights on the patent from Stanford University. The university received 1.8 million shares of Google in exchange for use of the patent; it sold the shares in 2005 for \$336 million.^{[31][32]}

PageRank was influenced by citation analysis, early developed by Eugene Garfield in the 1950s at the University of Pennsylvania, and by Hyper Search, developed by Massimo Marchiori at the University of Padua. In the same year PageRank was introduced (1998), Jon Kleinberg published his work on HITS. Google's founders cite Garfield, Marchiori, and Kleinberg in their original papers.^{[6][33]}

Algorithm

The PageRank algorithm outputs a probability distribution used to represent the likelihood that a person randomly clicking on links will arrive at any particular page. PageRank can be calculated for collections of documents of any size. It is assumed in several research papers that the distribution is evenly divided among all documents in the collection at the beginning of the computational process. The PageRank computations require several passes, called "iterations", through the collection to adjust approximate PageRank values to more closely reflect the theoretical true value.

A probability is expressed as a numeric value between 0 and 1. A 0.5 probability is commonly expressed as a "50% chance" of something happening. Hence, a document with a PageRank of 0.5 means there is a 50% chance that a person clicking on a random link will be directed to said document.

PageRank works on the assumption that a page is important if many other important pages link to it. This means the more quality backlinks a page has, the higher its PageRank score.

Simplified algorithm

Assume a small universe of four web pages: **A**, **B**, **C**, and **D**. Links from a page to itself are ignored. Multiple outbound links from one page to another page are treated as a single link. PageRank is initialized to the same value for all pages. In the original form of PageRank, the sum of PageRank over all pages was the total number of pages on the web at that time, so each page in this example would have an initial value of 1. However, later versions of PageRank, and the remainder of this section, assume a probability distribution between 0 and 1. Hence the initial value for each page in this example is 0.25.

The PageRank transferred from a given page to the targets of its outbound links upon the next iteration is divided equally among all outbound links.

If the only links in the system were from pages **B**, **C**, and **D** to **A**, each link would transfer 0.25 PageRank to **A** upon the next iteration, for a total of 0.75.

$$PR(A) = PR(B) + PR(C) + PR(D).$$

Suppose instead that page **B** had a link to pages **C** and **A**, page **C** had a link to page **A**, and page **D** had links to all three pages. Thus, upon the first iteration, page **B** would transfer half of its existing value (0.125) to page **A** and the other half (0.125) to page **C**. Page **C** would transfer all of its existing value (0.25) to the only page it links to, **A**. Since **D** had three outbound links, it would transfer one third of its existing value, or approximately 0.083, to **A**. At the completion of this iteration, page **A** will have a PageRank of approximately 0.458.

$$PR(A) = \frac{PR(B)}{2} + \frac{PR(C)}{1} + \frac{PR(D)}{3}.$$

In other words, the PageRank conferred by an outbound link is equal to the document's own PageRank score divided by the number of outbound links $L()$.

$$PR(A) = \frac{PR(B)}{L(B)} + \frac{PR(C)}{L(C)} + \frac{PR(D)}{L(D)}.$$

In the general case, the PageRank value for any page u can be expressed as:

$$PR(u) = \sum_{v \in B_u} \frac{PR(v)}{L(v)}$$

i.e. the PageRank value for a page u is dependent on the PageRank values for each page v contained in the set B_u (the set containing all pages linking to page u), divided by the number $L(v)$ of links from page v .

Damping factor

The PageRank theory holds that an imaginary surfer who is randomly clicking on links will eventually stop clicking. The probability, at any step, that the person will continue following links is a damping factor d . The probability that they instead jump to any random page is $1 - d$. Various studies have tested different damping factors, but it is generally assumed that the damping factor will be set around 0.85.^[6]

The damping factor is subtracted from 1 (and in some variations of the algorithm, the result is divided by the number of documents (N) in the collection) and this term is then added to the product of the damping factor and the sum of the incoming PageRank scores. That is,

$$PR(A) = \frac{1 - d}{N} + d \left(\frac{PR(B)}{L(B)} + \frac{PR(C)}{L(C)} + \frac{PR(D)}{L(D)} + \dots \right).$$

So any page's PageRank is derived in large part from the PageRanks of other pages. The damping factor adjusts the derived value downward. The original paper, however, gave the following formula, which has led to some confusion:

$$PR(A) = 1 - d + d \left(\frac{PR(B)}{L(B)} + \frac{PR(C)}{L(C)} + \frac{PR(D)}{L(D)} + \dots \right).$$

The difference between them is that the PageRank values in the first formula sum to one, while in the second formula each PageRank is multiplied by N and the sum becomes N . A statement in Page and Brin's paper that "the sum of all PageRanks is one"^[6] and claims by other Google employees^[34] support the first variant of the formula above.

Page and Brin confused the two formulas in their most popular paper "The Anatomy of a Large-Scale Hypertextual Web Search Engine", where they mistakenly claimed that the latter formula formed a probability distribution over web pages.^[6]

Google recalculates PageRank scores each time it crawls the Web and rebuilds its index. As Google increases the number of documents in its collection, the initial approximation of PageRank decreases for all documents.

The formula uses a model of a random surfer who reaches their target site after several clicks, then switches to a random page. The PageRank value of a page reflects the chance that the random surfer will land on that page by clicking on a link. It can be understood as a Markov chain in which the states are pages, and the transitions are the links between pages – all of which are all equally probable.

If a page has no links to other pages, it becomes a sink and therefore terminates the random surfing process. If the random surfer arrives at a sink page, it picks another URL at random and continues surfing again.

When calculating PageRank, pages with no outbound links are assumed to link out to all other pages in the collection. Their PageRank scores are therefore divided evenly among all other pages. In other words, to be fair with pages that are not sinks, these random transitions are added to all nodes in the Web. This residual probability, d , is usually set to 0.85, estimated from the frequency that an average surfer uses his or her browser's bookmark feature. So, the equation is as follows:

$$PR(p_i) = \frac{1 - d}{N} + d \sum_{p_j \in M(p_i)} \frac{PR(p_j)}{L(p_j)}$$

where p_1, p_2, \dots, p_N are the pages under consideration, $M(p_i)$ is the set of pages that link to p_i , $L(p_j)$ is the number of outbound links on page p_j , and N is the total number of pages.

The PageRank values are the entries of the dominant right eigenvector of the modified adjacency matrix rescaled so that each column adds up to one. This makes PageRank a particularly elegant metric: the eigenvector is

$$\mathbf{R} = \begin{bmatrix} PR(p_1) \\ PR(p_2) \\ \vdots \\ PR(p_N) \end{bmatrix}$$

where \mathbf{R} is the solution of the equation

$$\mathbf{R} = \begin{bmatrix} (1 - d)/N \\ (1 - d)/N \\ \vdots \\ (1 - d)/N \end{bmatrix} + d \begin{bmatrix} \ell(p_1, p_1) & \ell(p_1, p_2) & \cdots & \ell(p_1, p_N) \\ \ell(p_2, p_1) & \ddots & & \vdots \\ \vdots & & \ell(p_i, p_j) & \\ \ell(p_N, p_1) & \cdots & & \ell(p_N, p_N) \end{bmatrix} \mathbf{R}$$

where the adjacency function $\ell(p_i, p_j)$ is the ratio between number of links outbound from page j to page i to the total number of outbound links of page j . The adjacency function is 0 if page p_j does not link to p_i , and normalized such that, for each j

$$\sum_{i=1}^N \ell(p_i, p_j) = 1$$

i.e. the elements of each column sum up to 1, so the matrix is a stochastic matrix (for more details see the computation section below). Thus this is a variant of the eigenvector centrality measure used commonly in network analysis.

Because of the large eigengap of the modified adjacency matrix above,^[35] the values of the PageRank eigenvector can be approximated to within a high degree of accuracy within only a few iterations.

Google's founders, in their original paper,^[33] reported that the PageRank algorithm for a network consisting of 322 million links (in-edges and out-edges) converges to within a tolerable limit in 52 iterations. The convergence in a network of half the above size took approximately 45 iterations. Through this data, they concluded the algorithm can be scaled very well and that the scaling factor for extremely large networks would be roughly linear in $\log n$, where n is the size of the network.

As a result of Markov theory, it can be shown that the PageRank of a page is the probability of arriving at that page after a large number of clicks. This happens to equal t^{-1} where t is the expectation of the number of clicks (or random jumps) required to get from the page back to itself.

One main disadvantage of PageRank is that it favors older pages. A new page, even a very good one, will not have many links unless it is part of an existing site (a site being a densely connected set of pages, such as Wikipedia).

Several strategies have been proposed to accelerate the computation of PageRank.^[36]

Various strategies to manipulate PageRank have been employed in concerted efforts to improve search results rankings and monetize advertising links. These strategies have severely impacted the reliability of the PageRank concept, which purports to determine which documents are actually highly valued by the Web community.

Since December 2007, when it started *actively* penalizing sites selling paid text links, Google has combatted link farms and other schemes designed to artificially inflate PageRank. How Google identifies link farms and other PageRank manipulation tools is among Google's trade secrets.

Computation

PageRank can be computed either iteratively or algebraically. The iterative method can be viewed as the power iteration method ^{[37][38]} or the power method. The basic mathematical operations performed are identical.

Iterative

At $t = 0$, an initial probability distribution is assumed, usually

$$PR(p_i; 0) = \frac{1}{N}$$

where N is the total number of pages, and $\mathbf{p}_i; \mathbf{0}$ is page i at time 0 .

At each time step, the computation, as detailed above, yields

$$PR(\mathbf{p}_i; t+1) = \frac{1-d}{N} + d \sum_{p_j \in M(\mathbf{p}_i)} \frac{PR(\mathbf{p}_j; t)}{L(\mathbf{p}_j)}$$

where d is the damping factor,

or in matrix notation

$$\mathbf{R}(t+1) = d\mathcal{M}\mathbf{R}(t) + \frac{1-d}{N}\mathbf{1}, \quad (1)$$

where $\mathbf{R}_i(t) = PR(\mathbf{p}_i; t)$ and $\mathbf{1}$ is the column vector of length N containing only ones.

The matrix \mathcal{M} is defined as

$$\mathcal{M}_{ij} = \begin{cases} 1/L(\mathbf{p}_j), & \text{if } j \text{ links to } i \\ 0, & \text{otherwise} \end{cases}$$

i.e.,

$$\mathcal{M} := (\mathbf{K}^{-1}\mathbf{A})^T$$

where \mathbf{A} denotes the adjacency matrix of the graph and \mathbf{K} is the diagonal matrix with the outdegrees in the diagonal.

The probability calculation is made for each page at a time point, then repeated for the next time point. The computation ends when for some small ϵ

$$|\mathbf{R}(t+1) - \mathbf{R}(t)| < \epsilon$$

i.e., when convergence is assumed.

Power method

If the matrix \mathcal{M} is a transition probability, i.e., column-stochastic and \mathbf{R} is a probability distribution (i.e., $|\mathbf{R}| = 1$, $\mathbf{E}\mathbf{R} = \mathbf{1}$ where \mathbf{E} is matrix of all ones), then equation (2) is equivalent to

$$\mathbf{R} = \left(d\mathcal{M} + \frac{1-d}{N}\mathbf{E} \right) \mathbf{R} =: \widehat{\mathcal{M}}\mathbf{R}. \quad (3)$$

Hence PageRank \mathbf{R} is the principal eigenvector of $\widehat{\mathcal{M}}$. A fast and easy way to compute this is using the power method: starting with an arbitrary vector $\mathbf{x}(0)$, the operator $\widehat{\mathcal{M}}$ is applied in succession, i.e.,

$$\mathbf{x}(t+1) = \widehat{\mathcal{M}}\mathbf{x}(t)$$

until

$$|x(t + 1) - x(t)| < \epsilon$$

.

Note that in equation (3) the matrix on the right-hand side in the parenthesis can be interpreted as

$$\frac{1 - d}{N} \mathbf{E} = (1 - d) \mathbf{P} \mathbf{1}^t$$

,

where \mathbf{P} is an initial probability distribution. n the current case

$$\mathbf{P} := \frac{1}{N} \mathbf{1}$$

.

Finally, if \mathcal{M} has columns with only zero values, they should be replaced with the initial probability vector \mathbf{P} . In other words,

$$\mathcal{M}' := \mathcal{M} + \mathcal{D}$$

,

where the matrix \mathcal{D} is defined as

$$\mathcal{D} := \mathbf{P} \mathbf{D}^t$$

,

with

$$\mathbf{D}_i = \begin{cases} 1, & \text{if } L(p_i) = 0 \\ 0, & \text{otherwise} \end{cases}$$

In this case, the above two computations using \mathcal{M} only give the same PageRank if their results are normalized:

$$\mathbf{R}_{\text{power}} = \frac{\mathbf{R}_{\text{iterative}}}{|\mathbf{R}_{\text{iterative}}|} = \frac{\mathbf{R}_{\text{algebraic}}}{|\mathbf{R}_{\text{algebraic}}|}$$

.

Implementation

Python

```
import numpy as np

def pagerank(M, d: float = 0.85):
    """PageRank algorithm with explicit number of iterations. Returns ranking of nodes (pages) in the
    adjacency matrix.

    Parameters
    -----
    M : numpy array
```


adjacency matrix where $M_{i,j}$ represents the link from 'j' to 'i', such that for all 'j'
 $\sum(i, M_{i,j}) = 1$
d : float, optional
damping factor, by default 0.85

Returns

numpy array

a vector of ranks such that v_i is the i-th rank from [0, 1],

"""

```
N = M.shape[1]
w = np.ones(N) / N
M_hat = d * M
v = M_hat @ w + (1 - d) / N
while np.linalg.norm(w - v) >= 1e-10:
    w = v
    v = M_hat @ w + (1 - d) / N
return v
```

```
M = np.array([[0, 0, 0, .25],
              [0, 0, 0, .5],
              [1, 0.5, 0, .25],
              [0, 0.5, 1, 0]])
v = pagerank(M, 0.85)
```

Variations

PageRank of an undirected graph

The PageRank of an undirected graph G is statistically close to the degree distribution of the graph G ,^[39] but they are generally not identical: If R is the PageRank vector defined above, and D is the degree distribution vector

$$D = \frac{1}{2|E|} \begin{bmatrix} \deg(p_1) \\ \deg(p_2) \\ \vdots \\ \deg(p_N) \end{bmatrix}$$

where $\deg(p_i)$ denotes the degree of vertex p_i , and E is the edge-set of the graph, then, with $Y = \frac{1}{N} \mathbf{1}$,^[40] shows that:

$$\frac{1-d}{1+d} \|Y - D\|_1 \leq \|R - D\|_1 \leq \|Y - D\|_1,$$

that is, the PageRank of an undirected graph equals to the degree distribution vector if and only if the graph is regular, i.e., every vertex has the same degree.

Ranking objects of two kinds

A generalization of PageRank for the case of ranking two interacting groups of objects was described by Daugulis.^[41] In applications it may be necessary to model systems having objects of two kinds where a weighted relation is defined on object pairs. This leads to considering bipartite graphs. For such graphs

two related positive or nonnegative irreducible matrices corresponding to vertex partition sets can be defined. One can compute rankings of objects in both groups as eigenvectors corresponding to the maximal positive eigenvalues of these matrices. Normed eigenvectors exist and are unique by the Perron or Perron–Frobenius theorem. Example: consumers and products. The relation weight is the product consumption rate.

Distributed algorithm for PageRank computation

Sarma et al. describe two random walk-based distributed algorithms for computing PageRank of nodes in a network.^[42] One algorithm takes $O(\log n/\epsilon)$ rounds with high probability on any graph (directed or undirected), where n is the network size and ϵ is the reset probability ($1 - \epsilon$, which is called the damping factor) used in the PageRank computation. They also present a faster algorithm that takes $O(\sqrt{\log n}/\epsilon)$ rounds in undirected graphs. In both algorithms, each node processes and sends a number of bits per round that are polylogarithmic in n , the network size.

Google Toolbar

The Google Toolbar long had a PageRank feature which displayed a visited page's PageRank as a whole number between 0 (least popular) and 10 (most popular). Google had not disclosed the specific method for determining a Toolbar PageRank value, which was to be considered only a rough indication of the value of a website. The "Toolbar Pagerank" was available for verified site maintainers through the Google Webmaster Tools interface. However, on October 15, 2009, a Google employee confirmed that the company had removed PageRank from its *Webmaster Tools* section, saying that "We've been telling people for a long time that they shouldn't focus on PageRank so much. Many site owners seem to think it's the most important metric for them to track, which is simply not true."^[43]

The "Toolbar Pagerank" was updated very infrequently. It was last updated in November 2013. In October 2014 Matt Cutts announced that another visible pagerank update would not be coming.^[44] In March 2016 Google announced it would no longer support this feature, and the underlying API would soon cease to operate.^[45] On April 15, 2016, Google turned off display of PageRank Data in Google Toolbar,^[46] though the PageRank continued to be used internally to rank content in search results.^[47]

SERP rank

The search engine results page (SERP) is the actual result returned by a search engine in response to a keyword query. The SERP consists of a list of links to web pages with associated text snippets, paid ads, featured snippets, and Q&A. The SERP rank of a web page refers to the placement of the corresponding link on the SERP, where higher placement means higher SERP rank. The SERP rank of a web page is a function not only of its PageRank, but of a relatively large and continuously adjusted set of factors (over 200).^[48] Search engine optimization (SEO) is aimed at influencing the SERP rank for a website or a set of web pages.

Positioning of a webpage on Google SERPs for a keyword depends on relevance and reputation, also known as authority and popularity. PageRank is Google's indication of its assessment of the reputation of a webpage: It is non-keyword specific. Google uses a combination of webpage and website authority to determine the overall authority of a webpage competing for a keyword.^[49] The PageRank of the HomePage of a website is the best indication Google offers for website authority.^[50]

After the introduction of Google Places into the mainstream organic SERP, numerous other factors in addition to PageRank affect ranking a business in Local Business Results.^[51] When Google elaborated on the reasons for PageRank deprecation at Q&A #March 2016, they announced Links and Content as the Top Ranking Factors. RankBrain had earlier in October 2015 been announced as the #3 Ranking Factor, so the Top 3 Factors have been confirmed officially by Google.^[52]

Google directory PageRank

The Google Directory PageRank was an 8-unit measurement. Unlike the Google Toolbar, which showed a numeric PageRank value upon mouseover of the green bar, the Google Directory only displayed the bar, never the numeric values. Google Directory was closed on July 20, 2011.^[53]

False or spoofed PageRank

It was known that the PageRank shown in the Toolbar could easily be spoofed. Redirection from one page to another, either via a HTTP 302 response or a "Refresh" meta tag, caused the source page to acquire the PageRank of the destination page. Hence, a new page with PR 0 and no incoming links could have acquired PR 10 by redirecting to the Google home page. Spoofing can usually be detected by performing a Google search for a source URL; if the URL of an entirely different site is displayed in the results, the latter URL may represent the destination of a redirection.

Manipulating PageRank

For search engine optimization purposes, some companies offer to sell high PageRank links to webmasters.^[54] As links from higher-PR pages are believed to be more valuable, they tend to be more expensive. It can be an effective and viable marketing strategy to buy link advertisements on content pages of quality and relevant sites to drive traffic and increase a webmaster's link popularity. However, Google has publicly warned webmasters that if they are or were discovered to be selling links for the purpose of conferring PageRank and reputation, their links will be devalued (ignored in the calculation of other pages' PageRanks). The practice of buying and selling ^[55] is intensely debated across the Webmaster community. Google advised webmasters to use the nofollow HTML attribute value on paid links. According to Matt Cutts, Google is concerned about webmasters who try to game the system, and thereby reduce the quality and relevance of Google search results.^[54]

In 2019, Google announced two additional link attributes providing hints about which links to consider or exclude within Search: `rel="ugc"` as a tag for user-generated content, such as comments; and `rel="sponsored"` as a tag for advertisements or other types of sponsored content. Multiple `rel` values are also allowed, for example, `rel="ugc sponsored"` can be used to hint that the link came from user-generated content and is sponsored.^[56]

Even though PageRank has become less important for SEO purposes, the existence of back-links from more popular websites continues to push a webpage higher up in search rankings.^[57]

Directed Surfer Model

A more intelligent surfer that probabilistically hops from page to page depending on the content of the pages and query terms the surfer is looking for. This model is based on a query-dependent PageRank score of a page which as the name suggests is also a function of query. When given a multiple-term query, $Q = \{q_1, q_2, \dots\}$, the surfer selects a q according to some probability distribution, $P(q)$, and uses that term to guide its behavior for a large number of steps. It then selects another term according to the distribution to determine its behavior, and so on. The resulting distribution over visited web pages is QD-PageRank.^[58]

Other uses

The mathematics of PageRank are entirely general and apply to any graph or network in any domain. Thus, PageRank is now regularly used in bibliometrics, social and information network analysis, and for link prediction and recommendation. It is used for systems analysis of road networks, and in biology, chemistry, neuroscience, and physics.^[59]

Scientific research and academia

PageRank has been used to quantify the scientific impact of researchers. The underlying citation and collaboration networks are used in conjunction with PageRank algorithm in order to come up with a ranking system for individual publications which propagates to individual authors. The new index known as pagerank-index (Pi) is demonstrated to be fairer compared to h-index in the context of many drawbacks exhibited by h-index.^[60]

For the analysis of protein networks in biology PageRank is also a useful tool.^{[61][62]}

In any ecosystem, a modified version of PageRank may be used to determine species that are essential to the continuing health of the environment.^[63]

A similar newer use of PageRank is to rank academic doctoral programs based on their records of placing their graduates in faculty positions. In PageRank terms, academic departments link to each other by hiring their faculty from each other (and from themselves).^[64]

A version of PageRank has recently been proposed as a replacement for the traditional Institute for Scientific Information (ISI) impact factor,^[65] and implemented at Eigenfactor as well as at SCImago. Instead of merely counting total citations to a journal, the "importance" of each citation is determined in a PageRank fashion.

In neuroscience, the PageRank of a neuron in a neural network has been found to correlate with its relative firing rate.^[66]

Internet use

Personalized PageRank is used by Twitter to present users with other accounts they may wish to follow.^[67]

Swifttype's site search product builds a "PageRank that's specific to individual websites" by looking at each website's signals of importance and prioritizing content based on factors such as number of links from the home page.^[68]

A Web crawler may use PageRank as one of a number of importance metrics it uses to determine which URL to visit during a crawl of the web. One of the early working papers^[69] that were used in the creation of Google is *Efficient crawling through URL ordering*,^[70] which discusses the use of a number of different importance metrics to determine how deeply, and how much of a site Google will crawl. PageRank is presented as one of a number of these importance metrics, though there are others listed such as the number of inbound and outbound links for a URL, and the distance from the root directory on a site to the URL.

The PageRank may also be used as a methodology to measure the apparent impact of a community like the Blogosphere on the overall Web itself. This approach uses therefore the PageRank to measure the distribution of attention in reflection of the Scale-free network paradigm.

Other applications

In 2005, in a pilot study in Pakistan, *Structural Deep Democracy*, SD2^{[71][72]} was used for leadership selection in a sustainable agriculture group called Contact Youth. SD2 uses *PageRank* for the processing of the transitive proxy votes, with the additional constraints of mandating at least two initial proxies per voter, and all voters are proxy candidates. More complex variants can be built on top of SD2, such as adding specialist proxies and direct votes for specific issues, but SD2 as the underlying umbrella system, mandates that generalist proxies should always be used.

In sport the PageRank algorithm has been used to rank the performance of: teams in the National Football League (NFL) in the USA;^[73] individual soccer players;^[74] and athletes in the Diamond League.^[75]

PageRank has been used to rank spaces or streets to predict how many people (pedestrians or vehicles) come to the individual spaces or streets.^{[76][77]} In lexical semantics it has been used to perform Word Sense Disambiguation,^[78] Semantic similarity,^[79] and also to automatically rank WordNet synsets according to how strongly they possess a given semantic property, such as positivity or negativity.^[80]

How a traffic system changes its operational mode can be described by transitions between quasi-stationary states in correlation structures of traffic flow. PageRank has been used to identify and explore the dominant states among these quasi-stationary states in traffic systems.^[81]

nofollow

In early 2005, Google implemented a new value, "nofollow",^[82] for the rel attribute of HTML link and anchor elements, so that website developers and bloggers can make links that Google will not consider for the purposes of PageRank—they are links that no longer constitute a "vote" in the PageRank system. The nofollow relationship was added in an attempt to help combat spamdexing.

As an example, people could previously create many message-board posts with links to their website to artificially inflate their PageRank. With the nofollow value, message-board administrators can modify their code to automatically insert "rel='nofollow'" to all hyperlinks in posts, thus preventing PageRank

from being affected by those particular posts. This method of avoidance, however, also has various drawbacks, such as reducing the link value of legitimate comments. (See: [Spam in blogs#nofollow](#))

In an effort to manually control the flow of PageRank among pages within a website, many webmasters practice what is known as PageRank Sculpting^[83]—which is the act of strategically placing the nofollow attribute on certain internal links of a website in order to funnel PageRank towards those pages the webmaster deemed most important. This tactic had been used since the inception of the nofollow attribute, but may no longer be effective since Google announced that blocking PageRank transfer with nofollow does not redirect that PageRank to other links.^[84]

See also

- [Attention inequality](#)
- [CheiRank](#)
- [Domain authority](#)
- [EigenTrust](#) — a decentralized PageRank algorithm
- [Google bombing](#)
- [Google Hummingbird](#)
- [Google matrix](#)
- [Google Panda](#)
- [Google Penguin](#)
- [Google Search](#)
- [Hilltop algorithm](#)
- [Katz centrality](#) — a 1953 scheme closely related to pagerank
- [Link building](#)
- [Search engine optimization](#)
- [SimRank](#) — a measure of object-to-object similarity based on random-surfer model
- [TextRank](#) — Application of PageRank to Natural Language Processing
- [TrustRank](#)
- [VisualRank](#) — Google's application of PageRank to image-search
- [Webgraph](#)

References

Citations

1. "Facts about Google and Competition" (<https://web.archive.org/web/20111104131332/http://www.google.com/competition/howgooglesearchworks.html>). Archived from the original (<https://www.google.com/competition/howgooglesearchworks.html>) on 4 November 2011. Retrieved 12 July 2014.
2. Sullivan, Danny (2007-04-26). "What Is Google PageRank? A Guide For Searchers & Webmasters" (<http://searchengineland.com/what-is-google-pagerank-a-guide-for-searchers-webmasters-11068>). *Search Engine Land*. Archived (<https://web.archive.org/web/20160703031514/http://searchengineland.com/what-is-google-pagerank-a-guide-for-searchers-webmasters-11068>) from the original on 2016-07-03.

3. Cutts, Matt. "Algorithms Rank Relevant Results Higher" (<https://web.archive.org/web/20130702063520/https://www.google.com/competition/howgooglesearchworks.html>). Archived from the original (<https://www.google.com/competition/howgooglesearchworks.html>) on July 2, 2013. Retrieved 19 October 2015.
4. "US7058628B1 - Method for node ranking in a linked database - Google Patents" (<https://patents.google.com/patent/US7058628B1/en>). *Google Patents*. Archived (<https://web.archive.org/web/20200116103001/https://patents.google.com/patent/US7058628B1/en>) from the original on January 16, 2020. Retrieved September 14, 2019.
5. Avrachenkov, K., & Litvak, N. (2006). The Effect of New Links on Google Pagerank (<https://www.tandfonline.com/doi/citedby/10.1080/15326340600649052?scroll=top&needAccess=true>). *Stochastic Models*, 22(2), 319–331.
6. Brin, S.; Page, L. (1998). "The anatomy of a large-scale hypertextual Web search engine" (<http://infolab.stanford.edu/pub/papers/google.pdf>) (PDF). *Computer Networks and ISDN Systems*. **30** (1–7): 107–117. CiteSeerX 10.1.1.115.5930 (<https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.115.5930>). doi:10.1016/S0169-7552(98)00110-X (<https://doi.org/10.1016%2FS0169-7552%2898%2900110-X>). ISSN 0169-7552 (<https://search.worldcat.org/issn/0169-7552>). S2CID 7587743 (<https://api.semanticscholar.org/CorpusID:7587743>). Archived (<https://web.archive.org/web/20150927004511/http://infolab.stanford.edu/pub/papers/google.pdf>) (PDF) from the original on 2015-09-27.
7. Gyöngyi, Zoltán; Berkhin, Pavel; Garcia-Molina, Hector; Pedersen, Jan (2006), "Link spam detection based on mass estimation", *Proceedings of the 32nd International Conference on Very Large Data Bases (VLDB '06, Seoul, Korea)* (<http://ilpubs.stanford.edu:8090/697/1/2005-33.pdf>) (PDF), pp. 439–450, archived (<https://web.archive.org/web/20141203194914/http://ilpubs.stanford.edu:8090/697/1/2005-33.pdf>) (PDF) from the original on 2014-12-03.
8. "FAQ: All About The New Google "Hummingbird" Algorithm" (<https://searchengineland.com/google-hummingbird-172816>). *Search Engine Land*. 26 September 2013. Archived (<https://web.archive.org/web/20181223110045/https://searchengineland.com/google-hummingbird-172816>) from the original on 23 December 2018. Retrieved 18 December 2018.
9. Wang, Ziyang. "Improved Link-Based Algorithms for Ranking Web Pages" (<https://cs.nyu.edu/media/publications/TR2003-846.pdf>) (PDF). *cs.nyu.edu*. New York University, Department of Computer Science. Retrieved 7 August 2023.
10. Landau, Edmund (1895). "Zur relativen Wertbemessung der Turnierresultate". *Deutsches Wochensach*. **11** (42): 51–54.
11. Sinn, Rainer; Ziegler, Günter M. (2022-10-31). "Landau on Chess Tournaments and Google's PageRank". arXiv:2210.17300 (<https://arxiv.org/abs/2210.17300>) [math.HO (<https://arxiv.org/archive/math/HO>)].
12. Gabriel Pinski & Francis Narin (1976). "Citation influence for journal aggregates of scientific publications: Theory, with application to the literature of physics". *Information Processing & Management*. **12** (5): 297–312. doi:10.1016/0306-4573(76)90048-0 (<https://doi.org/10.1016%2F0306-4573%2876%2990048-0>).
13. Thomas Saaty (1977). "A scaling method for priorities in hierarchical structures". *Journal of Mathematical Psychology*. **15** (3): 234–281. doi:10.1016/0022-2496(77)90033-5 (<https://doi.org/10.1016%2F0022-2496%2877%2990033-5>). hdl:10338.dmlcz/101787 (<https://hdl.handle.net/10338.dmlcz%2F101787>).
14. Bradley C. Love & Steven A. Sloman. "Mutability and the determinants of conceptual transformability" (http://bradlove.org/papers/love_sloman_1995.pdf) (PDF). *Proceedings of the Seventeenth Annual Conference of the Cognitive Science Society*. pp. 654–659. Archived (https://web.archive.org/web/20171223215954/http://bradlove.org/papers/love_sloman_1995.pdf) (PDF) from the original on 2017-12-23. Retrieved 2017-12-23.

15. "How a CogSci undergrad invented PageRank three years before Google" (<http://bradlove.org/blog/cogsci-page-rank>). bradlove.org. Archived (<https://web.archive.org/web/20171211160207/http://bradlove.org/blog/cogsci-page-rank>) from the original on 2017-12-11. Retrieved 2017-12-23.
16. Li, Yanhong (August 6, 2002). "Toward a qualitative search engine". *IEEE Internet Computing*. **2** (4): 24–29. doi:10.1109/4236.707687 (<https://doi.org/10.1109%2F4236.707687>).
17. "The Rise of Baidu (That's Chinese for Google)" (<https://www.nytimes.com/2006/09/17/business/yourmoney/17baidu.html>). *The New York Times*. 17 September 2006. Archived (<https://web.archive.org/web/20190627071550/https://www.nytimes.com/2006/09/17/business/yourmoney/17baidu.html>) from the original on 27 June 2019. Retrieved 16 June 2019.
18. "About: RankDex" (<http://www.rankdex.com/about.html>) Archived (<https://web.archive.org/web/20150525015816/http://www.rankdex.com/about.html>) 2015-05-25 at the Wayback Machine, RankDex; accessed 3 May 2014.
19. USPTO, "Hypertext Document Retrieval System and Method" (<https://www.google.com/patents?hl=en&lr=&vid=USPAT5920859&id=x04ZAAAAEBAJ&oi=fnd&dq=yanhong+li&printsec=abstract#v=onepage&q=yanhong%20li&f=false>) Archived (<https://web.archive.org/web/20111205225726/http://www.google.com/patents?hl=en&lr=&vid=USPAT5920859&id=x04ZAAAAEBAJ&oi=fnd&dq=yanhong+li&printsec=abstract#v=onepage&q=yanhong%20li&f=false>) 2011-12-05 at the Wayback Machine, U.S. Patent number: 5920859, Inventor: Yanhong Li, Filing date: Feb 5, 1997, Issue date: Jul 6, 1999
20. Greenberg, Andy, "The Man Who's Beating Google" (https://www.forbes.com/forbes/2009/10/05/technology-baidu-robin-li-man-whos-beating-google_2.html) Archived (https://web.archive.org/web/20130308094445/http://www.forbes.com/forbes/2009/10/05/technology-baidu-robin-li-man-whos-beating-google_2.html) 2013-03-08 at the Wayback Machine, *Forbes* magazine, October 05, 2009
21. "About: RankDex" (<http://www.rankdex.com/about.html>) Archived (<https://web.archive.org/web/20120120002301/http://www.rankdex.com/about.html>) 2012-01-20 at the Wayback Machine, *rankdex.com*
22. "Method for node ranking in a linked database" (<https://patents.google.com/patent/US6285999>). Google Patents. Archived (<https://web.archive.org/web/20151015185034/http://www.google.com/patents/US6285999>) from the original on 15 October 2015. Retrieved 19 October 2015.
23. Altucher, James (March 18, 2011). "10 Unusual Things About Google" (<https://www.forbes.com/sites/jamesaltucher/2011/03/18/10-unusual-things-about-google-also-the-worst-vc-decision-i-ever-made/>). *Forbes*. Archived (<https://web.archive.org/web/20190616133656/https://www.forbes.com/sites/jamesaltucher/2011/03/18/10-unusual-things-about-google-also-the-worst-vc-decision-i-ever-made/>) from the original on 16 June 2019. Retrieved 16 June 2019.
24. Greg Wientjes. "Hector Garcia-Molina: Stanford Computer Science Professor and Advisor to Sergey" (https://www.podomatic.com/podcasts/gwientjes/episodes/2015-09-05T06_05_37-07_00). pp. minutes 25.45-32.50, 34.00 – 38.20. Retrieved 2019-12-06.
25. Page, Larry, "PageRank: Bringing Order to the Web" (<http://ilpubs.stanford.edu:8090/422/1/1999-66.pdf>) (PDF). Archived (<https://web.archive.org/web/20090126204112/http://ilpubs.stanford.edu:8090/422/1/1999-66.pdf>) (PDF) from the original on January 26, 2009. Retrieved 2022-10-06., Stanford Digital Library Project, talk. August 18, 1997 (archived 2002)
26. 187-page study from Graz University, Austria (http://www.iicm.tugraz.at/Ressourcen/Papers/dangers_google.pdf) Archived (https://web.archive.org/web/20140116234734/http://www.iicm.tugraz.at/Ressourcen/Papers/dangers_google.pdf) 2014-01-16 at the Wayback Machine, includes the note that also human brains are used when determining the page rank in Google.

27. "Our products and services" (<https://www.google.com/technology/>). Archived (<https://web.archive.org/web/20080623233116/http://www.google.com/technology/>) from the original on 2008-06-23. Retrieved 2011-05-27.
28. David Vise & Mark Malseed (2005). *The Google Story* (<https://archive.org/details/googlestory00vise/page/37>). Delacorte Press. p. 37 (<https://archive.org/details/googlestory00vise/page/37>). ISBN 978-0-553-80457-7.
29. "Google Press Center: Fun Facts" (<https://web.archive.org/web/20010715123343/https://www.google.com/press/funfacts.html>). Archived from the original (<https://www.google.com/press/funfacts.html>) on 2001-07-15.
30. U.S. patent 6,285,999 (<https://patents.google.com/patent/US6285999>)
31. Lisa M. Krieger (1 December 2005). "Stanford Earns \$336 Million Off Google Stock" (http://www.redorbit.com/news/education/318480/stanford_earns_336_million_off_google_stock/). *San Jose Mercury News*. Archived (https://web.archive.org/web/20090408084918/http://www.redorbit.com/news/education/318480/stanford_earns_336_million_off_google_stock/) from the original on 8 April 2009. Retrieved 2009-02-25 – via cited by redOrbit.
32. Richard Brandt. "Starting Up. How Google got its groove" (<http://www.stanfordalumni.org/news/magazine/2004/novdec/features/startup.html>). Stanford magazine. Archived (<https://web.archive.org/web/20090310190153/http://www.stanfordalumni.org/news/magazine/2004/novdec/features/startup.html>) from the original on 2009-03-10. Retrieved 2009-02-25.
33. Page, Lawrence; Brin, Sergey; Motwani, Rajeev; Winograd, Terry (1999). The PageRank citation ranking: Bringing order to the Web (<http://dbpubs.stanford.edu:8090/pub/showDoc.Fulltext?lang=en&doc=1999-66&format=pdf&compression=>) (Report). Archived (<https://web.archive.org/web/20060427013310/http://dbpubs.stanford.edu:8090/pub/showDoc.Fulltext?lang=en&doc=1999-66&format=pdf&compression=>) from the original on 2006-04-27., published as a technical report on January 29, 1998 PDF (<http://ilpubs.stanford.edu:8090/422/1/1999-66.pdf>) Archived (<https://web.archive.org/web/20110818093436/http://ilpubs.stanford.edu:8090/422/1/1999-66.pdf>) 2011-08-18 at the Wayback Machine
34. Matt Cutts's blog: Straight from Google: What You Need to Know (<http://www.mattcutts.com/blog/seo-for-bloggers/>) Archived (<https://web.archive.org/web/20100207231048/http://www.mattcutts.com/blog/seo-for-bloggers/>) 2010-02-07 at the Wayback Machine, see page 15 of his slides.
35. Taher Haveliwala & Sepandar Kamvar (March 2003). "The Second Eigenvalue of the Google Matrix" (<http://www-cs-students.stanford.edu/~taherh/papers/secondeigenvalue.pdf>) (PDF). *Stanford University Technical Report: 7056*. arXiv:math/0307056 (<https://arxiv.org/abs/math/0307056>). Bibcode:2003math.....7056N (<https://ui.adsabs.harvard.edu/abs/2003math.....7056N>). Archived (<https://web.archive.org/web/20081217094957/http://www-cs-students.stanford.edu/~taherh/papers/secondeigenvalue.pdf>) (PDF) from the original on 2008-12-17.
36. Gianna M. Del Corso; Antonio Gullí; Francesco Romani (2004). "Fast PageRank Computation Via a Sparse Linear System (Extended Abstract)". In Stefano Leonardi (ed.). *Algorithms and Models for the Web-Graph: Third International Workshop, WAW 2004, Rome, Italy, October 16, 2004. Proceedings*. pp. 118–130. CiteSeerX 10.1.1.58.9060 (<https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.58.9060>). doi:10.1007/978-3-540-30216-2_10 (https://doi.org/10.1007%2F978-3-540-30216-2_10). ISBN 978-3-540-23427-2.
37. Arasu, A.; Novak, J.; Tomkins, A.; Tomlin, J. (2002). "PageRank computation and the structure of the web: Experiments and algorithms". *Proceedings of the Eleventh International World Wide Web Conference, Poster Track*. Brisbane, Australia. pp. 107–117. CiteSeerX 10.1.1.18.5264 (<https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.18.5264>).
38. Massimo Franceschet (2010). "PageRank: Standing on the shoulders of giants". arXiv:1002.2858 (<https://arxiv.org/abs/1002.2858>) [cs.IR (<https://arxiv.org/archive/cs/IR>)].

39. Nicola Perra and Santo Fortunato; Fortunato (September 2008). "Spectral centrality measures in complex networks". *Phys. Rev. E*. **78** (3) 36107. arXiv:0805.3322 (<https://arxiv.org/abs/0805.3322>). Bibcode:2008PhRvE..78c6107P (<https://ui.adsabs.harvard.edu/abs/2008PhRvE..78c6107P>). doi:10.1103/PhysRevE.78.036107 (<https://doi.org/10.1103%2FPhysRevE.78.036107>). PMID 18851105 (<https://pubmed.ncbi.nlm.nih.gov/18851105>). S2CID 1755112 (<https://api.semanticscholar.org/CorpusID:1755112>).
40. Vince Grolmusz (2015). "A Note on the PageRank of Undirected Graphs". *Information Processing Letters*. **115** (6–8): 633–634. arXiv:1205.1960 (<https://arxiv.org/abs/1205.1960>). doi:10.1016/j.ipl.2015.02.015 (<https://doi.org/10.1016%2Fj.ipl.2015.02.015>). S2CID 9855132 (<https://api.semanticscholar.org/CorpusID:9855132>).
41. Peteris Daugulis; Daugulis (2012). "A note on a generalization of eigenvector centrality for bipartite graphs and applications". *Networks*. **59** (2): 261–264. arXiv:1610.01544 (<https://arxiv.org/abs/1610.01544>). doi:10.1002/net.20442 (<https://doi.org/10.1002%2Fnet.20442>). S2CID 1436859 (<https://api.semanticscholar.org/CorpusID:1436859>).
42. Atish Das Sarma; Anisur Rahaman Molla; Gopal Pandurangan; Eli Upfal (2015). "Fast Distributed PageRank Computation". *Theoretical Computer Science*. **561**: 113–121. arXiv:1208.3071 (<https://arxiv.org/abs/1208.3071>). doi:10.1016/j.tcs.2014.04.003 (<https://doi.org/10.1016%2Fj.tcs.2014.04.003>). S2CID 10284718 (<https://api.semanticscholar.org/CorpusID:10284718>).
43. Susan Moskwa. "PageRank Distribution Removed From WMT" (<https://www.google.com/support/forum/p/Webmasters/thread?tid=6a1d6250e26e9e48&hl=en>). Archived (<https://web.archive.org/web/20091017081100/http://www.google.com/support/forum/p/Webmasters/thread?tid=6a1d6250e26e9e48&hl=en>) from the original on October 17, 2009. Retrieved October 16, 2009.
44. Bartleman, Wil (2014-10-12). "Google Page Rank Update is Not Coming" (<https://managedadmin.com/phoenix-seo-services/google-page-rank-update-coming>). Managed Admin. Archived (<https://web.archive.org/web/20150402221831/https://managedadmin.com/phoenix-seo-services/google-page-rank-update-coming/>) from the original on 2015-04-02. Retrieved 2014-10-12.
45. Schwartz, Barry (March 8, 2016). "Google has confirmed it is removing Toolbar PageRank" (<http://searchengineland.com/google-has-confirmed-they-are-removing-toolbar-pagerank-244230>). *Search Engine Land*. Archived (<https://web.archive.org/web/20160310093405/http://searchengineland.com/google-has-confirmed-they-are-removing-toolbar-pagerank-244230>) from the original on March 10, 2016.
46. Schwartz, Barry (18 April 2016). "Google Toolbar PageRank officially goes dark" (<http://searchengineland.com/google-toolbar-pagerank-officially-goes-dark-247553>). *Search Engine Land*. Archived (<https://web.archive.org/web/20160421224919/http://searchengineland.com/google-toolbar-pagerank-officially-goes-dark-247553>) from the original on 2016-04-21.
47. Southern, Matt (2016-04-19). "Google PageRank Officially Shuts its Doors to the Public" (<https://www.searchenginejournal.com/google-pagerank-official-shuts-doors-public/161874/>). *Search Engine Journal*. Archived (<https://web.archive.org/web/20170413154033/https://www.searchenginejournal.com/google-pagerank-official-shuts-doors-public/161874/>) from the original on 2017-04-13.
48. Fishkin, Rand; Jeff Pollard (April 2, 2007). "Search Engine Ranking Factors - Version 2" (<http://www.seomoz.org/article/search-ranking-factors>). seomoz.org. Archived (<https://web.archive.org/web/20090507232106/http://www.seomoz.org/article/search-ranking-factors/>) from the original on May 7, 2009. Retrieved May 11, 2009.
49. Dover, D. *Search Engine Optimization Secrets* Indianapolis. Wiley. 2011.
50. Viniker, D. *The Importance of Keyword Difficulty Screening for SEO*. Ed. Schwartz, M. Digital Guidebook Volume 5. News Press. p 160–164.

51. "Ranking of listings: Ranking - Google Places Help" (<https://support.google.com/places/bin/answer.py?hl=en&answer=7091>). Archived (<https://web.archive.org/web/20120526230407/http://support.google.com/places/bin/answer.py?hl=en&answer=7091>) from the original on 2012-05-26. Retrieved 2011-05-27.
52. Clark, Jack. "Google Turning Its Lucrative Web Search Over to AI Machines" (<https://www.bloomberg.com/news/articles/2015-10-26/google-turning-its-lucrative-web-search-over-to-ai-machines>). Bloomberg. Archived (<https://web.archive.org/web/20160325232656/http://www.bloomberg.com/news/articles/2015-10-26/google-turning-its-lucrative-web-search-over-to-ai-machines>) from the original on 25 March 2016. Retrieved 26 March 2016.
53. Search Engine Watch: Google Directory Has Been Shut Down July 25, 2011 (<https://www.searchenginewatch.com/2011/07/25/google-directory-has-been-shut-down/>)
54. "How to report paid links" (<http://www.mattcutts.com/blog/how-to-report-paid-links/>). mattcutts.com/blog. April 14, 2007. Archived (<https://web.archive.org/web/20070528054725/http://www.mattcutts.com/blog/how-to-report-paid-links/>) from the original on May 28, 2007. Retrieved 2007-05-28.
55. "Google Link Schemes" (<https://support.google.com/webmasters/answer/66356?hl=en>) Archived (<https://web.archive.org/web/20200521035518/https://support.google.com/webmasters/answer/66356?hl=en>) 2020-05-21 at the Wayback Machine links
56. "Evolving" (<https://developers.google.com/search/blog/2019/09/evolving-nofollow-new-ways-to-identify>). *Google Developers*. Retrieved 2022-02-08.
57. "So...You Think SEO Has Changed" (<http://searchenginewatch.com/article/2334934/So...-You-Think-SEO-Has-Changed>). 19 March 2014. Archived (<https://web.archive.org/web/20140331043853/http://searchenginewatch.com/article/2334934/So...-You-Think-SEO-Has-Changed>) from the original on 31 March 2014.
58. Matthew Richardson & Pedro Domingos, A. (2001). *The Intelligent Surfer: Probabilistic Combination of Link and Content Information in PageRank* (<http://research.microsoft.com/pubs/66874/qd-pagerank.pdf>) (PDF). pp. 1441–1448. Archived (<https://web.archive.org/web/20160304112752/http://research.microsoft.com/pubs/66874/qd-pagerank.pdf>) (PDF) from the original on 2016-03-04.
59. Gleich, David F. (January 2015). "PageRank Beyond the Web". *SIAM Review*. **57** (3): 321–363. arXiv:1407.5107 (<https://arxiv.org/abs/1407.5107>). doi:10.1137/140976649 (<https://doi.org/10.1137/140976649>). S2CID 8375649 (<https://api.semanticscholar.org/CorpusID:8375649>).
60. Senanayake, Upul; Piraveenan, Mahendra; Zomaya, Albert (2015). "The Pagerank-Index: Going beyond Citation Counts in Quantifying Scientific Impact of Researchers" (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4545754>). *PLOS ONE*. **10** (8) e0134794. Bibcode:2015PLoSO..1034794S (<https://ui.adsabs.harvard.edu/abs/2015PLoSO..1034794S>). doi:10.1371/journal.pone.0134794 (<https://doi.org/10.1371/journal.pone.0134794>). ISSN 1932-6203 (<https://search.worldcat.org/issn/1932-6203>). PMC 4545754 (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4545754>). PMID 26288312 (<https://pubmed.ncbi.nlm.nih.gov/26288312>).
61. G. Ivan & V. Grolmusz (2011). "When the Web meets the cell: using personalized PageRank for analyzing protein interaction networks" (<https://doi.org/10.1093/bioinformatics/btq680>). *Bioinformatics*. **27** (3): 405–7. doi:10.1093/bioinformatics/btq680 (<https://doi.org/10.1093/bioinformatics/btq680>). PMID 21149343 (<https://pubmed.ncbi.nlm.nih.gov/21149343>).
62. D. Banky and G. Ivan and V. Grolmusz (2013). "Equal opportunity for low-degree network nodes: a PageRank-based method for protein target identification in metabolic graphs" (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3558500>). *PLOS ONE*. **8** (1): 405–7. Bibcode:2013PLoSO...854204B (<https://ui.adsabs.harvard.edu/abs/2013PLoSO...854204B>). doi:10.1371/journal.pone.0054204 (<https://doi.org/10.1371/journal.pone.0054204>). PMC 3558500 (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3558500>). PMID 23382878 (<https://pubmed.ncbi.nlm.nih.gov/23382878>).

63. Burns, Judith (2009-09-04). "Google trick tracks extinctions" (<https://news.bbc.co.uk/1/hi/sci/tech/8238462.stm>). *BBC News*. Archived (<https://web.archive.org/web/20110512060125/http://news.bbc.co.uk/2/hi/science/nature/8238462.stm>) from the original on 2011-05-12. Retrieved 2011-05-27.
64. Benjamin M. Schmidt & Matthew M. Chingos (2007). "Ranking Doctoral Programs by Placement: A New Method" (http://www.people.fas.harvard.edu/~gillum/rankings_paper.pdf) (PDF). *PS: Political Science and Politics*. **40** (July): 523–529. CiteSeerX 10.1.1.582.9402 (<https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.582.9402>). doi:10.1017/s1049096507070771 (<https://doi.org/10.1017%2Fs1049096507070771>). S2CID 6012229 (<https://api.semanticscholar.org/CorpusID:6012229>). Archived (https://web.archive.org/web/20150213104340/http://www.people.fas.harvard.edu/~gillum/rankings_paper.pdf) (PDF) from the original on 2015-02-13.
65. Johan Bollen; Marko A. Rodriguez; Herbert Van de Sompel (December 2006). "MESUR: Usage-based metrics of scholarly impact". *Proceedings of the 7th ACM/IEEE-CS joint conference on Digital libraries*. New York: Association for Computing Machinery. arXiv:cs.GL/0601030 (<https://arxiv.org/abs/cs.GL/0601030>). Bibcode:2006cs.....1030B (<https://ui.adsabs.harvard.edu/abs/2006cs.....1030B>). doi:10.1145/1255175.1255273 (<https://doi.org/10.1145%2F1255175.1255273>). ISBN 978-1-59593-644-8. S2CID 3115544 (<https://api.semanticscholar.org/CorpusID:3115544>).
66. Fletcher, Jack McKay; Wennekers, Thomas (2017). "From Structure to Activity: Using Centrality Measures to Predict Neuronal Activity" (<https://doi.org/10.1142%2FS0129065717500137>). *International Journal of Neural Systems*. **28** (2): 1750013. doi:10.1142/S0129065717500137 (<https://doi.org/10.1142%2FS0129065717500137>). hdl:10026.1/9713 (<https://hdl.handle.net/10026.1%2F9713>). PMID 28076982 (<https://pubmed.ncbi.nlm.nih.gov/28076982>).
67. Gupta, Pankaj; Goel, Ashish; Lin, Jimmy; Sharma, Aneesh; Wang, Dong; Zadeh, Reza (2013). "WTF: The Who to Follow Service at Twitter" (<http://dl.acm.org/citation.cfm?id=2488433>). *Proceedings of the 22nd International Conference on World Wide Web*. ACM. pp. 505–514. doi:10.1145/2488388.2488433 (<https://doi.org/10.1145%2F2488388.2488433>). ISBN 978-1-4503-2035-1. S2CID 207205045 (<https://api.semanticscholar.org/CorpusID:207205045>). Retrieved 11 December 2018.
68. Ha, Anthony (2012-05-08). "Y Combinator-Backed Swifttype Builds Site Search That Doesn't Suck" (<https://techcrunch.com/2012/05/08/swifttype-launch/>). *TechCrunch*. Archived (<https://web.archive.org/web/20140706215017/http://techcrunch.com/2012/05/08/swifttype-launch/>) from the original on 2014-07-06. Retrieved 2014-07-08.
69. "Working Papers Concerning the Creation of Google" (<http://dbpubs.stanford.edu:8091/diglib/pub/projectdir/google.html>). *Google*. Archived (<https://web.archive.org/web/20061128230555/http://dbpubs.stanford.edu:8091/diglib/pub/projectdir/google.html>) from the original on November 28, 2006. Retrieved November 29, 2006.
70. Cho, J.; Garcia-Molina, H.; Page, L. (1998). "Efficient crawling through URL ordering" (<http://web.archive.org/web/20080603171020/http://dbpubs.stanford.edu:8090/pub/1998-51>). *Proceedings of the Seventh Conference on World Wide Web*. Archived from the original (<http://dbpubs.stanford.edu:8090/pub/1998-51>) on 2008-06-03.
71. "Yahoo! Groups" (<https://web.archive.org/web/20131004234101/http://groups.yahoo.com/group/sd-2/>). Groups.yahoo.com. Archived from the original (<https://groups.yahoo.com/group/sd-2/>) on 2013-10-04. Retrieved 2013-10-02.
72. "Autopoietic Information Systems in Modern Organizations". CiteSeerX 10.1.1.148.9274 (<https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.148.9274>).
73. Zack, Laurie; Lamb, Ron; Ball, Sarah (2012-12-31). "An application of Google's PageRank to NFL rankings" (<https://doi.org/10.2140%2Finvolve.2012.5.463>). *Involve: A Journal of Mathematics*. **5** (4): 463–471. doi:10.2140/involve.2012.5.463 (<https://doi.org/10.2140%2Finvolve.2012.5.463>). ISSN 1944-4184 (<https://search.worldcat.org/issn/1944-4184>).

74. Peña, Javier López; Touchette, Hugo (2012-06-28). "A network theory analysis of football strategies". arXiv:1206.6904 (<https://arxiv.org/abs/1206.6904>) [math.CO (<https://arxiv.org/archive/math.CO>)].
75. Beggs, Clive B.; Shepherd, Simon J.; Emmonds, Stacey; Jones, Ben (2017-06-02). Zhou, Wei-Xing (ed.). "A novel application of PageRank and user preference algorithms for assessing the relative performance of track athletes in competition" (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5456068>). *PLOS ONE*. **12** (6) e0178458. Bibcode:2017PLoSO..1278458B (<https://ui.adsabs.harvard.edu/abs/2017PLoSO..1278458B>). doi:10.1371/journal.pone.0178458 (<https://doi.org/10.1371%2Fjournal.pone.0178458>). ISSN 1932-6203 (<https://search.worldcat.org/issn/1932-6203>). PMC 5456068 (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5456068>). PMID 28575009 (<https://pubmed.ncbi.nlm.nih.gov/28575009>).
76. B. Jiang (2006). "Ranking spaces for predicting human movement in an urban environment". *International Journal of Geographical Information Science*. **23** (7): 823–837. arXiv:physics/0612011 (<https://arxiv.org/abs/physics/0612011>). Bibcode:2009IJGIS..23..823J (<https://ui.adsabs.harvard.edu/abs/2009IJGIS..23..823J>). doi:10.1080/13658810802022822 (<https://doi.org/10.1080%2F13658810802022822>). S2CID 26880621 (<https://api.semanticscholar.org/CorpusID:26880621>).
77. Jiang B.; Zhao S. & Yin J. (2008). "Self-organized natural roads for predicting traffic flow: a sensitivity study". *Journal of Statistical Mechanics: Theory and Experiment*. P07008 (7): 008. arXiv:0804.1630 (<https://arxiv.org/abs/0804.1630>). Bibcode:2008JSMTE..07..008J (<https://ui.adsabs.harvard.edu/abs/2008JSMTE..07..008J>). doi:10.1088/1742-5468/2008/07/P07008 (<https://doi.org/10.1088%2F1742-5468%2F2008%2F07%2FP07008>). S2CID 118605727 (<https://api.semanticscholar.org/CorpusID:118605727>).
78. Roberto Navigli, Mirella Lapata. "An Experimental Study of Graph Connectivity for Unsupervised Word Sense Disambiguation" (http://www.dsi.uniroma1.it/~navigli/pubs/PAMI_2010_Navigli_Lapata.pdf) Archived (https://web.archive.org/web/20101214110431/http://www.dsi.uniroma1.it/~navigli/pubs/PAMI_2010_Navigli_Lapata.pdf) 2010-12-14 at the Wayback Machine. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 32(4), IEEE Press, 2010, pp. 678–692.
79. M. T. Pilehvar, D. Jurgens and R. Navigli. Align, Disambiguate and Walk: A Unified Approach for Measuring Semantic Similarity. (http://wwwusers.di.uniroma1.it/~navigli/pubs/ACL_2013_Pilehvar_Jurgens_Navigli.pdf) Archived (https://web.archive.org/web/20131001004443/http://wwwusers.di.uniroma1.it/~navigli/pubs/ACL_2013_Pilehvar_Jurgens_Navigli.pdf) 2013-10-01 at the Wayback Machine. *Proc. of the 51st Annual Meeting of the Association for Computational Linguistics (ACL 2013)*, Sofia, Bulgaria, August 4–9, 2013, pp. 1341-1351.
80. Andrea Esuli & Fabrizio Sebastiani. "PageRanking WordNet synsets: An Application to Opinion-Related Properties" (<http://nmis.isti.cnr.it/sebastiani/Publications/ACL07.pdf>) (PDF). *In Proceedings of the 35th Meeting of the Association for Computational Linguistics, Prague, CZ, 2007*, pp. 424–431. Archived (<https://web.archive.org/web/20070628113247/http://nmis.isti.cnr.it/sebastiani/Publications/ACL07.pdf>) (PDF) from the original on June 28, 2007. Retrieved June 30, 2007.
81. Wang S.; Schreckenberg M.; Guhr T (2023). "Transitions between quasi-stationary states in traffic systems: Cologne orbital motorways as an example" (<https://doi.org/10.1088%2F1742-5468%2F2023i3401W>). *Journal of Statistical Mechanics: Theory and Experiment*. **2023** (9): 093401. arXiv:2302.14596 (<https://arxiv.org/abs/2302.14596>). Bibcode:2023JSMTE2023i3401W (<https://ui.adsabs.harvard.edu/abs/2023JSMTE2023i3401W>). doi:10.1088/1742-5468/acf210 (<https://doi.org/10.1088%2F1742-5468%2Facf210>). S2CID 257232659 (<https://api.semanticscholar.org/CorpusID:257232659>).
82. "Preventing Comment Spam" (<http://googleblog.blogspot.com/2005/01/preventing-comment-spam.html>). *Google*. Archived (<https://web.archive.org/web/20050612001742/http://googleblog.blogspot.com/2005/01/preventing-comment-spam.html>) from the original on June 12, 2005. Retrieved January 1, 2005.

83. "PageRank Sculpting: Parsing the Value and Potential Benefits of Sculpting PR with Nofollow" (<http://www.seomoz.org/blog/pagerank-sculpting-parsing-the-value-and-potential-benefits-of-sculpting-pr-with-nofollow>). SEOMoz. 14 October 2008. Archived (<https://web.archive.org/web/20110514201211/http://www.seomoz.org/blog/pagerank-sculpting-parsing-the-value-and-potential-benefits-of-sculpting-pr-with-nofollow>) from the original on 2011-05-14. Retrieved 2011-05-27.
84. "PageRank sculpting" (<http://www.mattcutts.com/blog/pagerank-sculpting/>). Mattcutts.com. 2009-06-15. Archived (<https://web.archive.org/web/20110511180227/http://www.mattcutts.com/blog/pagerank-sculpting/>) from the original on 2011-05-11. Retrieved 2011-05-27.

Sources

- Altman, Alon; Moshe Tennenholtz (2005). "Ranking Systems: The PageRank Axioms" (<http://www.eecs.harvard.edu/cs286r/courses/fall11/papers/AT%2705.pdf>) (PDF). *Proceedings of the 6th ACM conference on Electronic commerce (EC-05)*. Vancouver, BC. Retrieved 29 September 2014.
- Cheng, Alice; Eric J. Friedman (2006-06-11). "Manipulability of PageRank under Sybil Strategies" (<http://www.cs.duke.edu/nicl/netecon06/papers/ne06-sybil.pdf>) (PDF). *Proceedings of the First Workshop on the Economics of Networked Systems (NetEcon06)*. Ann Arbor, Michigan. Archived (<https://web.archive.org/web/20100821011759/http://www.cs.duke.edu/nicl/netecon06/papers/ne06-sybil.pdf>) (PDF) from the original on 2010-08-21. Retrieved 2008-01-22.
- Farahat, Ayman; LoFaro, Thomas; Miller, Joel C.; Rae, Gregory; Ward, Lesley A. (2006). "Authority Rankings from HITS, PageRank, and SALSA: Existence, Uniqueness, and Effect of Initialization". *SIAM Journal on Scientific Computing*. **27** (4): 1181–1201. Bibcode:2006SJSC...27.1181F (<https://ui.adsabs.harvard.edu/abs/2006SJSC...27.1181F>). CiteSeerX 10.1.1.99.3942 (<https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.99.3942>). doi:10.1137/S1064827502412875 (<https://doi.org/10.1137/S1064827502412875>).
- Haveliwala, Taher; Jeh, Glen; Kamvar, Sepandar (2003). "An Analytical Comparison of Approaches to Personalizing PageRank" (<http://www-cs-students.stanford.edu/~taherh/papers/comparison.pdf>) (PDF). *Stanford University Technical Report*. Archived (<https://web.archive.org/web/20101216084254/http://www-cs-students.stanford.edu/~taherh/papers/comparison.pdf>) (PDF) from the original on 2010-12-16. Retrieved 2008-11-13.
- Langville, Amy N.; Meyer, Carl D. (2003). "Survey: Deeper Inside PageRank". *Internet Mathematics*. **1** (3).
- Langville, Amy N.; Meyer, Carl D. (2006). *Google's PageRank and Beyond: The Science of Search Engine Rankings*. Princeton University Press. ISBN 978-0-691-12202-1.
- Richardson, Matthew; Domingos, Pedro (2002). "The intelligent surfer: Probabilistic combination of link and content information in PageRank" (<http://www.cs.washington.edu/homes/pedrod/papers/nips01b.pdf>) (PDF). *Proceedings of Advances in Neural Information Processing Systems*. Vol. 14. Archived (<https://web.archive.org/web/20100628120615/http://www.cs.washington.edu/homes/pedrod/papers/nips01b.pdf>) (PDF) from the original on 2010-06-28. Retrieved 2004-09-18.

Relevant patents

- Original PageRank U.S. Patent—Method for node ranking in a linked database (<http://patft.uspto.gov/netacgi/nph-Parser?patentnumber=6,285,999>) Archived (<https://web.archive.org/web/20140829134755/http://patft.uspto.gov/netacgi/nph-Parser?patentnumber=6,285,999>) 2014-08-29 at the Wayback Machine—Patent number 6,285,999—September 4, 2001

- PageRank U.S. Patent—Method for scoring documents in a linked database (<http://patft1.uspto.gov/netacgi/nph-Parser?Sect1=PTO1&Sect2=HITOFF&d=PALL&p=1&u=%2Fnetahhtml%2FPTO%2Fsrchnum.htm&r=1&f=G&l=50&s1=6,799,176.PN.&OS=PN/6,799,176&RS=PN/6,799,176>)—Patent number 6,799,176—September 28, 2004
- PageRank U.S. Patent—Method for node ranking in a linked database (<http://patft.uspto.gov/netacgi/nph-Parser?Sect1=PTO2&Sect2=HITOFF&u=%2Fnetahhtml%2FPTO%2Fsearch-adv.htm&r=1&p=1&f=G&l=50&d=PTXT&S1=7,058,628.PN.&OS=pn/7,058,628&RS=PN/7,058,628>) Archived (<https://web.archive.org/web/20190828223450/http://patft.uspto.gov/netacgi/nph-Parser?Sect1=PTO2&Sect2=HITOFF&u=%2Fnetahhtml%2FPTO%2Fsearch-adv.htm&r=1&p=1&f=G&l=50&d=PTXT&S1=7%2C058%2C628.PN.&OS=pn%2F7%2C058%2C628&RS=PN%2F7%2C058%2C628>) 2019-08-28 at the [Wayback Machine](https://web.archive.org/web/20190828223450/http://patft.uspto.gov/netacgi/nph-Parser?Sect1=PTO2&Sect2=HITOFF&u=%2Fnetahhtml%2FPTO%2Fsearch-adv.htm&r=1&p=1&f=G&l=50&d=PTXT&S1=7,058,628.PN.&OS=pn/7,058,628&RS=PN/7,058,628)—Patent number 7,058,628—June 6, 2006
- PageRank U.S. Patent—Scoring documents in a linked database (<http://patft.uspto.gov/netacgi/nph-Parser?Sect1=PTO2&Sect2=HITOFF&u=%2Fnetahhtml%2FPTO%2Fsearch-adv.htm&r=1&p=1&f=G&l=50&d=PTXT&S1=7,269,587.PN.&OS=pn/7,269,587&RS=PN/7,269,587>) Archived (<https://web.archive.org/web/20180331090147/http://patft.uspto.gov/netacgi/nph-Parser?Sect1=PTO2>) 2018-03-31 at the [Wayback Machine](https://web.archive.org/web/20180331090147/http://patft.uspto.gov/netacgi/nph-Parser?Sect1=PTO2)—Patent number 7,269,587—September 11, 2007

External links

- Algorithms (<https://www.google.com/insidesearch/howsearchworks/algorithms.html>) by Google
 - Our products and services (<https://www.google.com/technology/>) by Google
 - How Google Finds Your Needle in the Web's Haystack (<https://www.ams.org/featurecolumn/archive/pagerank.html>) by the American Mathematical Society
-

Retrieved from "<https://en.wikipedia.org/w/index.php?title=PageRank&oldid=1334279166>"