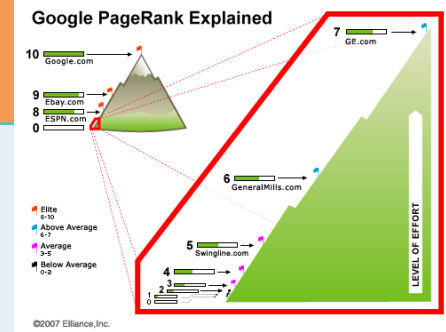


Apply It.

The math behind... Google PageRank™



Technical terms used:

Graph theory, random walk, Markov chain, linear system, eigenvector, power iteration

Uses and applications:

PageRank™ is one of many factors that influence what pages appear at the top of the list of results of a Google™ search. The PageRank value of a webpage is a measure of that page’s relative importance according to the hyperlink structure of the surrounding web.

How it works:

Imagine a person surfing the web by randomly clicking on hyperlinks to travel between pages. The path of this “random surfer” can be viewed as a random walk on the vertices of the web graph and can be described mathematically as a trajectory of the Markov chain whose transition probabilities are determined by the hyperlink structure of the web. The likelihood that, in the long run, the surfer ends up on a specific page is entirely dependent on this hyperlink structure. In general, this long-term likelihood is called the stationary distribution of the Markov chain, but in the context of a random surfer traversing the web graph it is called PageRank.

Under certain assumptions about the structure of the web graph, PageRank turns out to be the unique unit eigenvector corresponding to the largest eigenvalue of the hyperlink matrix. In practice, this eigenvector can be computed via the power iteration, a common technique in linear systems theory.

Google originally applied the PageRank algorithm to a web of 24 million pages in 1998. By 2014, the size of the indexed web had grown thousand-fold to the order of tens of billions. As the web continues to grow, challenges arise in the computation of PageRank. Although these challenges are somewhat mitigated by improving computer technology, they are also being addressed through academic research on algorithm efficiency. Recent approaches include a distributed randomized updating of PageRank within a web-aggregation framework.

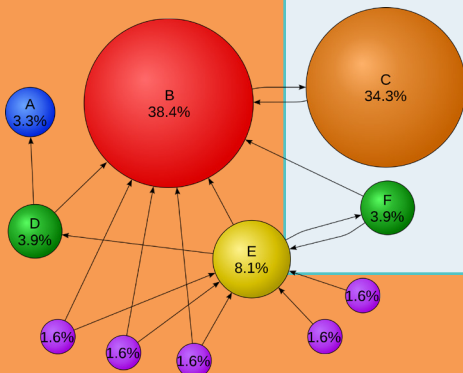
Interesting fact:

While the term PageRank is self-referential, describing a measure of rank associated with each page in the indexed web, it is also an intentional play on the name of Google co-founder and CEO, Larry Page.

References:

[1] S. Brin and L. Page, The anatomy of a large-scale hypertextual web search engine, *Computer Networks and ISDN Sys.* 30(1-7), 107-117 (April 1998).

[2] H. Ishii and R. Tempo, The PageRank problem, multiagent consensus, and web aggregation: A systems and control viewpoint, *IEEE Control Sys. Mag.* 34(3), 34-53 (June 2014).



Submitted by Cody Clifton, University of Kansas, Third Place, Math Matters, Apply it! Contest, February 2015.