



The Illusion in the Presentation of the Rank of a Web Page with Dangling Links

*¹FELIX UKPAI OGBAN; PRINCE OGHENEKARO ASAGBA (PH.D.); OLUMIDE OWOLABI (PH.D.)

¹Department of Mathematics, Statistics, & Computer Science, Faculty of Science,
University of Calabar, Nigeria
Email: felix.ogban@gmail.com

²Department of Computer Science, Faculty of Physical & Information Technology, University of Port Harcourt, Nigeria
Email: pasagba@yahoo.com

³Computer Center, University of Abuja, Nigeria

Keywords: Hyperlinked, Page Rank algorithm, Page Rank calculation, Google Toolbar, in-bound and out-bound links, web page.

ABSTRACT: The hyperlinked display list of search results from any given search engine is an illusion of the real order in priority if its position is based on a pageRank computation. This is because, several factors governs the pageRank computation. Besides the fact that different search engines fashions their ranking model, it is an established fact that, the **hub** and **authority** factor must be considered. This paper considered the effects of an in-bound (authority) and out-bound (hub) links on the rank of a page. **Hyperlink-Induced Topic Search (HITS)** (also known as **hubs and authorities**) is a link analysis algorithm that rates Web pages, developed by Jon Kleinberg (1999). It was a precursor to PageRank. The idea behind Hubs and Authorities stemmed from a particular insight into the creation of web pages when the Internet was originally forming; that is, certain web pages, known as hubs, served as large directories that were not actually authoritative in the information that it held, but were used as compilations of a broad catalog of information that led users directly to other authoritative pages. In other words, a good hub represented a page that pointed to many other pages, and a good authority represented a page that was linked by many different hubs. A good hub page is one that points to many good authorities; a good authority page is one that is pointed to by many good hub pages. We focused on the Google's Toolbar with regard to pages given a certain toolbar Page Rank, but with an inbound link from a page that has a toolbar Page Rank, which is higher by one. This is done to alleviate the effect of the removal from the database(s), pages with no outbound links as proposed by Page and Brin, and applied by Google for the normalization of the dangling links. We considered, six (6) Web site clips scenarios to show the effects of inbound and outbound links, vis a vis the number of Pages in the Web, and the influence of the Damping Factor. We observed that the linkage of sets of dangling sites/pages (PDF, MS-word infested pages) and the application of a new ranking model for them is better in smoothening, the hub and authority complimentary effects in page ranking. We gave recommendations on how search engines and crawlers could weight and produce a better ranking of pages for users. Thus, users can within the first few search results, get to their expected search results. © JASEM

Some people may have the idea that generating millions of pages is a good way to produce Page Rank and improve ranking of their website after studying the Page Rank algorithm. Theoretically, this should work, assuming that an appropriate linking structure is chosen. However, it does not practically work that way. **Google** changed their algorithm several years ago. One of the changes prevents the generation of Page Rank by weighted suggestions from page designers (self-links) (Googlewebmaster, 2012). Moreover in practice, you even need Page Rank to get larger websites completely crawled. Again, Google made some minor changes of the algorithm, that is, they changed the value of the damping factor which originally was $d=0.85$ to 0.50

Another aspect is the question: which links are counted for Page Rank calculation by Google? If page **A** is linking two times to page **B** and one time to page **C**, then there are different possibilities how Page Rank is split between page **B** and **C**. The current implementation of the Page Rank algorithm is ignoring multiple links, that is, in the example above, the transferred Page Rank is split 1 by 1 between page **B** and page **C**. Self-links are another example of ambiguousness. Currently, Google is counting self-links for Page Rank calculation. However, links with the attribute *rel="nofollow"* are ignored just as robots.txt file contents are ignored during crawling.

Currently, there are just two ways to get independent information about the Page Rank: (1) the Google toolbar, and (2) Google directory. Of course, there are