Limitations of PageRank

PageRank is one of the algorithms used by Google Search to determine order of pages in the search results. The PageRank algorithm determines the score of a page based on the scores of the pages that link to it, whose scores in turn depend on the scores of the pages that link to them and so on. The PageRank algorithm has been extended to ordering papers in citation networks, predicting traffic in traffic networks, finding the best teams and athletes in sport networks, and ranking systems that contributed to anomalous behavior in debugging. Based on the wide range of its applications, it is necessary that the PageRank algorithm gives the right results.

One of the limitations of the PageRank algorithm is that it's ordering does not favor current events (Cornell, 2015). According to the algorithm, old pages typically have more votes because they have more links from other reputable pages. This means that a new page will not be as reputable until it has gained exposure and links from other pages. As a web user, I would like to see the most recent research papers in a certain field.

Another limitation is that in its use in ranking papers in the citation network, it does not account for the size of a field. The number of citations per paper in each field varies widely depending on the discipline, for example, an average paper is cited about 6 times in life sciences, 3 times in physics, and about 1 time in mathematics (Maslov, 2008). The algorithm is therefore more likely to give a paper in a mathematics field a lower score than a paper in a life sciences field.

Another limitation is that it is prone to manipulation through link selling. Originally, Google used to display the rank of a page on the browser, but this caused a lot of controversy and Google made this metric private. During the era of public PageRank, a market emerged for link selling as Search Engine Optimizers (SEOs) found a way to manipulate the algorithm by creating more traffic to their pages. The impact of artificially creating more traffic to a page was that it the page was unfairly higher up in scores according the algorithm.

Another limitation is that it unleashed link spamming. Link spamming refers to the practice of leaving links to a page unnecessarily on various platforms. To counter this development, Google introduced the nofollow attribute in the link tag in HTML to prevent such links from passing PageRank votes along (Niechai, 2018).

In conclusion, the PageRank algorithm opened up web search to both advancement and manipulation but with optimizations, we can take full advantage of the benefits it offers.

References

1. Maslov, S., & Redner, S. (2008, October 29). Promise and Pitfalls of Extending Google's PageRank Algorithm to Citation Networks. Retrieved from http://www.jneurosci.org/content/28/44/11103.short
2. Niechai, V. (n.d.). Beginner's Guide to Google PageRank: How It Works & Why It Still Matters in 2018. Retrieved from https://www.link-assistant.com/news/page-rank-2018.html
3. Networks. (n.d.). Retrieved from https://blogs.cornell.edu/info2040/2015/10/28/limitations-of-pagerank/