

Page Rank Optimization interim report

Group 1

Huo Zhenkun
Student ID:120090822

Li Nanxuan
Student ID:120090860

Liu Hengrui
Student ID:120090609

Zhao Dule
Student ID:120090080

I. ABSTRACT

In this project, we tend to develop methods to increase our homepage's importance score, based on our knowledge of Google's page ranking algorithm. We are going to compare several methods by reading some relevant papers and Google's guidance and find out which ones are most effective. In addition, we develop a website to do some experiments to adjust and prove our methods.

II. REPORT INTRODUCTION

This interim report for our Google Page Rank Optimization project is divided into three parts. Firstly, the literature review section helps us identify the factors that impact page ranking. Next, we set up experiments based on a website we created and outline the expected results for the entire project. Finally, we'll summarize our findings in the last section.

III. TEXTBOOK REVIEW

According to our textbook, hyperlinked web pages can be treated as a network, and Google determines the importance of web pages using their connectivity patterns. If many important web pages point to a particular web page, it may be deemed important. Each outgoing neighbor of a node i receives an importance score:

$$\frac{\pi_i}{O_i}$$

A node's importance score is the sum of the importance scores received from all incoming neighbors, indexed by j . The set of scores can be viewed as the dominant eigenvector of a Google matrix:

$$G = \theta \hat{H} + (1 - \theta) \frac{1}{N} \mathbf{1}\mathbf{1}^T$$

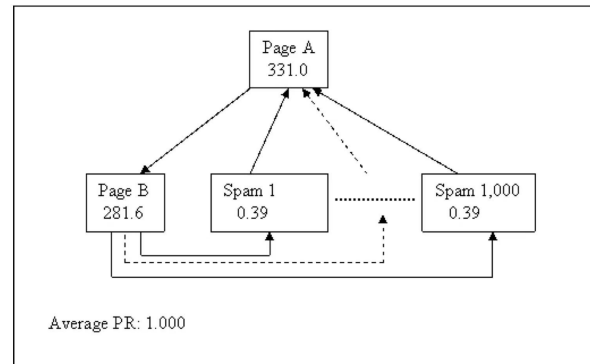
IV. LITERATURE REVIEW

A. Paper review part 1

We have read some papers regarding the Google page rank algorithm and found some factors that may have impacts on the page ranking of websites. [1] There was a paper published in 2002 by Ian Rogers from IPR Computing Ltd that indicated how the Google page rank algorithm works. According to him, Google specifies their page rank algorithm for a website A , $PR(A)$ is:

$$PR(A) = (1-d) + d \left(\frac{PR(T1)}{C(T1)} + \frac{PR(T2)}{C(T2)} + \dots + \frac{PR(Tn)}{C(Tn)} \right)$$

where d is a dumping factor usually set to 0.85 and $T1$ to Tn are the web pages pointing to A while $C(x)$ is the number of outgoing links of web page x . Ian also did some experiments on the structure of the web page node graph. He tried calculating page rank based on different structures of the graph and compared the result. He found out that the best structure for increasing the page rank of a single web page is shown in the following graph:



This hierarchical layout makes sure pages that contribute to our target page vote the most. Because Ian mentioned in his paper that Google may not count the vote of web pages that has no incoming links, so at least there should be one incoming link to the pages voting for our target web page. Ian also mentioned content quality as an important criterion for high page rank. However, he didn't specify how the content quality affects the page rank. From what Ian has discovered, there are at least two factors. One is the incoming links and the other is the outgoing links.

B. Paper review part 2

We have read a paper that discusses the factors that may impact the Page ranking on Google. [2] The author conducted an analysis of several factors that may affect website page rank, including backlinks count, website age, content relevance, traffic, keyword usage, mobile responsiveness, contact pages, and sitemaps. The author reviewed SEO updates, created web pages, and collected search results from Google to assess these factors. The following findings were obtained:

- 1) A website's backlinks count and quality, both incoming and outgoing, play a crucial role in determining its page rank.
- 2) The age of a page has only a limited impact on its ranking.
- 3) The quality of a page's content is a significant factor in ranking.
- 4) Higher website traffic and visits can increase a website's ranking by attracting web crawlers.
- 5) Keyword usage on websites is important, but excessive use and stuffing can lead to Google penalties.
- 6) Mobile responsiveness influences page ranking, with faster load times resulting in higher page rank.
- 7) Creating contact pages and submitting sitemaps can lead to a higher website ranking.
- 8) Google does not assign significant weight to the frequency of search terms on a website's full content.

C. Google's Guidance

Google has released a [guidance](#) that can help website owners find their sites on search engines and improve their page ranks. According to this paper, website owners should first upload a sitemap to help Google's crawler identify their website. They should also make their website's contents attractive to the crawler by using techniques such as creating unique, accurate pages or using meta descriptions. A well-designed website hierarchy can also boost page rank. Ultimately, the most important way to increase page rank is to improve the website's content quality for the target audience. Finally, adding links to relevant pages is highly encouraged by Google in this paper.

V. EXPERIMENT DESIGN

A. Factors

After discussion and evaluation, we choose some important factors which are not difficult to apply to our pages.

- 1) In degree
- 2) Out degree
- 3) Keyword and Content quality
- 4) Website visits
- 5) Website responsiveness

B. Experiment

1) *Experimental Platform:* We created 5 web pages(including 1 homepage and 4 sub-pages (check fig.1))for CUHK(SZ) student homework on a rented server located in Hong Kong to improve responsiveness for our customers in Shenzhen and Hong Kong. A Google search showed only about 1000 results related to our content, making it easy to detect page rank changes. As newly created web pages, they are likely to appear near the bottom of search results, and small changes in factors can affect their location on the page. As we created all 5 pages, we can easily modify factors to observe their effects on page rank.

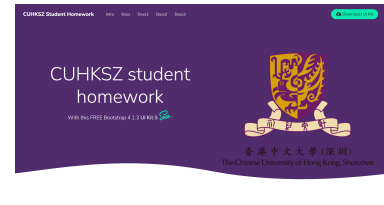


Fig. 1. The screenshot of our website's homepage(the bar on top indicate the 4 subpages)

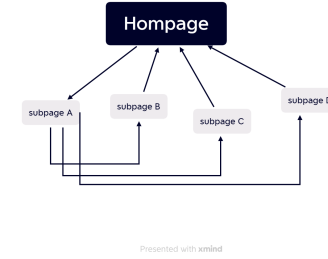


Fig. 2. Original graph

2) *Experiment Methodology:* The original graph structure is shown below.

1) In-degree

There will be one pair of experiments. One with an incoming link from a mega website like The Chinese Software Developer Network (CSDN). Therefore, our page rank could get a huge increase from the incoming links like Fig.3. Another one will be kept unchanged as the original graph as a comparison.

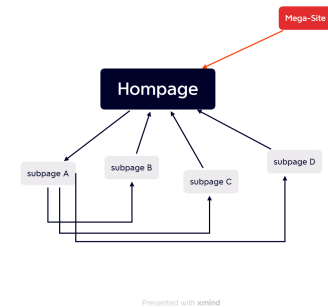


Fig. 3. Network with web page A has an incoming link from mega-site

2) Out-degree

There will be three experiments compared to each other. Three network architectures are listed in Fig.4. and Fig.5. and the original graph. We first start from the original graph. Delete one outgoing link to one of our sub-pages to achieve the graph where the homepage has no outgoing link like Fig. 3. Finally, we add an

outgoing link to a mega-site and compare the three results in page rank.

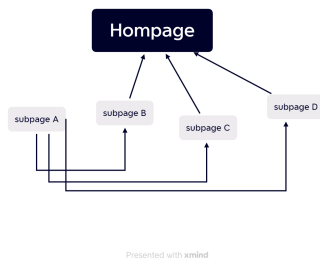


Fig. 4. Network with web page A has no outgoing link from mega-site

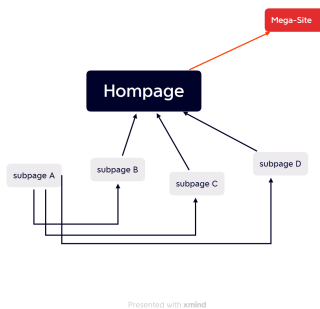


Fig. 5. Network with web page A has an outgoing link toward one of the mega-site

- Zhao Dule: Read and summarize the textbook. Find the website structure of the homepage from the Internet. Modify the origin code of the homepage. (25%)

VII. APPENDIX

- We have uploaded our materials including website source code and reference paper to GitHub. The GitHub link is: https://github.com/GeorgeMac/EIE3280_Project_1.git
- Our page link is: www.lgustu.top

REFERENCES

- [1] I. Rogers, "The google pagerank algorithm and how it works," 2002.
- [2] M. A. Joshi and P. Patel, "Google page rank algorithm and it's updates," in *International Conference on Emerging Trends in Science, Engineering and Management, ICETSEM-2018*, 2018.

3) Keyword and Content quality

We will try our best to write high-quality content to improve the page rank. Since the overuse of keywords may attract penalties from Google, we will use some appropriate methods to demonstrate the contents of the web pages.

4) Website visits

Scripts are used to change the number of visitors to the site. We will use a series of successive visit change values to reflect the changes in page ranks.

5) Website responsiveness

To reduce the loading time of our website, we bought a server in Hong Kong, which is very close to our customers.

VI. CONTRIBUTION

- Huo Zhenkun: Read and summarize the paper. Write the paper review part 1. (25%)
- Li Nanxuan: Read and summarize the paper. Write the paper review part 2. Design some parts of the experiment. (25%)
- Liu Hengrui: Purchase and set up the cloud server. Deploy the web control panel. Design some parts of the experiment. The Polish part of the report. (25%)