# COMP 543, Tools and Models for Data Science

Chengyin Liu, cl93

## *Research #2, PageRank*

This paper [1] was written by L. Page, who described PageRank, a method for rating pages objectively and measuring the human interest devoted to the web pages. This paper deals primarily with the approximation of the overall importance of web pages and gives us a glimpse into the inner mechanics of Google's core search algorithms.

The basic form of PageRank works by factoring in the number and quality of backlinks to a page to roughly determine a website's importance. It considers the whole web as a directed graph. Each web page is a node and there is a directed edge from page A to page B if there is a link/citation from A to B. Thus, we can get the probability that the user will stay at a certain web page when the random walk converges after iterations. Additionally, to avoid the rank sink problem, this algorithm solves the problem by jumping to random pages which is independent to the rank sink.

In the paper [2] at 1998, S. Brin and L. Page already presented the design of Google. By incorporating data regarding the link/citation that points to a given page, the authors were able to better refine search results and greatly increase the precision of PageRank. So, we would say that one of Google's greatest advantages over existing search methods at that time was that it indicated that hyperlink structure of the web is extremely useful in determining search engine query relevance and efficiently utilized it.

However, the original PageRank algorithm does not consider the topic relevance between web pages. Therefore, Haveliwala proposed an improvement for PageRank called Topic-Sensitive PageRank [3] at 2002. This paper defines some topics and computes the score of each page for each topic. Then the algorithm will maintain the vectors for each topic so that the web pages of the same topic will tend to have a high rank when the user switch between pages.

**References:**

[1] Page, Lawrence, et al. *The PageRank citation ranking: Bringing order to the web*. Stanford InfoLab, 1999.

[2] Brin, Sergey, and Lawrence Page. "The anatomy of a large-scale hypertextual web search engine." *Computer networks and ISDN systems* 30.1-7 (1998): 107-117.

[3] Haveliwala, Taher H. "Topic-sensitive pagerank." *Proceedings of the 11th international conference on World Wide Web*. ACM, 2002.