1. What is the key need for this algorithm/approach/system? What are some short comings of existing approaches?

Backlink is the most important element to implement the PageRank. Because the PageRank is the algorithm which use the graph to calculate the weight to determine which page is more significant than other pages. However, more backlinks do not mean the web page is more important than pages which has less backlinks. Some pages (ex: google) might have more important meanings, so it might have larger weight when this link point to other webpages. Moreover, the graph could not contain all the pages in the Internet. That is, when deciding the PageRank, some important pages might miss. Hence, the score of each page might be not precise, so the rank of the pages is not what users want to see.

2. What is the key object (term) that is the solution to this need? (e.g. tweet, RDD, Tensor) Describe this object in a few sentences. This is the fundamental concept or "thing" proposed to solve the need addressed in question #1.

According to what paper mentioned, the graph, which is produced by using backlinks, is the fundamental approach to determine the PageRank. Each page has many backlinks to other pages, so the fact that the pages have lots of backlinks basically implies that the page is relative important. The score which is based on the number of backlinks will be propagated to the pages which they linked to. By this method, we could create a structure graph which contains scores on each page to rank pages.

3. What has the author identified as a weakness or limitation of the proposed algorithm / approach / system? Or what has the author proposed as next steps? If the author does not provide this information, what do you think could be improved?

When the graph has loops in it, it might accumulate the scores and would not link to other pages. That is, the scores will not distribute to other pages which are outside the loops. It is also called as rank sink, and it will lead users to less relevant pages. If we want to solve this problem, we could add the regularization to implement random surf such that the loop will have an exit. Hence, the links will have chance to point to any other pages and break the loop.

Moreover, if someone want to get higher rank, they can create new pages which have many links which points to that page. That page will have more backlinks it should get. The new page might have less backlinks, so the links on it will have less scores. When it

distributes to next layers, the less scores that page can get from it. Therefore, it might still have low rank.

4. What is something interesting your learned from this paper, or your thoughts about its strengths and/or weaknesses. Is there anything else interesting about this paper, or your interpretation of it that you want to share?

What interests me is not only the PageRank could show the rank what users want to read, but how to apply it to other areas. For example, in my opinion, it can be applied to predict what advertisements the website should show to each specific person. By analyzing the link on the graph people clicked, we might be able to find some patterns, and then the machine would understand what users currently need. Hence, the website could promote correct products to right people.

5. Read an additional related paper. Provide a citation for this paper. How does this paper relate to the assigned one? What new information did you learn by reading it? This answer should be a decent sized paragraph describing the content of this paper (be specific) and how it relates to the assigned paper.

The Anatomy of a Large-Scale Hypertextual Web Search Engine

Sergey Brin and Lawrence Page

Computer Science Department, Stanford University, Stanford, CA 94305, USA
sergey@cs.stanford.edu and page@cs.stanford.edu

BRIN, S., AND PAGE, L. 1998. Anatomy of a large-scale hypertextual web search engine. In Proceedings of the 7th International World Wide Web Conference

The authors present Google, a search engine which makes use of the structure present in hypertext. The analysis of link structure via PageRank allows Google to evaluate the quality of web page, and link text as a description of what the link points to also enables the search engine to return high quality result. Therefore, the use of proximity information helps the Google return relevant result. Through this paper, I understand the application of PageRank and how it could help search engineer return the result what users really want. Moreover, the paper also demonstrates how the search engine, Google, gets high quality pages such that those pages typically have high PageRank. Therefore, this paper helps me understand how to improve the accuracy of search result by employing not only PageRank but also anchor text, and proximity information.