**ASSIGNMENT 2**

**Authors:**

|  |  |
| --- | --- |
| **Name** | **ID** |
| *2015A7PS0011H* | *R.Shriram* |
| *2015A7PS0021H* | *Bhavitej* |
| *2015A7PS0967H* | *Akash Sebastian* |

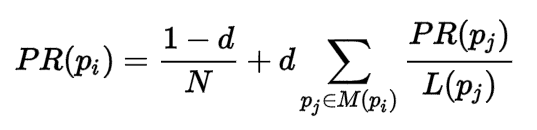
**Code files:**

* main.cpp
* generate\_random\_graph.java
* result.py

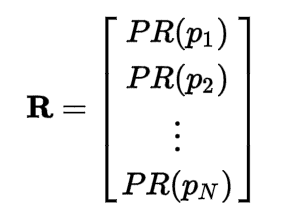
The algorithm implemented in this assignment was Page Rank. PageRank (PR) is an [algorithm](https://en.wikipedia.org/wiki/Algorithm) used by [Google Search](https://en.wikipedia.org/wiki/Google_Search) to rank [websites](https://en.wikipedia.org/wiki/Websites) in their [search engine](https://en.wikipedia.org/wiki/Search_engine) results. PageRank was named after [Larry Page](https://en.wikipedia.org/wiki/Larry_Page), one of the founders of Google. PageRank is a way of measuring the importance of website pages. According to Google,

*“PageRank works by counting the number and quality of links to a page to determine a rough estimate of how important the website is. The underlying assumption is that more important websites are likely to receive more links from other websites.”*

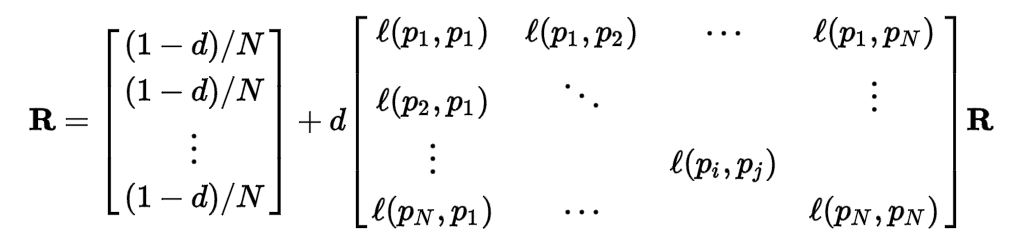
The algorithm is given by the formula,



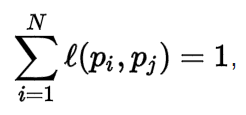
Where d is called the dampening factor or the jump factor, where p1,p2,p3,…,pn are the pages under consideration, (pi) is the set of pages that link to pi, L(pj) is the number of outbound links on page pj, and N is the total number of pages.



Where R is the solution of the equation,



Where the adjacency function l (pi, pj) is 0 if page p(j) does not link to p(i), and normalized such that, for each j



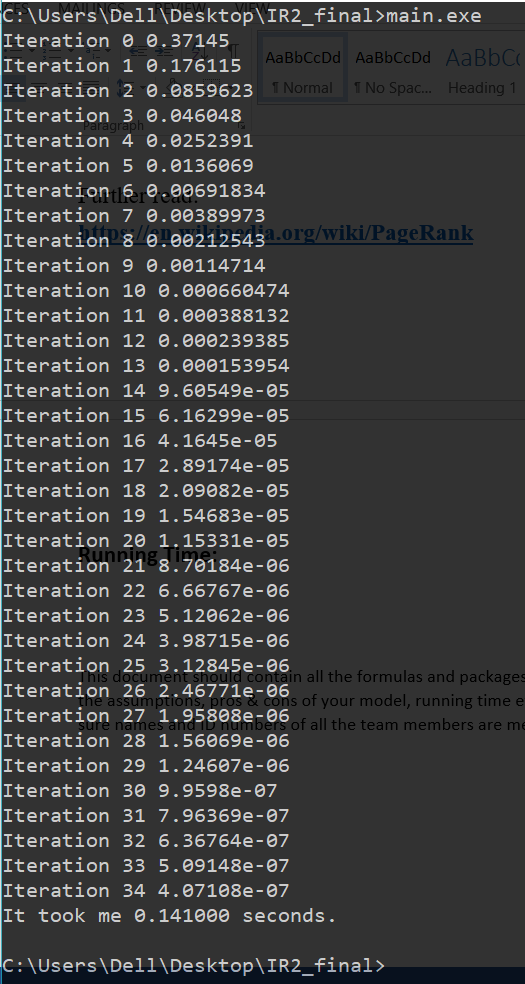
One main disadvantage of PageRank is that it favors older pages. A new page, even a very good one, will not have many links unless it is part of an existing site (a site being a densely connected set of pages, such as [Wikipedia](https://en.wikipedia.org/wiki/Wikipedia)).

Advantage of PageRank is that Dead ends and spider traps are taken care by the jump factor. And any leak in rank is added back into the network.

Further read:

[**https://en.wikipedia.org/wiki/PageRank**](https://en.wikipedia.org/wiki/PageRank)

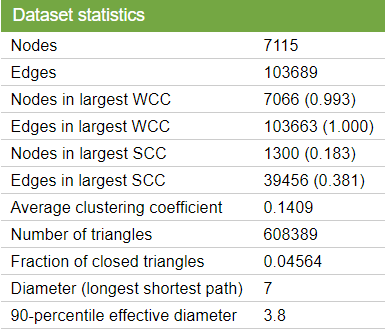
**Running Time:**



**Dataset Used:**

Wikipedia is a free encyclopedia written collaboratively by volunteers around the world. A small part of Wikipedia contributors are administrators, who are users with access to additional technical features that aid in maintenance. In order for a user to become an administrator a Request for adminship (RfA) is issued and the Wikipedia community via a public discussion or a vote decides who to promote to adminship. Using the latest complete dump of Wikipedia page edit history (from January 3 2008) we extracted all administrator elections and vote history data. This gave us 2,794 elections with 103,663 total votes and 7,066 users participating in the elections (either casting a vote or being voted on). Out of these 1,235 elections resulted in a successful promotion, while 1,559 elections did not result in the promotion. About half of the votes in the dataset are by existing admins, while the other half comes from ordinary Wikipedia users.

The network contains all the Wikipedia voting data from the inception of Wikipedia till January 2008. Nodes in the network represent Wikipedia users and a directed edge from node i to node j represents that user i voted on user j.



**REFERENCE:**

