

Assignment 1: Implementation of Pagerank

Lecturer: Dr. Wan-Lei Zhao
Xiamen University

1 Problem

Pagerank is one of the most successful ranking strategies that has been widely adopted by nowadays search engine. The basic idea behind this re-ranking scheme is to make use of the links between different websites. Intuitively, the website which receives many links from other websites should be more important. As a result, higher re-ranking weight should be assigned to it. In this assignment, you are required to implement the pagerank algorithm that has been introduced in **Lecture-4**. The key equation that is used to calculate the pagerank is given as follows.

$$PR(i) = \sum_{\substack{j \\ j \neq i}} \frac{PR(j)}{L(j)} \cdot d + \frac{1-d}{N}, \quad (1)$$

where i and j are the websites, and $PR(\cdot)$ is the pagerank for a website. Function $L(\cdot)$ in the equation returns the total number of outward links from a website. d is the damping factor and N is the number of websites in total. In your implementation, you are recommended to set d to 0.85. You are provided with a matrix of 8×8 . “1” in the matrix indicates an outward link from a row element to a column element. For instance, “1” in the $2nd$ row and $4th$ column indicates an outward link from the $2nd$ element to the $4th$ element.

$$M = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

Hints: When you are updating your $PR(\cdot)$ for one website, you should use the $PR(\cdot)$ values of the others' from the previous round.

2 Requirements

- You are required to implement it in Python;
- Please send the source file ONLY to me. Indicate your student number and student name in both the source code comments and the email;
- The deadline is **2024-Oct.-10**.
- Email: stonescx@gmail.com

Along with the source code, you are required to submit a report in English (PDF format). Basically, there should be three sections. In the first section, the pagerank algorithm should be reviewed, its background, its idea and its impact. In the second section, please elaborate how you implement this algorithm and any improvement you make over this algorithm. In the section followed, please present the visualized results. For instance, you can show the pageranks of the websites by a histogram. If you'd like to, give some comments about the figure presented in the report.