**暨南大学本科实验报告专用纸**

课程名称 数值分析 成绩评定

实验项目名称 ComputingProblems 指导教师 LiangdaFang 实验项目编号 04 实验项目类型 验证 实验地点 N117

学生姓名 陈圣尹 学号 2016051519

学院 信息科学技术学院 系 计算机系 专业 计算机科学与技术

实验时间 年 月 日 午～ 月 日 午 温度 ℃湿度

**Ⅰ、Problem**

1. Implement a simple web crawler;

2. Acquire the google matrix of the first 500 web pages from the main webpage of any

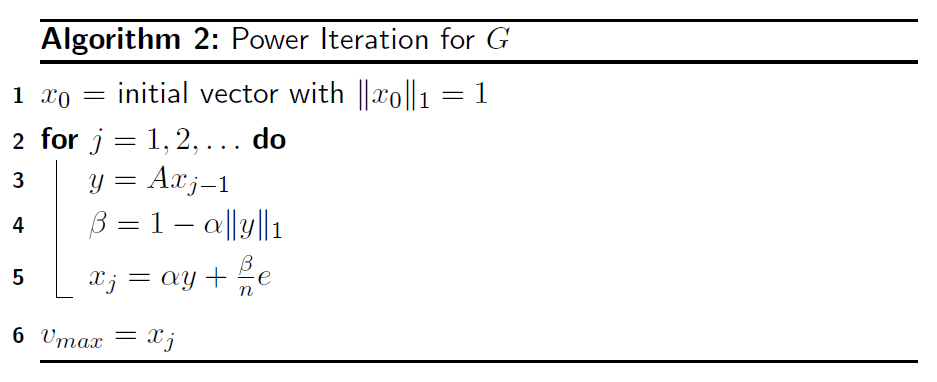
university via the above web crawler, and give their adjacencymatrix;

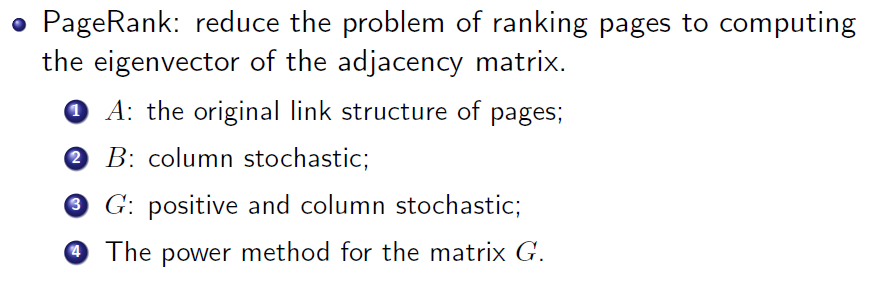
3. Implement the Power Method;

4. Compute the dominant eigenvector of the google matrix;

5. List the top 20 web pages.

**Ⅱ、Algorithm Summary**





**Ⅲ、Experimental procedures**

Step1. Implement a web crawler using Python.

Step2. Acquire the google matrix of the first 500 web pages from the main webpage of JNU University via the above web crawler, give their adjacency matrix and draw the scatter plot of it.

Step3. Implement the Power Method for G.

Step4. Compute the dominant eigenvector of the google matrix.

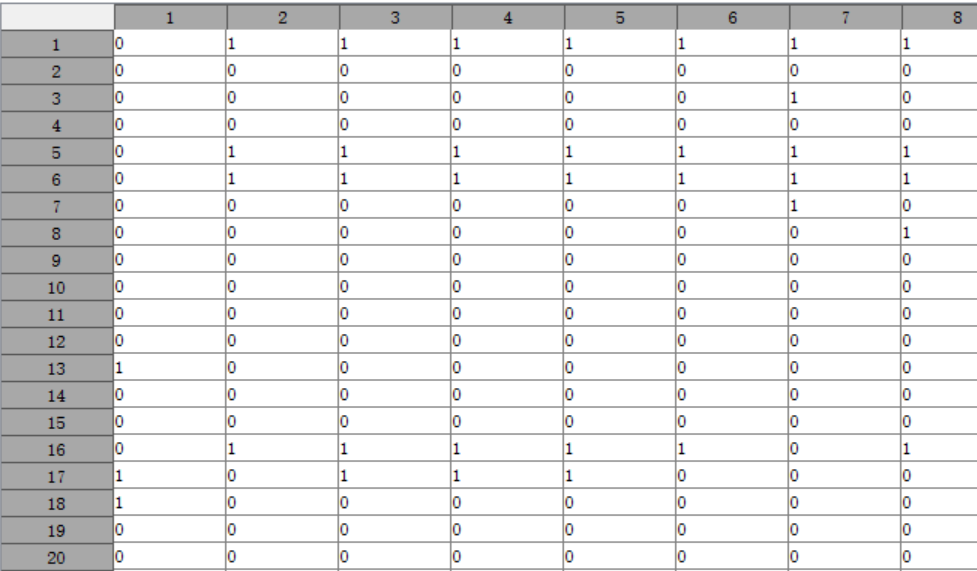
Step5. List the top 20 web pages.

**Ⅳ、Result Analysis**

The web crawler using Python:



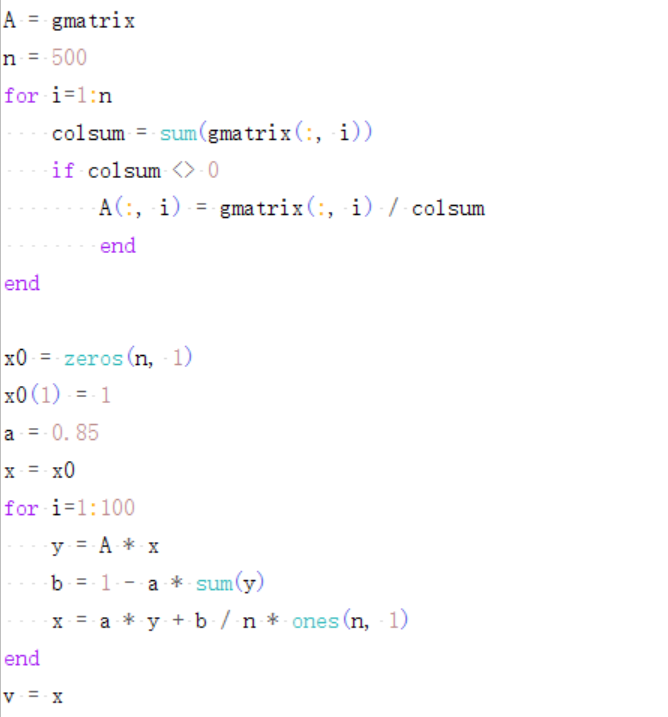
The part of the google matrix:



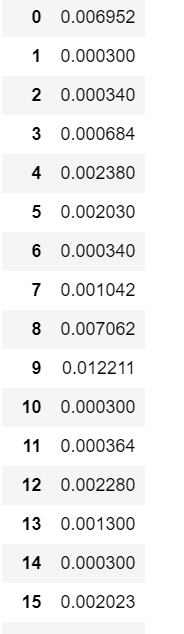
And here is its scatter plot:



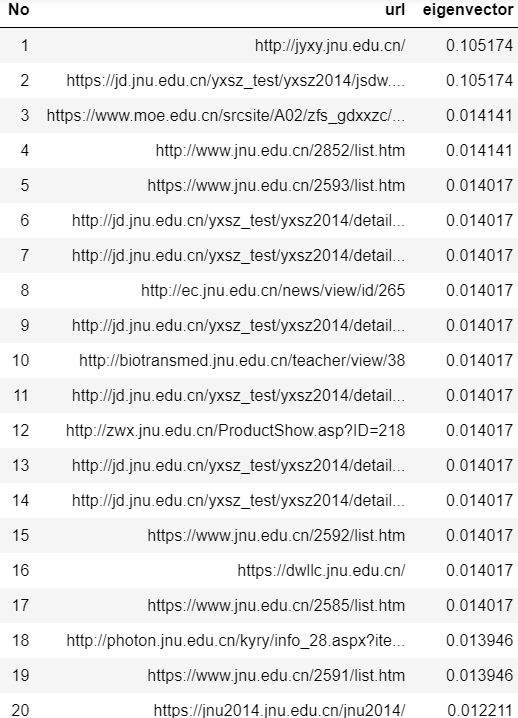
The Power Method for G:



Import the google matrix and run the Power Method for G, and we get the dominant eigenvector:



Sort the eigenvector on the descending order, and we get the top 20 web pages:



**Ⅴ、Experimental Summary**

In this experiment, we use the eigenvector to rank the relative importance of the website. As we can see, the most important website of our university is <http://jyxy.jnu.edu.cn/>. It is the official website of Education College of Jinan University. No matter whether the result is right or wrong, this method to measure the relative importance of each web page is worthwhile for us to learn from.