

# 第一次大作业: **Compute the PageRank scores on the given dataset**

Dataset: Data.txt

The format of the lines in the file is as follow:

**FromNodeID ToNodeID**

In this project, you need to report the **Top 100 NodeID** with their PageRank scores. You can choose different parameters, such as the teleport parameter, to compare different results.

One result you must report is that when setting the teleport parameter to **0.85**.

In addition to the basic **PageRank algorithm**, you need to implement the **Block-Stripe Update algorithm**.

Deadline: 2023.4.30

# 第一次大作业: Compute the PageRank scores on the given dataset

## 作业具体要求:

1. 语言: C/C++/JAVA/Python
2. 考虑dead ends 和spider trap 节点
3. 优化稀疏矩阵
4. 实现分块计算
5. 程序需要迭代至收敛
6. 不可直接调接口, 例如实现pagerank时, 调用Python的networkx包
7. 结果格式(.txt文件): [NodeID] [Score]

**截止日期: 2023年4月30日 23:59**

## 作业提交:

1. 作业发送至邮箱: bigdatacomputing@163.com
2. 实验报告内容 (包括但不限于): 数据集说明、关键代码细节、实验结果、结果分析等。
3. 程序源码。C, C++, JAVA, Python, 四者之一。无需提交数据集。
4. 程序执行结果文件。必须与实验报告中讨论的内容一致, 严格按照作业要求生成结果文件。格式不一致将无法进行正确率的计算, 影响最终成绩。
5. 可执行文件。请说明具体运行方式, 提交前请确认在其他电脑上也可以运行。C/C++ 编译选择release 方式  
进行静态编译 (debug方式生成的exe文件在其他电脑上运行可能会有缺少dll文件等问题)。JAVA 和 Python请借助第三方软件生成exe可执行文件, 并集成相关依赖包。
6. 上传文件命名:  
学号1+姓名1+学号2+姓名2 (同一组的同学上传相同的文件) 例: 00000\_李明\_11111\_王芳
7. 各小组独立完成, 请勿抄袭