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 optimize your memory use** via Block Matrix, Sparse Matrix and some other approches.
- Deadline: 2025.4.30

Compute the PageRank scores on the given dataset

Specific Requirement:

- Language: You can use C/C++/Python.
- For C/C++, It's recommended to code in a single file and minimize the use of third-party libraries. You'd better to use gcc or g++ for compiling your code and give the compilation parameters in *compile-parameter.txt* with the following format.

  g++ -fopenmp PageRank.cpp -o PageRank.exe

- For Python, It's recommended to use commonly-used libraries or package (e.g. numpy and scipy is accepted). To confirm the availability of a package, you can email *bigdatacomputing@163.com*. It's recommended to code in a single file. Ensure the entry of your code should be named as 'main.py'.

Compute the PageRank scores on the given dataset

Specific Requirement:

- Consider of dead-ends and spider-traps
- Optimize your memory use **as much as possible.**
- You can optimize the memory use in other ways**, but Block Matrix and Sparse Matrix(**https://zhuanlan.zhihu.com/p/557231877 **) optimization is compulsorily required**
- Your program is required to iterate until reaching convergence
- Using existing PageRank API is prohibited. (e.g. *networkx.pagerank* in Python)
- Your program need to read *Data.txt*, conduct PageRank and Return the Top-100 nodes in *Res.txt* with its score as following format:

  NodeID Score

Compute the PageRank scores on the given dataset

Specific Requirement:

- You'd better make sure the maximum memory use during the whole life of your program should be lower than **80MB.**
- You can't to sacrifice too much time performance to minimize your memory usage. The program need to complete its runtime under **60s**.
- Maybe you can use some Parallel or/and Distributed Computing techniques. **But it's not the main content of this assignment**.

Compute the PageRank scores on the given dataset

Submitting:

- *bigdatacomputing@163.com*
- *Report:* Include but not limited: description of dataset, key code details, **how to optimize the memory usage** and result with analysis
- *Code:* Meet the above requirements, Input Dataset is not required
- *Result:* Named as *Res.txt* and Meet the above requirements
- *Executable File:* Compile with all Static link library.
  - For C/C++, maybe you can use [-*static*] setting in gcc/g++
  - For Python, using third-party packages to generate and integrate the package used in your code.
  - **Ensure your code can run in other computers (and maybe other OS)**. (You can ask LLMs or TAs for more).

Compute the PageRank scores on the given dataset

Submitting:

- Indicate team division of work and contributions.
- If no team division indicated, whole group will be given the same score.
- Compress your submission (zip format), with following naming:

  2000000_小张_2111111_小王_2222222_小林_第一次作业.zip

- Deadline: 2025.4.30
- **Failure to comply with the format will result in failure to grade.**
- **Code plagiarism and academic fraud are absolutely intolerable.**