BLG202E Numerical Methods in Comp. Eng.

Spring 2024 - Term Project

Due: May 5, 2023

By turning in this assignment, I agree by the ITU honor code and declare that all of this is my own work.

Important Notes

- Upload your solutions through **Ninova**. Projects sent via e-mail and late submissions **will not be accepted**.
- Please make sure that you write your **full name** and student identification **number** to **every file** you submit.
- Cheating is highly discouraged. It will be punished by a negative grade. Also disciplinary actions will be taken. Please do your homework on your own. Team work is not allowed. Pattern of your **solutions must belong only to you**.
- All codes and reports will be run through **plagiarism check**s. Please **do not copy any text or code** from other sources.
- If you have any questions, please contact with **T.A. Yunus Emre Cebeci** (cebeci16@itu.edu.tr).
- Remember, there are only 10 types of people in the world – those who understand binary, and those who don't.

Project No 1: Implementation of the PageRank on the Cora dataset.

Please read the following prerequisites.

- Install a Conda environment if you do not have it already.
- Install Jupyter Notebook.
- The project should be done in Jupyter Notebook.
- Do not forget to format your code and leave comments for non-trivial sections.

Instructions.

- Implement the PageRank algorithm using the citations in the "cora.citations" file. In the coursebook (Example 8.3), the PageRank algorithm is implemented using links between web pages. In this project, if a paper cites another paper, it will be considered as if there is a link from the citing paper to the cited paper.
- In this project, the citations in the "cora.citations" file will be used. "cora.citations" file contains the citation graph of the corpus. Each line describes a link in the following format:
  <ID of cited paper> <ID of citing paper>
  Each line contains two paper IDs. The first entry is the ID of the paper being cited and the second ID stands for the paper which contains the citation. The direction of the link is from right to left. If a line is represented by "paper1 paper2" then the link is "paper2 -> paper1".
  - **Cora Dataset:**
    https://linqs-data.soe.ucsc.edu/public/lbc/cora.tgz
- Method Implementation and Experimental Setup:
  - You need to implement PageRank from scratch. You can't use built-in libraries or methods like 'numpy.linalg.svd', 'scipy.linalg.svd', 'sklearn.decomposition' etc.
  - You are expected to write code and get numerical results. Submissions without a working code will not be graded.
  - You should write the following information in the report.
    * To apply the PageRank algorithm to the Cora.cites dataset, how do you create the link matrix.
    * Have you encountered the "dangling node" problem? If so, how do you solve this problem?

    - Have you encountered the "dead end" problem? If so, how do you solve this problem (You can use the methods in the coursebook to solve these problems)?
    - Write the 20 biggest and 20 smallest entries of the PageRank vector.
    - Draw the graph of the 10 papers with the highest rank. In the graph, the nodes will be the papers and the arrows will be the connections between them. Don't forget to write the id and rank of each paper in the nodes. You can use built-in libraries like networkx to draw the graph.
    - If there are other significant challenges, please do not forget to write them down in the report.

- Report:
  - Write a maximum of 3 pages report using IEEE Latex Template. State the problem, implementation details, dataset and experiments.

- Submission
  - Submit a zip file that includes your Python code, plots, and results until the deadline through Ninova.
  - Upload your project report until the deadline through Ninova.