Madison Chester
Numerical Linear Algebra
Project 3
22 December 2023

Introduction:

The PageRank algorithm is a link analysis algorithm used by search engines to rank web pages in search results. It assigns a numerical weighting to each element of a hyperlinked set of documents, representing the likelihood of a user following the links at random. Exercise 1 entails extracting the dataset and computing PageRank on a given link matrix. Exercise 2 requires implementing a modified PageRank algorithm that avoids storing matrices. Finally, these two methods are compared.

Methods:

For Exercise 1, first the dataset is extracted from the compressed dataset file into a specified directory. Then, the compute_pagerank function computes the PageRank vector for the given link matrix using the power iteration method with a damping factor of m=0.15. It takes a link matrix, *G*, as input and allows for the determination of damping factor (*m*, *tol*, and *max_iter*). The damping factor represents the random behavior of the user. The lower the *m*, the more random clicks by the user. Dangling nodes, nodes with no outgoing links, are handled to ensure convergence. Sparse matrix operations are used.

For Exercise 2, matrix operations are performed without storage. For this, the build_diag function constructs a diagonal matrix *D* based on the out-degrees of nodes. The compute_PR_no_store function completes PageRank without explicitly storing matrices, which significantly reduces memory usage. It uses set operations to identify neighbors and iteratively updates PageRank scores.

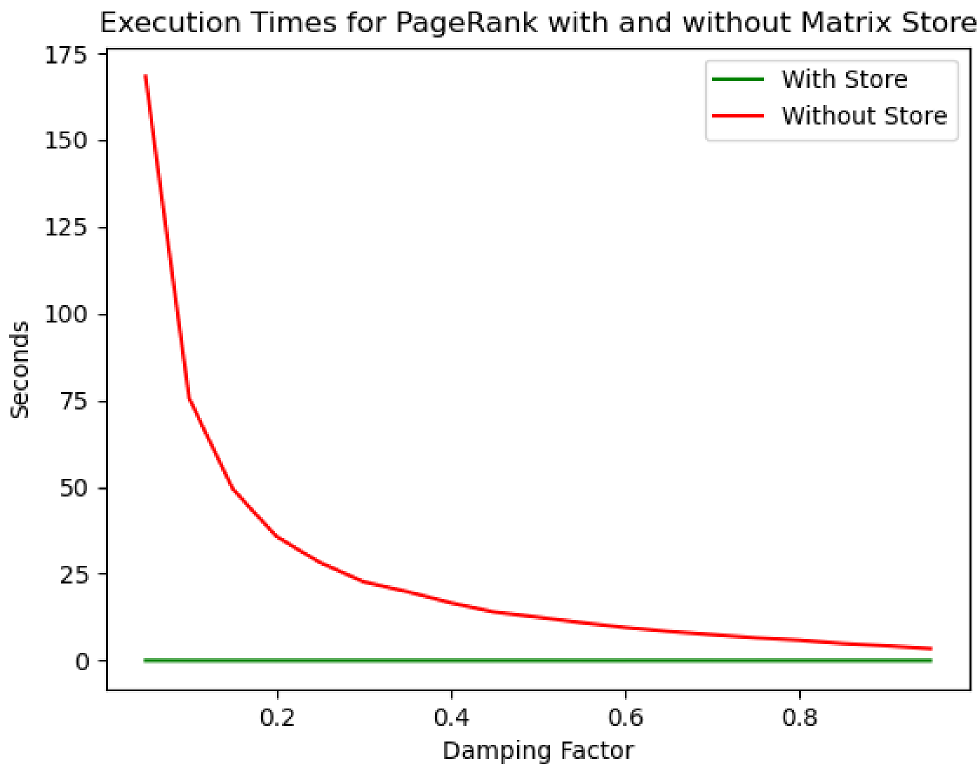
Outcomes:

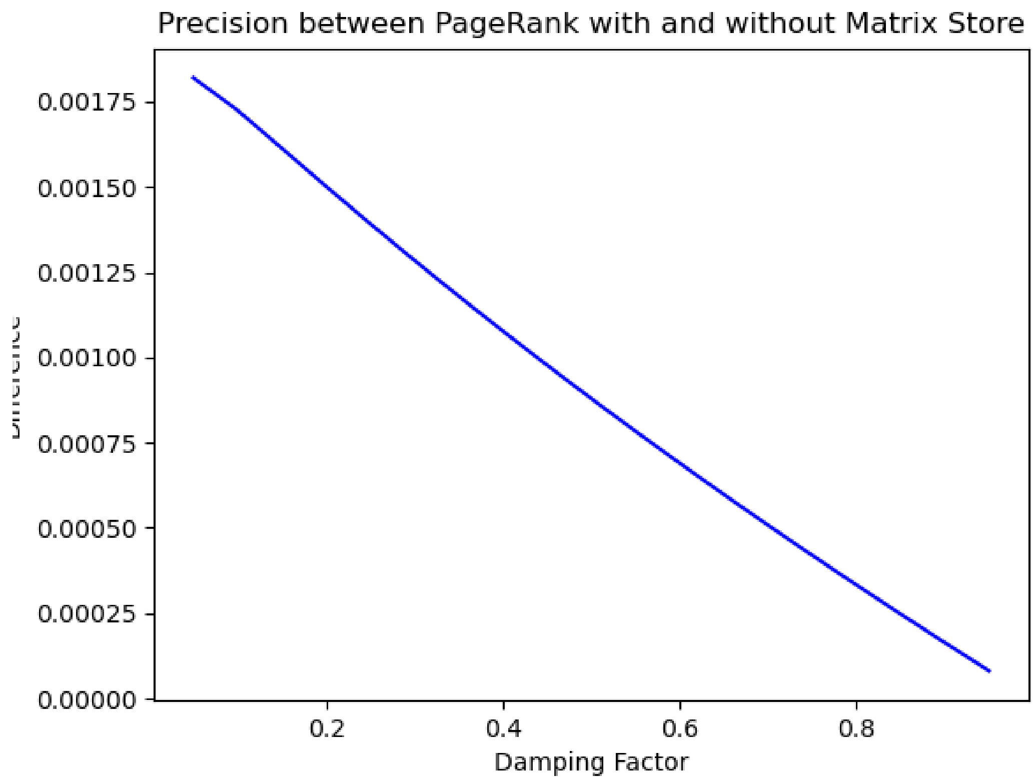
For Exercise 1, the dataset is successfully extracted, and the contents of the subdirectory are listed. The PageRank vector for the given link matrix is computed successfully, and the sum of PageRank scores is verified to be 1.

For Exercise 2, the modified PageRank algorithm without matrix storage produces a PageRank vector, and once again, the sum of PageRank scores is verified to be 1 (or very close to it).

Comparison:

Finally, the two methods are compared. The script iterates over different damping factors, *m*, and compares execution times and precision differences between the compute_pagerank and compute_PR_no_store functions. The comparison shows the tradeoff between execution time and memory consumption for different damping factors. Plots are also generated to visualize the results.





We can see that the execution time decreases as the damping factor increases in computations that do not store the matrices. Otherwise, computation times remain constant regardless of damping factor. Lastly, we can see that the difference in precision between storing and not storing the matrices in our algorithm decreases almost linearly with the damping factor. This makes sense as the randomness introduced in the algorithm is decreasing.

Conclusion:

The implementation of PageRank and its modified version demonstrates the versatility of the algorithm in handling large datasets with potential memory optimizations. The comparison provides valuable insights for choosing the appropriate damping factor based on memory storage limitations.

Appendix:

Code Output

```
p2p-Gnutella30.tar\p2p-Gnutella30\p2p-Gnutella30.mtx

PageRank Vector for Mm:
 [2.74354856e-05 2.72045201e-05 2.74077979e-05 ... 2.72045201e-05
 2.72045201e-05 2.72045201e-05]
Sum of PageRank scores: 1.0

PR vector without storing matrices:
 [9.00855109e-05 1.40978984e-05 3.32410694e-05 ... 1.40978984e-05
 1.40978984e-05 1.40978984e-05]
Sum of PageRank scores: 0.9999999999999997
Execution time: 57.10869646072388
```

```
Damping Factor: 0.05
- Execution time (with store): 168.30347657203674
- Precision (difference between store and no-store): 0.0018198748202384169
Damping Factor: 0.1
- Execution time (with store): 75.5927402973175
- Precision (difference between store and no-store): 0.0017243307310197567
Damping Factor: 0.15
- Execution time (with store): 49.454667806625366
- Precision (difference between store and no-store): 0.0016132971495340202
Damping Factor: 0.2
- Execution time (with store): 35.78685212135315
- Precision (difference between store and no-store): 0.0015016945158313184
Damping Factor: 0.25
- Execution time (with store): 28.238062620162964
- Precision (difference between store and no-store): 0.0013919795861273264
Damping Factor: 0.3
- Execution time (with store): 22.713937282562256
- Precision (difference between store and no-store): 0.0012846704010929097
Damping Factor: 0.35
- Execution time (with store): 19.83895492553711
- Precision (difference between store and no-store): 0.0011798447080598358
Damping Factor: 0.39999999999999997
- Execution time (with store): 16.606772661209106
- Precision (difference between store and no-store): 0.001077450081087822
Damping Factor: 0.44999999999999996
- Execution time (with store): 13.958806276321411
- Precision (difference between store and no-store): 0.00097739035159656
Damping Factor: 0.49999999999999994
- Execution time (with store): 12.497222661972046
- Precision (difference between store and no-store): 0.000879554847869697
Damping Factor: 0.5499999999999999
- Execution time (with store): 10.930372953414917
- Precision (difference between store and no-store): 0.0007838296321355671
```

```
Damping Factor: 0.6
- Execution time (with store): 9.514686584472656
- Precision (difference between store and no-store): 0.0006901021639498518
Damping Factor: 0.65
- Execution time (with store): 8.344971418380737
- Precision (difference between store and no-store): 0.0005982632851687499
Damping Factor: 0.7
- Execution time (with store): 7.4206626415252686
- Precision (difference between store and no-store): 0.0005082080211177124
Damping Factor: 0.75
- Execution time (with store): 6.5081493854522705
- Precision (difference between store and no-store): 0.0004198358269862103
Damping Factor: 0.7999999999999999
- Execution time (with store): 5.8334715366363525
- Precision (difference between store and no-store): 0.00033305056442216346
Damping Factor: 0.85
- Execution time (with store): 4.870642900466919
- Precision (difference between store and no-store): 0.0002477603451228419
Damping Factor: 0.9
- Execution time (with store): 4.2168872356414795
- Precision (difference between store and no-store): 0.00016387731019261884
Damping Factor: 0.95
- Execution time (with store): 3.4322378635406494
- Precision (difference between store and no-store): 8.131738093227073e-05
```