HOMEWORK No: 2

Page Rank Algorithm & the modified version

Course: Computational Linear Algebra 2023-2024

Baharak Qaderi s296876, Silva Bashllari s299317

Exercise 1:

Figure 1 shows the updated web after adding page 5 and its associated links.

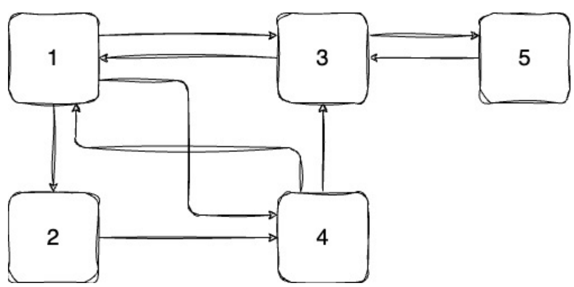

Figure 1: Adding Page 5

Let's begin by examining the initial matrix A_{initial}:

$$A_{\text{initial}} = \begin{bmatrix} 0 & 0 & 1 & \frac{1}{2} \\ \frac{1}{3} & 0 & 0 & 0 \\ \frac{1}{3} & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{3} & \frac{1}{2} & 0 & 0 \end{bmatrix}$$

The new link matrix A after adding page 5 is:

$$A = \begin{bmatrix} 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{3} & 0 & 0 & 0 & 0 \\ \frac{1}{3} & \frac{1}{2} & 0 & \frac{1}{2} & 1 \\ \frac{1}{3} & \frac{1}{2} & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{2} & 0 & 0 \end{bmatrix}$$

To rank the pages by their importance, we use the PageRank algorithm, which involves solving an eigenvector problem:

$$Ax = x$$

In this equation, x represents the vector of PageRank scores for each page, and A is the link matrix of the web. A higher score signifies a more important page.
To solve the eigenvector problem, we set $\lambda = 1$ and calculate this equation:

$$(A - I)x = \begin{bmatrix} -1 & 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{3} & -1 & 0 & 0 & 0 \\ \frac{1}{3} & \frac{1}{2} & -1 & \frac{1}{2} & 1 \\ \frac{1}{3} & \frac{1}{2} & 0 & -1 & 0 \\ 0 & 0 & \frac{1}{2} & 0 & -1 \end{bmatrix} x = 0$$

There are several methods for solving a system of equations. In the example above, we used two different methods, Gaussian elimination by hand and some online software to verify the results and obtained the following:

$$V_{\mathbf{1}}(A) = \text{span}\left\{ \begin{pmatrix} \frac{12}{49} \\ \frac{4}{49} \\ \frac{18}{49} \\ \frac{6}{49} \\ \frac{9}{49} \end{pmatrix} \right\}$$

Recall that any non-zero multiple of an eigenvector is again an eigenvector. Then comparing it to the original scoring:

$$V_1(A_{\text{initial}}) = \text{span}\left\{\begin{pmatrix} \frac{12}{31} \\ \frac{4}{31} \\ \frac{9}{31} \\ \frac{6}{31} \end{pmatrix}\right\}$$

From the vectors above, we observe that the creation of page 5 increased the PageRank score of page 3 from $\frac{9}{31}$ to $\frac{18}{49}$, so an increase 7 percent.
Additionally, the PageRank score of page 1 decreased from $\frac{12}{31}$ to $\frac{12}{49}$. This indicates that page 3 has become more important than page 1.Consequently, the owners of page 3 have successfully enhanced its importance by creating page 5 and linking it to page 3. They have been successful also due to the fact page 5 didn't have many outgoing links, because we have to remember this version of the page rank algorithm penlaizes links coming from pages that have many outgoing links.

Exercise 2:

For simplicity, we created the required subwebs with only two pages per each and create a link matrix A accordingly.

Example Web Structure:

- **Subweb 1:** Pages 1 and 2
- **Subweb 2:** Pages 3 and 4
- **Subweb 3:** Pages 5 and 6

Each subweb is fully connected internally, but not to other subwebs. This means pages in one subweb link to each other but not to pages in other subwebs. Figure 2 shows the setting.

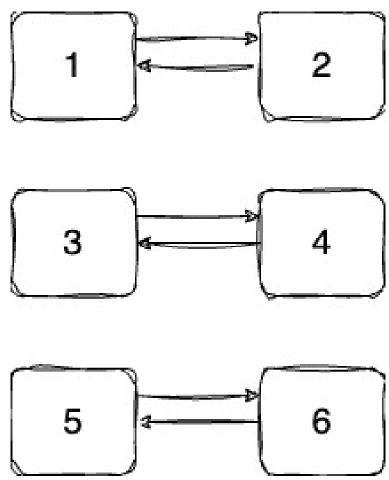

Figure 2: An example of a web with three subwebs

To find $V_1(A) = \text{span}\{\text{eigenvectors}\}$, we first need to solve the eigenvalue problem $Ax = x$ or equivalently $(A - I)x = 0$. This will give us the eigenvectors corresponding to the eigenvalue $\lambda = 1$.

Let's take the link matrix A:

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

First, subtract the identity matrix I from A:

$$A - I = \begin{bmatrix} -1 & 1 & 0 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 1 \\ 0 & 0 & 0 & 0 & 1 & -1 \end{bmatrix}$$

Now, solve $(A - I)x = 0$. The system is:

$$\begin{bmatrix} -1 & 1 & 0 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 1 \\ 0 & 0 & 0 & 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

From this system, we can find the relationships between the variables:

$$x_1 = x_2$$
$$x_3 = x_4$$
$$x_5 = x_6$$

Let:

$$x_1 = x_2 = a$$
$$x_3 = x_4 = b$$
$$x_5 = x_6 = c$$

The general solution vector x can be written as:

$$x = \begin{pmatrix} a \\ a \\ b \\ b \\ c \\ c \end{pmatrix} = a \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} + b \begin{pmatrix} 0 \\ 0 \\ 1 \\ 1 \\ 0 \\ 0 \end{pmatrix} + c \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 1 \end{pmatrix}$$

Thus, the eigenspace $V_1(A)$ corresponding to the eigenvalue $\lambda = 1$ is spanned by the vectors:

$$V_1(A) = \text{span} \left\{ \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 1 \end{pmatrix} \right\}$$

This shows that the dimension of $V_1(A)$ is 3, which corresponds to the number of subwebs (or connected components) in the web.

Exercise 3:

Given the link matrix A_{new} after adding a link from page 5 to page 1:

$$A_{\text{new}} = \begin{bmatrix} 0 & 0 & 1 & 0 & \frac{1}{3} \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & \frac{1}{3} \\ 0 & 0 & 1 & 0 & \frac{1}{3} \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

We subtract the identity matrix I from A_{new}:

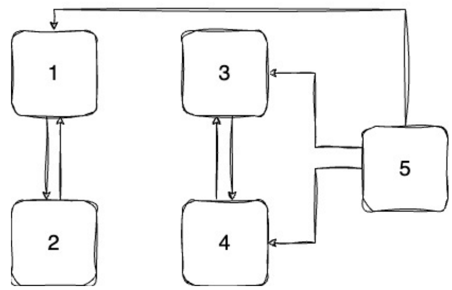

Figure 3: Updated the web with new link 5 to 1

$$A_{\text{new}} - I = \begin{bmatrix} -1 & 0 & 1 & 0 & \frac{1}{3} \\ 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & -1 & 1 & \frac{1}{3} \\ 0 & 0 & 1 & -1 & \frac{1}{3} \\ 0 & 0 & 0 & 0 & -1 \end{bmatrix}$$

Next, we solve the system $(A_{\text{new}} - I)x = 0$:

$$\begin{bmatrix} -1 & 0 & 1 & 0 & \frac{1}{3} \\ 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & -1 & 1 & \frac{1}{3} \\ 0 & 0 & 1 & -1 & \frac{1}{3} \\ 0 & 0 & 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

This system simplifies to the following relationships:

$$x_1 = x_2 = x_3 = x_4, \quad x_5 = 0$$

Thus, the general solution vector x is:

$$x = a \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} + b \begin{pmatrix} 0 \\ 0 \\ 1 \\ 1 \\ 0 \end{pmatrix}$$

The eigenspace $V_1(A_{\text{new}})$, corresponding to the eigenvalue $\lambda = 1$, is therefore:

$$V_1(A_{\text{new}}) = \text{span} \left\{ \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \\ 1 \\ 0 \end{pmatrix} \right\}$$

This indicates that the dimension of $V_1(A_{\text{new}})$ is **still** 2, meaning there are two linearly independent eigenvectors. Linking page 5 to page does not alter the dimension eigenspace

$V_1(A_{\text{new}})$ to the PageRank because it lacks backlinks. This means that if consider pages 5,3, and 4 as WEB1 and pages 1, and 2 as WEB2 there's a way to go from WEB1 to WEB2 but not the other way around.

Exercise 4:

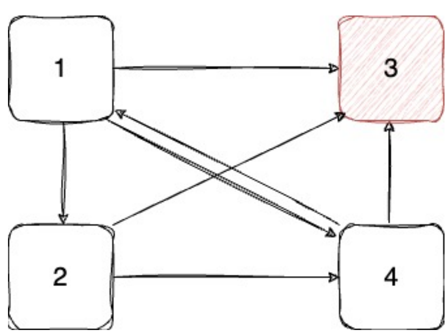

Figure 4: Updated web, making page 3 a a dangling node

Consider the initial link matrix:

$$A_{\text{initial}} = \begin{bmatrix} 0 & 0 & 1 & \frac{1}{2} \\ \frac{1}{3} & 0 & 0 & 0 \\ \frac{1}{3} & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{3} & \frac{1}{2} & 0 & 0 \end{bmatrix}$$

Removing the link from page 3 to page 1 results in:

$$A_{\text{new}} = \begin{bmatrix} 0 & 0 & 0 & \frac{1}{2} \\ \frac{1}{3} & 0 & 0 & 0 \\ \frac{1}{3} & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{3} & \frac{1}{2} & 0 & 0 \end{bmatrix}$$

Note that the third column is all zeros, reflecting that page 3 is now a dangling node.
To find the largest eigenvalue and the Perron eigenvector of A_{new}, we apply numerical methods (such as the power method). The results are:

- Largest Eigenvalue: $\lambda_{\max} = 0.5614$
- Corresponding Perron Eigenvector after normalization:

$$x = \begin{bmatrix} 0.2355 \\ 0.1399 \\ 0.3601 \\ 0.2645 \end{bmatrix}$$

The Perron eigenvector x is already normalized such that the sum of its components equals 1:

$$0.2355 + 0.1399 + 0.3601 + 0.2645 = 1.$$

The largest eigenvalue, $\lambda_{\max} = 0.5614$, is less than 1, which is expected for a substochastic matrix. The components of the Perron eigenvector provide the following importance scores for each page:

- Page 1: 0.2355
- Page 2: 0.1399
- **Page 3 (Dangling Node):** 0.3601
- Page 4: 0.2645

As expected, page 3 has the highest score since it has the most backlinks. Pages 1 and 4 have nearly identical scores, while page 2 has the lowest score. This ranking makes sense given the structure of the graph and the number of backlinks each page has.

Exercise 5:

Let $\mathbf{v}$ be the importance score vector for a web with n pages, and let B be the link matrix of the web. Suppose page k has no backlinks. We want to show that the importance score of page k is zero, i.e., $v_k = 0$.
Given that page k has no incoming links, the k-th row of B is zero:

$$B_{\cdot k} = \begin{bmatrix} 0 & 0 & \dots & 0 \end{bmatrix}$$

The importance score vector $\mathbf{v}$ must satisfy:

$$\mathbf{v} = B\mathbf{v}$$

Specifically, for page k:

$$v_k = \sum_{j \in L_k} \frac{v_j}{d_j}$$

where L_k denotes the set of pages linking to page k and d_j is the number of outgoing links from page j. Since $L_k = \emptyset$, we have:

$$v_k = 0$$

Thus, a page with no backlinks has an importance score of zero.

Exercise 6:

We consider $\tilde{A}$, the link matrix in which there is a new indexing of a given web whose original link matrix was A. In this new indexing, exactly two indexes have been swapped. For example, if the original web indexing was the following:

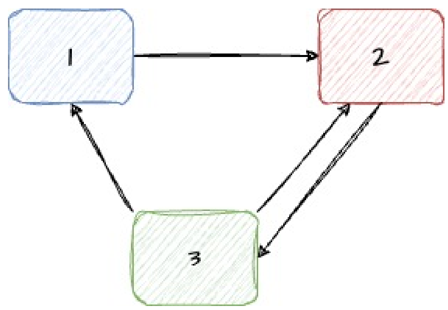

Figure 5: An example of a web with 3 pages

The web with the new indexing is the following:

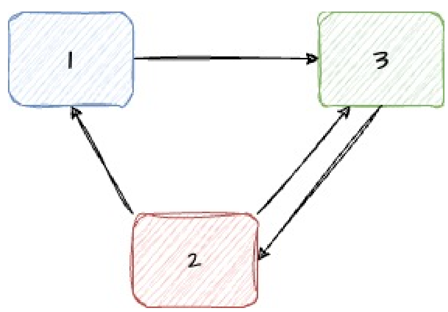

Figure 6: The web with 3 pages reindexed

We claim that $\tilde{A} = PAP$, where P is the elementary matrix obtained by swapping rows i and j of the $n \times n$ identity matrix I. For example, if we have the square three dimensional matrix its identity I_3

$$I_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

and P will be the following, assuming we swap the rows 3 and 2:

$$P = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$$

We also know that:

$$P^2 = I$$

We also know that: x is an eigenvector for A, so

$$Ax = \lambda x$$

We can also observe that:

- $A \to PA$: Multiplying A on the left by P swaps rows i and j of A.
- $A \to AP$: Multiplying A on the right by P swaps columns i and j of A.

We need to show that $\mathbf{y} = P\mathbf{x}$ is an eigenvector of $\tilde{A}$ with the same eigenvalue λ.

Consider the following:

$$\tilde{A}\mathbf{y} = (PAP)\mathbf{y} = PAP(P\mathbf{x}) = P^2A(P\mathbf{x}) = IA(P\mathbf{x}) = A(P\mathbf{x}) = P(A\mathbf{x}) = P(\lambda\mathbf{x}) = \lambda(P\mathbf{x}) = \lambda\mathbf{y}$$

So, we have:

$$\tilde{A}\mathbf{y} = \lambda\mathbf{y}$$

Thus, $\mathbf{y} = P\mathbf{x}$ is indeed an eigenvector of $\tilde{A}$ corresponding to the same eigenvalue λ.
Since $\mathbf{y} = P\mathbf{x}$ corresponds to the same eigenvalue λ as $\mathbf{x}$, transposing the indices of pages i and j does not affect the eigenvalues or the relative importance scores (the components of the eigenvector).
This argument can be generalized. If any permutation π of the page indices is applied, it can be represented as a product of such transpositions. Each transposition leaves the importance scores unchanged, so the overall permutation also leaves the importance scores unchanged. Hence, the indexing of pages has no effect on the importance scores assigned to any page in the web.

Exercise 7:

To prove that $M = (1-m)A + mS$ is a column-stochastic matrix, given that A is an $n \times n$ column-stochastic matrix and $0 \leq m \leq 1$:
By definition, for any column j of A,

$$\sum_{i=1}^{n} A_{ij} = 1.$$

The matrix S, where each entry $S_{ij} = \frac{1}{n}$, also satisfies

$$\sum_{i=1}^{n} S_{ij} = \sum_{i=1}^{n} \frac{1}{n} = 1.$$

Hence, S is column-stochastic.
The sum of the elements in column j of M is:

$$\sum_{i=1}^{n} M_{ij} = \sum_{i=1}^{n} \left[(1-m)A_{ij} + mS_{ij}\right].$$

Distributing the sum:

$$\sum_{i=1}^{n} M_{ij} = (1-m)\sum_{i=1}^{n} A_{ij} + m\sum_{i=1}^{n} S_{ij}.$$

Since both A and S are column-stochastic:

$$\sum_{i=1}^{n} M_{ij} = (1-m) \times 1 + m \times 1 = 1.$$

Since the sum equals 1 for all columns j, M is column-stochastic. This is an intuitive result given the fact that simply put we have two column stochastic matrices that are being scaled by scalars whose sum is 1 (1-m, m) and then added and will invariably produce a column stochastic matrix.

Exercise 8:

Let A and B be $n \times n$ column-stochastic matrices. We need to show that $C = AB$ is column-stochastic. By definition, for each column j,

$$\sum_{i=1}^{n} A_{ij} = 1 \quad \text{and} \quad \sum_{i=1}^{n} B_{ij} = 1$$

Every single element of C can be described by

$$C_{ij} = \sum_{k=1}^{n} A_{ik} B_{kj}$$

The sum of the elements in column j of C is

$$\sum_{i=1}^{n} C_{ij} = \sum_{i=1}^{n} \sum_{k=1}^{n} A_{ik} B_{kj} = \sum_{k=1}^{n} B_{kj} \sum_{i=1}^{n} A_{ik}$$

B_{kj} does not depend on i. So, you can factor B_{kj} out of the inner sum. Since A is column-stochastic, $\sum_{i=1}^{n} A_{ik} = 1$, and since B is column-stochastic, $\sum_{k=1}^{n} B_{kj} = 1$. Therefore, $\sum_{i=1}^{n} C_{ij} = 1$, so C is column-stochastic.

Exercise 9:

Formula (3.2) is given by:

$$x = (1 - m)Ax + ms,$$

where A is the link matrix, m is a scalar such that $0 \leq m \leq 1$, and s is a vector where each entry is $\frac{1}{n}$.

Consider a page i with no backlinks to it, meaning the i-th row of A consists entirely of zeros. Therefore, the i-th entry of the importance score vector x is:

$$x_i = (1 - m)\sum_{j=1}^{n} A_{ij}x_j + ms_i.$$

Since $A_{ij} = 0$ for all j (no backlinks),

$$x_i = (1 - m) \times 0 + m \times \frac{1}{n} = \frac{m}{n}.$$

Thus, the importance score for a page with no backlinks is $\frac{m}{n}$.

Exercise 10:

In General terms, we assume that $(A^k)_{ij}$ means that page i can be reached from page j in k step, if and and only $(A^k)_{ij} > 0$.
then we can claim that $(A^2)_{ij} > 0$ if and only if page i can be reached from page j in exactly two steps.
This means that the entry $(A^2)_{ij}$ represents the number of ways to go from page j to page i in exactly two steps.
To compute $(A^2)_{ij}$, we have:

$$(A^2)_{ij} = \sum_k A_{ik} A_{kj}$$

This expression is a sum over all possible intermediate pages k.
So we can state that:

- If $(A^2)_{ij} > 0$, then there exists at least one intermediate page k such that both $A_{ik} > 0$ and $A_{kj} > 0$. This means there's a link from j to k and a link from k to i, establishing a path of exactly two steps.
- Conversely, if page i can be reached from page j in exactly two steps, there exists some intermediate page k such that $A_{kj} > 0$ and $A_{ik} > 0$. Therefore, $(A^2)_{ij} > 0$ due to the presence of this path.

To generalize this to A^p:

$$(A^p)_{ij} > 0$$

if and only if page i can be reached from page j in exactly p steps. We already showed it holds for $p = 2$.

Now consider

$$A^{p+1} = A \cdot A^p.$$

The entry

$$(A^{p+1})_{ij} = \sum_k A_{ik} (A^p)_{kj}$$

By the inductive hypothesis,

$$(A^p)_{kj} > 0$$

if there is a path from j to k in exactly p steps, and multiplying by $A_{ik} > 0$ implies a path of $p + 1$ steps from j to i. Thus, $(A^{p+1})_{ij} > 0$ if and only if there's a path from j to i in exactly $p + 1$ steps.
Then we can claim:

$$(I + A + A^2 + \cdots + A^p)_{ij} > 0$$

if and only if page i can be reached from page j in p or fewer steps.
This sum matrix considers all possible path lengths from 0 to p. We should keep in mind:

- The diagonal matrix I represents 0 steps (a page can always reach itself with 0 steps).

- Each matrix A^k (for $k \geq 1$) represents paths of exactly k steps.
- If $(I + A + \cdots + A^p)_{ij} > 0$, there is at least one k such that $(A^k)_{ij} > 0$, meaning there is a path from j to i of length k where $0 \leq k \leq p$.

Another thing we have to consider is the matrix

$$I + A + A^2 + \cdots + A^{n-1}$$

is positive if the web is strongly connected.
Because:

- In a strongly connected web, any page i can be reached from any page j within $n - 1$ steps (since the web has n pages).
- Therefore, $(I + A + A^2 + \cdots + A^{n-1})_{ij} > 0$ for all i, j since any page i can be reached from page j in at most $n - 1$ steps.

Now consider the matrix B:

$$B = \frac{1}{n}(I + A + A^2 + \cdots + A^{n-1})$$

positive and column-stochastic.

- Since $I + A + A^2 + \cdots + A^{n-1}$ is positive, B is also positive.
- Since $I + A + \cdots + A^{n-1}$ captures all paths and ensures that any page contributes its probability to every other page through normalization by n, each column sum becomes 1.

By Lemma 3.2, since B is a positive column-stochastic matrix, we have $\dim(V_1(B)) = 1$

Consider:

$$x \in V_1(A)$$

Then

$$x \in V_1(B)$$

This implies

$$\dim(V_1(A)) = 1$$

So:

- If $x \in V_1(A)$, then $Ax = x$. When multiplying x by B:

$$Bx = \frac{1}{n}(I + A + A^2 + \cdots + A^{n-1})x$$

- Since x is in $V_1(A)$, $Ax = x$, $A^2x = A(Ax) = Ax = x$, and similarly $A^kx = x$ for all $k \geq 0$.

- Therefore:
$$Bx = \frac{1}{n}(I + A + A^2 + \cdots + A^{n-1})x = \frac{1}{n}(nx) = x$$
- This shows that $x \in V_1(B)$.

Since $\dim(V_1(B)) = 1$ and any vector in $V_1(A)$ is also in $V_1(B)$, it follows that $\dim(V_1(A))$ must also be 1.

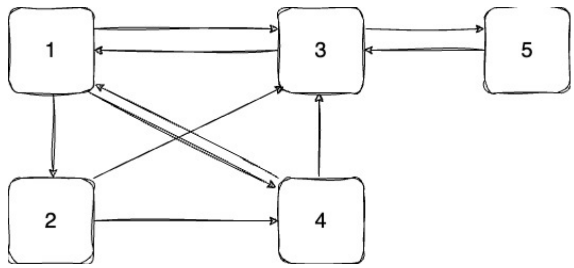

Figure 7: Adding Page 5

Exercise 11:

Given the link matrix A and $m = 0.15$,

$$A = \begin{bmatrix} 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{3} & 0 & 0 & 0 & 0 \\ \frac{1}{3} & \frac{1}{2} & 0 & \frac{1}{2} & 1 \\ \frac{1}{3} & \frac{1}{2} & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{2} & 0 & 0 \end{bmatrix}$$

we calculate M as follows:

$$0.85A = \begin{bmatrix} 0 & 0 & 0.425 & 0.425 & 0 \\ 0.283 & 0 & 0 & 0 & 0 \\ 0.283 & 0.425 & 0 & 0.425 & 0.85 \\ 0.283 & 0.425 & 0 & 0 & 0 \\ 0 & 0 & 0.425 & 0 & 0 \end{bmatrix}$$

$$0.15S = \begin{bmatrix} 0.03 & 0.03 & 0.03 & 0.03 & 0.03 \\ 0.03 & 0.03 & 0.03 & 0.03 & 0.03 \\ 0.03 & 0.03 & 0.03 & 0.03 & 0.03 \\ 0.03 & 0.03 & 0.03 & 0.03 & 0.03 \\ 0.03 & 0.03 & 0.03 & 0.03 & 0.03 \end{bmatrix}$$

$$M = 0.85A + 0.15S = \begin{bmatrix} 0.03 & 0.03 & 0.455 & 0.455 & 0.03 \\ 0.313 & 0.03 & 0.03 & 0.03 & 0.03 \\ 0.313 & 0.455 & 0.03 & 0.455 & 0.88 \\ 0.313 & 0.455 & 0.03 & 0.03 & 0.03 \\ 0.03 & 0.03 & 0.455 & 0.03 & 0.03 \end{bmatrix}$$

To find the eigenvector corresponding to $\lambda = 1$, solve:

$$(M - I)x = 0$$

Normalize the eigenvector x so that its components sum to 1 (as denoted in Example 2 of Point 3.1 Modification in the PageRank paper). Solving the equation using python(Numpy),the normalized principal eigenvector $\mathbf{x}$ is given by:

$$\mathbf{x} = \begin{bmatrix} 0.2371 \\ 0.0972 \\ 0.3489 \\ 0.1385 \\ 0.1783 \end{bmatrix}$$

According to this new ranking, the importance of the pages is as follows:

1. Page 3 (importance score: 0.3489)
2. Page 1 (importance score: 0.2371)
3. Page 5 (importance score: 0.1783)
4. Page 4 (importance score: 0.1385)
5. Page 2 (importance score: 0.0972)

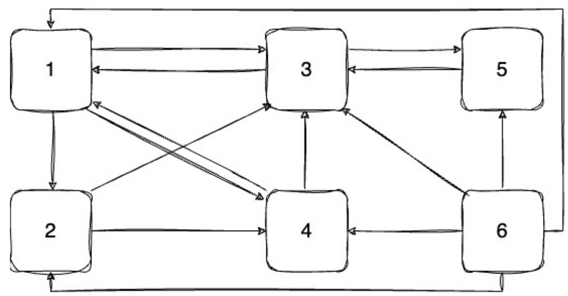

Figure 8: Adding Page 6, and links it to all the pages

Exercise 12:

We start with the updated link matrix A for 6 pages:

$$A = \begin{bmatrix} 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 & \frac{1}{5} \\ \frac{1}{3} & 0 & 0 & 0 & 0 & \frac{1}{5} \\ \frac{1}{3} & \frac{1}{2} & 0 & \frac{1}{2} & 1 & \frac{1}{5} \\ \frac{1}{3} & \frac{1}{2} & 0 & 0 & 0 & \frac{1}{5} \\ 0 & 0 & \frac{1}{2} & 0 & 0 & \frac{1}{5} \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

The principal eigenvector for the matrix A is:

$$\mathbf{x}_A = \begin{bmatrix} 0.2449 \\ 0.0816 \\ 0.3673 \\ 0.1224 \\ 0.1837 \\ 0 \end{bmatrix}$$

The matrix M is defined as:

$$M = (1 - 0.15)A + 0.15 \cdot S$$

where S is the matrix with all entries $\frac{1}{6}$.
The principal eigenvector for the matrix M is:

$$\mathbf{x}_M = \begin{bmatrix} 0.2312 \\ 0.0948 \\ 0.3402 \\ 0.1350 \\ 0.1738 \\ 0.0250 \end{bmatrix}$$

Observation:
Rankings based on matrices A and M are very close numerically in terms of importance scores for pages 1-5 (inclusive) , with the exception of page 6. As it can be observed, for the

eigenvector produced by A page 6 has a score of 0, whereas for the eigenvector produced by M, page 6 has a small allbeit above 0 value. The ranking is:

1. Page 3
2. Page 1
3. Page 5
4. Page 4
5. Page 2
6. Page 6

- Original Matrix A: Page 6 has zero importance, reflecting its lack of incoming links.
- Modified Matrix M: Page 6 now has a non-zero importance score due to the inclusion of the teleportation factor $m = 0.15$, which redistributes importance more evenly.

Exercise 13:

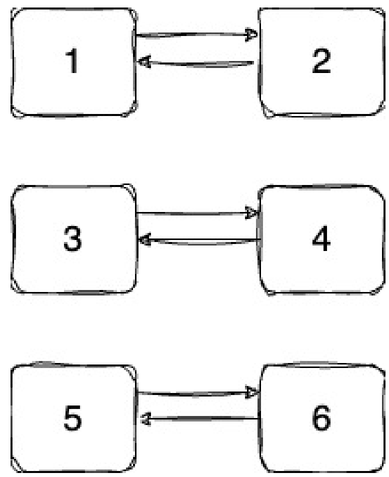

Figure 9: An example of a web with three subwebs

Given a web consisting of three subwebs where each subweb is fully connected internally but not connected to other subwebs, we are to determine the ranking using the modified matrix M defined by:

$$M = (1 - m)A + mS$$

The matrix A:

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

The uniform matrix S for a 6-page web is:

$$S = \frac{1}{6} \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}$$

Using $m = 0.15$, the modified matrix M is calculated as:

$$M = 0.85A + 0.15S$$

Substituting the matrices A and S into the formula:

$$M = 0.85 \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix} + 0.15 \begin{bmatrix} \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \\ \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \\ \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \\ \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \\ \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \\ \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \end{bmatrix}$$

Calculating each element of M, we get:

$$M = \begin{bmatrix} 0.025 & 0.875 & 0.025 & 0.025 & 0.025 & 0.025 \\ 0.875 & 0.025 & 0.025 & 0.025 & 0.025 & 0.025 \\ 0.025 & 0.025 & 0.025 & 0.875 & 0.025 & 0.025 \\ 0.025 & 0.025 & 0.875 & 0.025 & 0.025 & 0.025 \\ 0.025 & 0.025 & 0.025 & 0.025 & 0.025 & 0.875 \\ 0.025 & 0.025 & 0.025 & 0.025 & 0.875 & 0.025 \end{bmatrix}$$

To determine the page rankings, we compute the principal eigenvector (the eigenvector corresponding to the eigenvalue $\lambda = 1$) of matrix M.(We used python, Numpy to calculate it)
The principal eigenvector, normalized to sum to 1, is:

$$\mathbf{x} : \begin{bmatrix} 0.1667 \\ 0.1667 \\ 0.1667 \\ 0.1667 \\ 0.1667 \\ 0.1667 \end{bmatrix}$$

This result was expected as you can see all pages in this web are equally important, with each page having an equal PageRank of approximately 0.1667. This result is consistent also with the one we observe in Exercise 2, given that we have taken the same web-structure but computed the eigenvectors derived from matrix A. Were we to linearly combine the 3 eigenvectors produced, we would get a single one that also attributes the same importance score to all the pages involved. However, that score would be of 1, whereas using the modified version of link matrix, we get a different scalar.

Exercise 14:

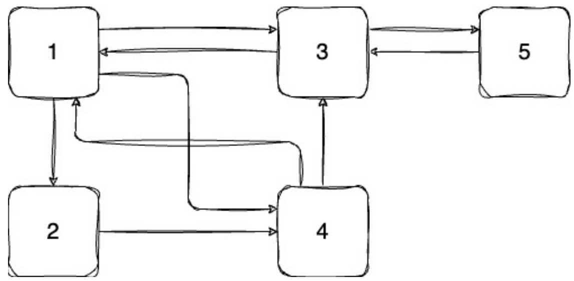

Figure 10: A web with five pages

For the web structure described in Exercise 11, we aim to compute the values of $\|M^k x_0 - q\|_1$ and the error ratio

$$\frac{\|M^k x_0 - q\|_1}{\|M^{k-1} x_0 - q\|_1}$$

for $k = 1, 5, 10, 50$, using the modified matrix M and an initial guess x_0. We will also determine the constant c and the absolute value of the second largest eigenvalue of M.
The link matrix A for the web is:

$$A = \begin{bmatrix} 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{3} & 0 & 0 & 0 & 0 \\ \frac{1}{3} & \frac{1}{2} & 0 & \frac{1}{2} & 1 \\ \frac{1}{3} & \frac{1}{2} & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{2} & 0 & 0 \end{bmatrix}$$

We modify A by adding a weighted average of the matrix S with all entries equal to $\frac{1}{5}$ as follows:

$$M = (1 - m)A + mS$$

where $m = 0.15$. This results in:

$$M = \begin{bmatrix} 0.03 & 0.03 & 0.455 & 0.455 & 0.03 \\ 0.3133 & 0.03 & 0.03 & 0.03 & 0.03 \\ 0.3133 & 0.455 & 0.03 & 0.455 & 0.88 \\ 0.3133 & 0.455 & 0.03 & 0.03 & 0.03 \\ 0.03 & 0.03 & 0.455 & 0.03 & 0.03 \end{bmatrix}$$

The initial guess is:

$$x_0 = \begin{bmatrix} 0.4 \\ 0.1 \\ 0.1 \\ 0.3 \\ 0.1 \end{bmatrix}$$

and the dominant eigenvector q is:

$$q = \begin{bmatrix} 0.23 \\ 0.10 \\ 0.35 \\ 0.14 \\ 0.18 \end{bmatrix}$$

The errors $\|M^k x_0 - q\|_1$ and the error ratios are shown in Table 2.

Table 1: Error and Error Ratio for Different Values of k

k	$\|M^k x_0 - q\|_1$	Error Ratio $\frac{\|M^k x_0 - q\|_1}{\|M^{k-1} x_0 - q\|_1}$
0	0.6543	N/A (No previous norm)
1	0.2858	0.4367
5	0.0352	0.6380
10	0.0029	0.6245
50	8.116018327315733e-05	1.0000001418756281

We aslo tried with a different value of initial guess: 0.24, 0.31, 0.08, 0.18, 0.19

$$x_0 = \begin{bmatrix} 0.24 \\ 0.31 \\ 0.08 \\ 0.18 \\ 0.19 \end{bmatrix}$$

As you can see, the results are consistent:

Table 2: Error and Error Ratio for Different Values of k

k	$\|M^k x_0 - q\|_1$	Error Ratio $\frac{\|M^k x_0 - q\|_1}{\|M^{k-1} x_0 - q\|_1}$
0	0.5378	N/A (No previous norm)
1	0.4217	0.78430
5	0.0496	0.5618
10	0.0042	0.62394
50	8.116018510999357e-05	1.0000002015342655

To find the constant c, we use:

$$c = \max_{1 \le j \le n} \left| 1 - 2 \min_{1 \le i \le n} M_{ij} \right|$$

For matrix M, the minimum entry of each column is 0.03. Therefore:

$$c = |1 - 2 \times 0.03| = |1 - 0.06| = 0.94$$

The error ratio converges to approximately 0.62, which represents the absolute value of the second largest eigenvalue $|\lambda_2|$ of M.
The power method shows rapid convergence to the dominant eigenvector q, with the error decreasing significantly as k increases. The error ratio converges to the absolute value of the second largest eigenvalue, which is approximately 0.62. The constant c is computed as 0.94, indicating the convergence rate of the power method.

Exercise 15:

Given that M is diagonalizable, we can express M as:

$$M = PDP^{-1},$$

where D is a diagonal matrix containing the eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ of M, and P is a matrix whose columns are the corresponding eigenvectors $q, v_1, \ldots, v_{n-1}$. The eigenvector q

corresponds to the eigenvalue $\lambda_1 = 1$ and represents the steady-state vector of M.
Let x_0 be any vector with non-negative components that sum to one. We can express x_0 as a linear combination of the eigenvectors of M:

$$x_0 = aq + \sum_{k=1}^{n-1} b_k v_k,$$

where a and b_k are coefficients to be determined. Since x_0 and q are probability vectors (i.e., their components sum to one), it follows that:

$$a = 1.$$

Moreover, because the sum of the components of x_0 is 1, and v_k are eigenvectors corresponding to the eigenvalues $\lambda_k \neq 1$, the sum of the components of each v_k must be zero:

$$\sum_{i=1}^{n} (v_k)_i = 0 \quad \text{for } k \geq 2.$$

Applying the diagonalization of M, we find:

$$M^k x_0 = M^k \left(q + \sum_{k=1}^{n-1} b_k v_k \right).$$

Since $M^k q = q$ (because $Mq = q$ and $\lambda_1 = 1$), and $M^k v_k = \lambda_k^k v_k$, we obtain:

$$M^k x_0 = q + \sum_{k=1}^{n-1} b_k \lambda_k^k v_k.$$

Since M is a positive column-stochastic matrix, Proposition 4 tells us that for all $k \geq 2$, we have $|\lambda_k| < 1$. Therefore, as $k \to \infty$, $\lambda_k^k \to 0$ for all $k \neq 1$. Consequently, $M^k x_0$ converges to q as $k \to \infty$. Given that x_0 has non-negative components that sum to one, we must have:

$$\sum_{k=1}^{n-1} b_k v_k = 0.$$

Thus, the convergence is given by:

$$\lim_{k\to\infty} M^k x_0 = q + \sum_{k=1}^{n-1} b_k \lambda_k^k v_k.$$

To analyze the convergence rate, consider:

$$\lim_{k\to\infty} \frac{\|M^k x_0 - q\|_1}{\|M^{k-1} x_0 - q\|_1}.$$

For large k, the term with the largest eigenvalue less than 1 dominates:

$$M^k x_0 - q \leq b_2 \lambda_2^k v_2.$$

Hence,

$$\|M^k x_0 - q\|_1 \leq |b_2||\lambda_2|^k \|v_2\|_1,$$

and similarly,

$$\|M^{k-1} x_0 - q\|_1 \leq |b_2||\lambda_2|^{k-1} \|v_2\|_1.$$

Thus, the ratio is:

$$\frac{\|M^k x_0 - q\|_1}{\|M^{k-1} x_0 - q\|_1} \leq \frac{|b_2||\lambda_2|^k \|v_2\|_1}{|b_2||\lambda_2|^{k-1} \|v_2\|_1} = |\lambda_2|.$$

Therefore, we conclude that:

$$\lim_{k\to\infty} \frac{\|M^k x_0 - q\|_1}{\|M^{k-1} x_0 - q\|_1} = |\lambda_2|.$$

This result shows that the second-largest eigenvalue in magnitude $|\lambda_2|$ dictates the convergence rate of the power iterations $M^k x_0$ to the steady-state vector q. The closer $|\lambda_2|$ is to 1, the slower the convergence; conversely, the smaller $|\lambda_2|$, the faster the convergence.

Exercise 16:

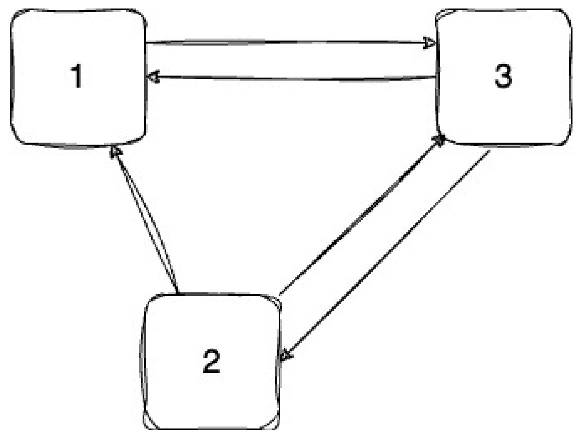

Figure 11: A 3-page graph

We are given the link matrix:

$$A = \begin{bmatrix} 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & \frac{1}{2} \\ 1 & \frac{1}{2} & 0 \end{bmatrix}$$

and the matrix S with all entries equal to $\frac{1}{3}$:

$$S = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{bmatrix}.$$

We want to show that the modified matrix

$$M = (1-m)A + mS,$$

where $0 \leq m < 1$, is not diagonalizable.
First, let's compute the matrix M:

$$M = (1-m)A + mS = \begin{bmatrix} \frac{m}{3} & \frac{3-m}{6} & \frac{3-m}{6} \\ \frac{m}{3} & \frac{m}{3} & \frac{3-m}{6} \\ \frac{3-2m}{3} & \frac{3-m}{6} & \frac{m}{3} \end{bmatrix}.$$

To find the eigenvalues of M, we solve the characteristic equation:

$$\det(M - \lambda I) = 0.$$

This determinant is given by:

$$\det(M - \lambda I) = \begin{vmatrix} \frac{m}{3} - \lambda & \frac{3-m}{6} & \frac{3-m}{6} \\ \frac{m}{3} & \frac{m}{3} - \lambda & \frac{3-m}{6} \\ \frac{3-2m}{3} & \frac{3-m}{6} & \frac{m}{3} - \lambda \end{vmatrix}.$$

Simplifying, we find:

$$\det(M - \lambda I) = -(\lambda - 1)^2(\lambda - \frac{m-1}{2}).$$

Thus, the eigenvalues of M are:

$$\lambda_1 = 1 \quad , \quad \lambda_2 = \frac{m-1}{2}$$

With algebraic multiplicities

$$m(\lambda_1) = 2 \quad , \quad m(\lambda_2) = 1 \quad .$$

Next, we find the eigenvectors and corresponding eigenspaces.

For $\lambda_1 = 1$, we solve:

$$(M - I)v_1 = 0.$$

The resulting eigenspace V_{λ_1} is spanned by:

$$V_{\lambda_1} = \text{span}\left\{\begin{bmatrix} \frac{m-3}{2m-4} \\ \frac{-1}{m-2} \\ 1 \end{bmatrix}\right\}.$$

The dimension of this eigenspace is 1, which is less than the algebraic multiplicity of $\lambda_1 = 1$. Therefore, M is not diagonalizable at this stage since a matrix is diagonalizable if and only if the algebraic and geometric multiplicities of each eigenvalue are equal.

For $\lambda_2 = \frac{m-1}{2}$, we solve:

$$(M - (\frac{m-1}{2})I)v_2 = 0.$$

The resulting eigenspace V_{λ_2} is spanned by:

$$V_{\lambda_2} = \text{span}\left\{\begin{bmatrix} 0 \\ -1 \\ 1 \end{bmatrix}\right\}.$$

The dimension of this eigenspace is 1, which matches its algebraic multiplicity.
Since the geometric multiplicity of $\lambda_1 = 1$ is less than its algebraic multiplicity, the matrix M is not diagonalizable for $0 \leq m < 1$.

Exercise 17:

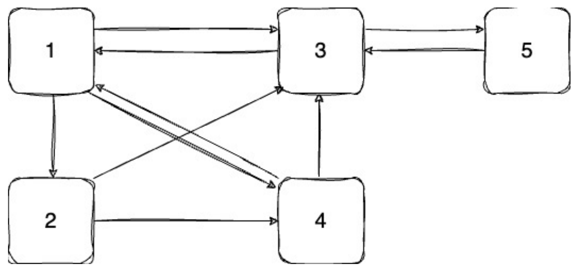

Figure 12: An web example with 5 pages

The parameter m in the modified PageRank matrix

$$M = (1 - m)A + mS,$$

where A is the link matrix and S is a stochastic matrix with all entries equal, plays a critical role in determining both the ranking results and the computational efficiency of the PageRank algorithm.
The value of m affects the balance between the influence of the web graph's link structure and the random teleportation component. When m is closer to 1, the matrix M emphasizes the random jumps represented by S, leading to more uniform PageRank scores across pages. Conversely, when m is closer to 0, the matrix M is more influenced by the link structure of A, resulting in rankings that favor pages with more incoming links, thus reflecting the web's actual connectivity more closely.
In terms of computation time, a higher m results in a matrix M that is closer to a stochastic matrix with a uniform distribution, which generally improves the convergence rate of iterative methods used to compute the PageRank vector. This leads to faster convergence and reduced computation time. On the other hand, a lower m makes M more dominated by the link structure matrix A, which can slow down convergence due to increased sparsity and irregularity in the matrix.
Choosing the optimal m involves balancing accuracy in reflecting the link structure with efficiency in computation. A commonly chosen value is $m = 0.15$, which balances these aspects well by ensuring a reasonable convergence rate while still accounting for the web's link structure. However, the optimal value of m may vary depending on the specific characteristics of the web graph and the objectives of the ranking algorithm, requiring empirical testing to determine the best choice.

We consider the web shown in Figure 12 and calculate the ranking using 15 random values for m to see how the rankings change. The results are plotted in Figure 13, where we observe the following points:

Observation: Effect of different m values

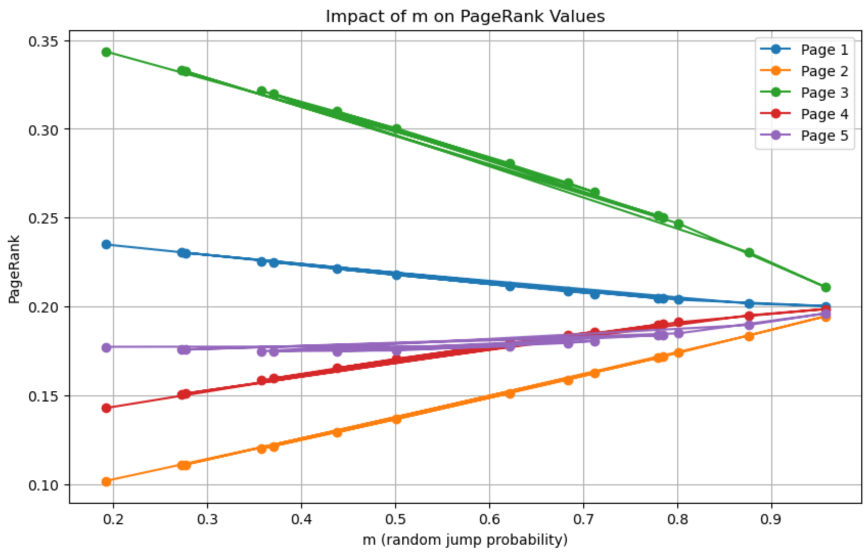

Figure 13: impact of m on the ranking

Low m Values (Closer to 0): When m is small, the matrix M gives more weight to the link structure of the web graph (matrix A). This means that the rank of each page is primarily influenced by the number and quality of incoming links. Pages with more incoming links or links from highly ranked pages have a higher PageRank. In our case page 3.

High m Values (Closer to 1): When m is large, the matrix M gives more weight to the random jump matrix S. This results in a more uniform distribution, where the PageRank is less influenced by the link structure and more by the uniform probability of jumping to any page. The differences in PageRank between pages become smaller. This is clearly illustrated in Figure 13.