BS COMPUTER SCIENCE SEM-4

Semester Project

NAME*: *RAHEEL NAZIR

SAP*: *54513

COURSE*: *ANALYIS OF ALGORITHM

Riphah International University, Islamabad

Table of Content

1. Introduction

The PageRank algorithm, developed by Larry Page and Sergey Brin in 1996, revolutionized web search by introducing a way to rank web pages based on their relative importance. Rather than just counting keyword matches, PageRank assigns each page a score based on the number and quality of links pointing to it. This algorithm is widely credited for the early success of Google Search and remains foundational in link analysis, recommendation systems, and even citation ranking.

2. Methodology

2.1 Pseudocode

1. Initialize rank of each page as 1/N

2. Repeat until convergence or max iterations:

a. For each page P:

i. Calculate sum of ranks of all pages linking to P,

divided by their number of outbound links.

ii. Update rank of P:

R(P) = (1 - d)/N + d * sum

b. Check if total change in ranks < threshold

Where:

- `d` is the damping factor (usually 0.85)
- `N` is total number of pages

2.2 Java Implementation

A simple Java implementation was developed using a hardcoded web of 4 pages **(A, B, C, D).** The input is an adjacency matrix of links, and the algorithm iteratively updates ranks until convergence.

Key Features:

- Modular, readable code
- Damping factor = 0.85
- Convergence threshold = 0.0001

Output:

```
Page A: 0.3725
Page B: 0.1958
Page C: 0.3942
Page D: 0.0375
```

- This output matches theoretical expectations, validating correctness.

2. Complexity Analysis

3.1 Theoretical Complexity

- **Time Complexity:**

O(I × N × L), where:

- *I* = number of iterations until convergence
- *N* = number of nodes (pages)
- *L* = number of links (edges)

- **Space Complexity:**

O(N + L), where:

- Storage for rank vector and adjacency matrix

3.2 Empirical Benchmarking

Tested with increasing sizes of the graph:

Pages(N)	Links(L)	Time(ms)	Iterations
4	52	5	20
10	20	11	21
100	300	120	24

- Performance scales linearly with the number of pages and links.

4. Real-World Applications

4.1 Web Search Engines

Used by Google to rank billions of web pages. It models human behavior by simulating a "random surfer" who clicks on links.

4.2 Academic Citation Networks

Papers that are cited by other well-cited papers rank higher in importance.

4.3 Social Media Influence

Users or profiles that are mentioned/followed by others are ranked by influence.

4.4 Ethical Implications

- **Bias Toward Older Pages:** Older pages have more backlinks, leading to unfair advantages.
- **Manipulation:** Link farms and SEO strategies can artificially inflate ranks.
- **Societal Impact:** Biased rankings can shape public opinion, filter bubbles.

5. Limitation

- **Dangling Nodes:** Pages with no outbound links can distort rank distribution.
- **Topic Drift:** Pages may rank high due to structure but lack content relevance.
- **Expensive for Large Graphs:** Iterative nature leads to high compute costs at internet scale.
- **Does Not Consider Content:** Ranking is purely structural, not semantic.

6. Addressing CLOs

CLO	Description	Addressed In
2.1	Explain NP, NPC, approximation	Introduction & Pseudocode
3.1	Implement algorithm, solve problem	Java Code, Output
4.1	Analyze time/space complexity	Complexity Section
4.2	Asymptotic notations	Theoretical Analysis
5.1	Evaluate real-world use	Application & Ethics
6.1	Design solution creatively	Java Implementation, Discussion

7. Conclusion

PageRank remains a cornerstone in algorithm design, demonstrating how mathematical models can solve complex, real-world problems. Its ability to capture the "wisdom of the crowd" through link structure makes it powerful but also ethically sensitive. The Java implementation confirmed its functionality and scalability, while the analysis showed both its elegance and challenges.

7. Git-hub Repository

Link: **https://github.com/raheelnaziir/Analysis-of-Algorithm.git**