



DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING (ARTIFICIAL INTELLIGENCE & MACHINE LEARNING)

PageRank Algorithm

PageRank is an algorithm developed by Google to rank web pages in search results based on their importance, which is determined by the number and quality of links to the page. The idea is that a page linked by many important pages is itself considered important.

Working:

Initialization: Every page starts with an equal rank.

Calculation: The PageRank of a page is calculated by summing the contributions from pages linking to it, weighted by their own PageRanks and the number of links they have.

Iteration: This process is repeated until the PageRanks stabilize.

Example:

Consider 4 web pages: A, B, C, and D.

Links

A → B, C

B → C, D

C → A

D → C

Initial PageRank: $PR(A)=PR(B)=PR(C)=PR(D)=0.25$

Iteration Example:

Page A:

$$PR(A)=0.0375+0.85\times PR(C)/1=0.0375+0.2125=0.25$$

Page B: $PR(B)=0.0375+0.85\times PR(A)/2=0.14375$

Page C: $PR(C)=0.4625$ (calculated considering links from A, B, and D).

Page D: $PR(D)=0$.

After several iterations, Page C has the highest PageRank due to more valuable inbound links.



DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING (ARTIFICIAL INTELLIGENCE & MACHINE LEARNING)

Advantages of PageRank:

1. **Quality over Quantity:** It values the quality of links over sheer quantity, meaning a link from a highly ranked page is more valuable.
2. **Resilience to Manipulation:** The algorithm is less prone to manipulation by link farms or spammy tactics because it takes into account the importance of linking pages.
3. **Simple and Efficient:** PageRank can be implemented and computed relatively efficiently, even for large datasets.
4. **Global Influence:** It considers the global link structure, giving a comprehensive view of a page's importance within the entire web.

Disadvantages of PageRank:

1. **Rank Sinks:** Pages with many outbound links can dilute their influence, leading to lower PageRank for linked pages.
2. **Damping Factor Issue:** The damping factor d is an arbitrary constant and might not reflect real-world linking behavior accurately.
3. **Doesn't Consider Content:** PageRank relies solely on link structure, ignoring the actual content quality of the pages.
4. **Susceptible to Link Manipulation:** While resilient, it can still be manipulated through complex linking schemes, such as link pyramids.
5. **Outdated for Modern Web:** The web has evolved, and the PageRank algorithm doesn't account for modern ranking signals like social media, user engagement, or content freshness.

Applications of the Page-Rank algorithm

The PageRank algorithm has found several applications beyond its original use in ranking web pages. Some notable applications include Ranking in search engines:

1. **search engines:** It helps determine the importance and importance of web pages based on the website's link structure. Search engines such as Google include PageRank as one of many factors to rank search results and provide users with more accurate and helpful information.
2. **Recommender systems:** PageRank can recommend relevant items to users based on their preferences and similarity. Applying an algorithm to a network of objects and analyzing their relationships can identify essential and influential objects that may interest the user.
3. **Social Network Analysis:** PageRank analyzes social networks to identify influential individuals or network nodes. The algorithm can classify users based on their connections and network influence by treating individuals as nodes and connections as links. This information can be valuable in various areas, such as marketing, identifying key opinion leaders, or understanding the spread of information.
4. **Citation analysis:** In academic research, the PageRank algorithm can be applied to Analyze citation networks. The algorithm can identify influential articles or researchers in a given field by treating academic articles as nodes and



DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING (ARTIFICIAL INTELLIGENCE & MACHINE LEARNING)

citations as links. This information helps to understand the impact and importance of scientific work

5. Content Recommendation: PageRank can recommend related or similar content on a website or platform. By analyzing the link structure between different pages or articles, the algorithm can identify related pages and recommend them to users as related or recommended

6. Fraud detection: PageRank can be used in fraud detection systems to identify suspicious fraud patterns or By analyzing connections between entities, such as financial transactions or network communications, the algorithm can flag potentially fraudulent nodes or transactions based on their impact on the network.