

Trabalho 1

Vinicius B. Guerreiro 1

Carlos Thiago 2

cc24488@g.unicamp.br

cc24459@g.unicamp.br

27 de março de 2025

1 Introdução

O algoritmo PageRank é uma técnica usada para classificar páginas da web com base na estrutura de links entre elas. O objetivo do algoritmo é priorizar páginas relevantes, avaliando os links que apontam para elas. Neste trabalho, foram implementadas duas abordagens principais para calcular o PageRank: uma abordagem iterativa, onde a pontuação de uma página é recalculada a cada iteração, e uma abordagem de amostragem de páginas, que estima o PageRank por meio de uma técnica de amostragem aleatória.

2 Objetivo

O objetivo principal deste trabalho foi implementar as duas abordagens do algoritmo PageRank e compará-las em precisão e eficiência. As abordagens envolvem:

- **PageRank Iterativo:** A pontuação de cada página é recalculada

com base nas pontuações das páginas que a linkam, e o processo de atualização é repetido até que as pontuações se estabilizem, utilizando toda a estrutura de links.

- **PageRank por Amostragem de Páginas:** Usa uma técnica de amostragem aleatória para estimar o PageRank. A ideia central é simular um passeio aleatório pelas páginas e usar a probabilidade de estar em uma página específica após vários passos para estimar seu PageRank, levando apenas o estado atual para calcular o próximo.

3 Dificuldades Encontradas

Durante a implementação das funções, dificuldades foram encontradas.

3.1 Como cada fórmula é usada em cada função

As fórmulas do PageRank envolvem a distribuição de valores de uma página para outras páginas, com base nos links que ela contém.

Entender como essas fórmulas se aplicam no código foi desafiador, especialmente ao implementá-la de forma iterativa. A dificuldade foi em saber como usar corretamente os valores anteriores de PageRank para recalculá-los os novos valores. Foi necessário estudar e revisar a fórmula para garantir que ela fosse implementada corretamente dentro da função `iterate_pagerank`.

3.2 Ferramentas usadas para desempenhar cada papel

Durante a implementação, uma das dificuldades foi identificar quais ferramentas e funções seriam mais adequadas para desempenhar cada parte do trabalho. Por exemplo, a função `random.choices` foi essencial na implementação do PageRank por amostragem de páginas, pois ela permite realizar uma distribuição com pesos. Vimos na documentação do Python que essa função poderia ser utilizada para amostrar páginas com base em probabilidades definidas pela distribuição de probabilidade.

No início, a função `random.choices` não era uma escolha óbvia, pois a tarefa envolvia uma amostragem ponderada, o que não era claro. Após revisar a documentação do Python, percebi que essa função poderia ser usada para selecionar páginas de forma aleatória, considerando as probabilidades definidas pela função de transição.

3.3 Distribuição dos valores

A distribuição dos valores de PageRank foi uma das tarefas mais complicadas, pois exige que use a fórmula matemática aplicada no código, juntamente com o conceito de navegação pelas páginas, tudo de uma só vez.

4 Soluções e Abordagens Adotadas

Para resolver as dificuldades mencionadas, algumas soluções e abordagens foram adotadas:

4.1 Estudo das fórmulas de PageRank

Foi fundamental compreender totalmente as fórmulas utilizadas no algoritmo, especialmente a fórmula iterativa. Para isso, revisões teóricas sobre o PageRank foram feitas, garantindo que a implementação seguisse a lógica matemática correta.

4.2 Pesquisa sobre ferramentas adequadas

Em especial, a pesquisa sobre a função `random.choices` e o uso de dicionários para armazenar os valores de PageRank foi fundamental para o desempenho do algoritmo.

4.3 Normalização e verificação de convergência

Para que os valores fossem distribuídos corretamente, a normalização das visitas foi feita ao final de cada iteração. O código foi ajustado para verificar a convergência com base em uma tolerância, garantindo que o processo fosse interrompido quando os valores estivessem estáveis o suficiente.

5 Comparação entre as abordagens

5.1 PageRank Iterativo

Esta abordagem demonstrou uma precisão maior, uma vez que calcula os valores de PageRank com base na estrutura completa de links. No entanto, ela é mais lenta, pois é necessário várias iterações até a convergência.

5.2 PageRank por Amostragem de Páginas

Embora esta abordagem seja mais rápida, pois utiliza amostragem aleatória, ela apresentou uma precisão um pouco menor em comparação com a abordagem iterativa. Para alcançar um resultado mais preciso, é necessário um número maior de amostras.

6 Conclusão

A implementação do algoritmo PageRank utilizando duas abordagens foi concluída com sucesso. O maior desafio esteve na compreensão das fórmulas usadas no algoritmo e na escolha das ferramentas adequadas. A abordagem iterativa mostrou ser mais precisa, enquanto a abordagem de amostragem de páginas foi mais rápida, mas com um custo em precisão.