



DEPARTAMENTO
DE COMPUTACION

Facultad de Ciencias Exactas y Naturales - UBA

TP 1: PageRank

September 5, 2022

Métodos Numéricos

Grupo 18

Integrante	LU	Correo electrónico
Vekselman, Natán	338/21	natanvek11@gmail.com
Arienti, Federico	316/21	fa.arianti@gmail.com
Barcos, Juan Cruz	463/20	juancruzbarcos@hotmail.com



Facultad de Ciencias Exactas y Naturales

Universidad de Buenos Aires

Ciudad Universitaria - (Pabellón I/Planta Baja)

Intendente Güiraldes 2610 - C1428EGA

Ciudad Autónoma de Buenos Aires - Rep. Argentina

Tel/Fax: (+54 +11) 4576-3300

<http://www.exactas.uba.ar>

RESUMEN

El Ranking de Page, *PageRank* [1], es un método propuesto por Sergey Brin y Larry Page —co-fundadores de Google—, para establecer la importancia de una página web dentro del internet, o dentro de un subconjunto de las páginas que lo componen. Holísticamente, considera que un sitio va a ser más relevante si es más probable que un usuario lo acceda. Por ello, el ranking define para cada par de páginas a y b la probabilidad condicional de que un navegante acceda a a , dado que actualmente se encuentre en b , y busca resolver un sistema de ecuaciones lineal donde la i -ésima ecuación representa la *importancia* de la i -ésima página: la esperanza probabilística de que el usuario la acceda.

Este trabajo propone una implementación eficiente de PageRank a través del uso de distintas representaciones de matriz (acorde a la cantidad de valores no nulos), y el empleo de iteradores específicos, para reducir el costo espacial y temporal de la eliminación gaussiana, método utilizado para la resolución del sistema.

Se buscará dar una presentación teórica y una evaluación cuantitativa y cualitativa de los resultados del método propuesto.

Palabras clave: *Ranking de Page, Eliminación Gaussiana, Matrices ralas*

CONTENIDOS

1. Introducción Teórica	2
1.1. Aridad	2
1.2. Matriz de probabilidad	2
2. Desarrollo	3
3. Resultados y Discusión	4
4. Conclusiones	5
Referencias	6
5. Apéndice	7

1. INTRODUCCIÓN TEÓRICA

1.1. Aridad.

Consideremos primero el dominio y la imagen del problema. Tenemos un conjunto de páginas web conectadas unas a otras por hipervínculos. Como única condición vamos a ignorar las auto-referencias. Si nos abstraemos, podemos considerar que nuestro dominio es el conjunto de todo grafo direccionado sin auto-direccionamiento (los nodos son los sitios y los ejes, los links). En consecuencia, podemos proponer la siguiente representación:

$$(1) \quad W_n = \{w_n \mid w_n \in \{0, 1\}^{n \times n}, (w_n)_{ij} = \begin{cases} 1 & \text{si } i \neq j \wedge j \xrightarrow{l} i \\ 0 & \text{sino} \end{cases} \quad \forall i, j : 1 \dots n \}$$

donde la notación $j \xrightarrow{l} i$ representa un link de la página j a la página i , y las filas y columnas de w_n , denominada *matriz de conectividad*, representan, indexadas por posición, las páginas del grafo.

PageRank define, además, un parámetro de entrada $p \in (0, 1)$, que representa la probabilidad de que un usuario decida navegar aleatoriamente a otra página en el grafo. Notar que p se puede interpretar como el parámetro de un variable aleatoria bernoulli, y que no está definida para sus extremos.

La imagen, por su parte, es un vector $\vec{s} \in [0, 1]^n$, donde \vec{s}_i representa el Ranking de Page para la i -ésima página del grafo de entrada. Los valores están normalizados, tal que:

$$\sum_{i=0}^n \vec{s}_i = 1$$

Tenemos entonces:

$$(2) \quad \text{PageRank} : W_n \times (0, 1) \longrightarrow [0, 1]^n \quad \forall n \in \mathbb{N}$$

1.2. Matriz de probabilidad.

La resolución de PageRank requiere resolver un sistema de ecuaciones lineales sobre una matriz de probabilidades. Esta se define de la siguiente manera:

2. DESARROLLO

3. RESULTADOS Y DISCUSIÓN

4. CONCLUSIONES

REFERENCIAS

- [1] Sergey Brin and Lawrence Page. *The anatomy of a large-scale hypertextual Web search engine*, volume Computer networks and ISDN systems, 30 (1-7):107–117. Stanford, 1998.

5. APÉNDICE