



DEPARTAMENTO
DE COMPUTACION

Facultad de Ciencias Exactas y Naturales - UBA

Ranking web

11 de octubre de 2014

Métodos Numéricos
Trabajo Práctico Nro. 2

Integrante	LU	Correo electrónico
Martin Carreiro	45/10	martin301290@gmail.com
Kevin Kujawski	459/10	kevinkuja@gmail.com
Juan Manuel Ortíz de Zárate	403/10	jmanuoz@gmail.com



Facultad de Ciencias Exactas y Naturales
Universidad de Buenos Aires

Ciudad Universitaria - (Pabellón I/Planta Baja)

Intendente Güiraldes 2160 - C1428EGA

Ciudad Autónoma de Buenos Aires - Rep. Argentina

Tel/Fax: (54 11) 4576-3359

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

Índice

1. Resumen	3
2. Introducción teórica	4
2.1. Matriz Dispersa	4
2.2. DOK vs CRS vs CSC	4
3. Desarrollo	5
3.1. Page Rank	5
3.2. HITS	5
3.3. Indeg	5
4. Experimentación Y Resultados	6
4.1. Casos de prueba	6
4.2. Comparación de Normas	6
4.3. PageRank	6
4.4. Comparación de Tiempos	6
5. Discusión	7
6. Conclusiones	8

1. Resumen

2. Introducción teórica

2.1. Matriz Dispersa

Se define una matrix dispersa aquella a la que la mayoría de sus elementos son cero.

$$\begin{bmatrix} 0 & 0 & 0 & 0 & a_{04} \\ 0 & a_{11} & a_{12} & 0 & 0 \\ 0 & 0 & 0 & a_{23} & 0 \\ 0 & 0 & 0 & a_{33} & 0 \\ a_{40} & 0 & 0 & 0 & 0 \end{bmatrix}$$

2.2. DOK vs CRS vs CSC

La matriz dispersa al tener la propiedad de tener muy pocos valores no—cero es conveniente solo guardar estos y asumir el resto como cero. Existen varias estructuras como Dictionary of Keys (dok), Compressed Sparse Row (CSR) o Compressed Sparse Column (CSC). En el desarrollo de este TP, utilizamos DOK por facilidad en el uso del mismo. Tanto CSR o CSC se basan en la estructura Yale y se diferencian en como guardan los mismos valores, uno priorizando las columnas y otro las filas respectivamente.

La estructura Yale consiste en a partir de la matriz original obtener tres vectores que contengan

- A = los elementos no—cero de arriba-abajo,izquierda-derecha
- IA = los indices para cada fila i del primer elemento no-cero de dicha fila
- JA = los indices de columna para cada valor de A

Si bien en caso de que haya en una fila muchos números no-ceros es más beneficioso la utilización de esta estructura, la facilidad con DOK permite hacer pruebas más rápido.

3. Desarrollo

3.1. Page Rank

3.2. HITS

3.3. Indeg

Algorithm 1 Indeg

```
1: result  $\leftarrow []$ ;  
2: result  $\leftarrow \text{initializeWithZeros}(\text{sizeOfLinksMatrix}[0]); \text{linksReferencedForAPageinlinksMatrix}$   
   linkinlinksReferencedForAPage result[link]+ = 1;
```

Algorithm 2 randomKill()

```
1: randomRemoveFrom(allLeaches); ▷ All Leaches is loaded with matrix  
2: recalculateByBandMatrix();
```

4. Experimentación Y Resultados

4.1. Casos de prueba

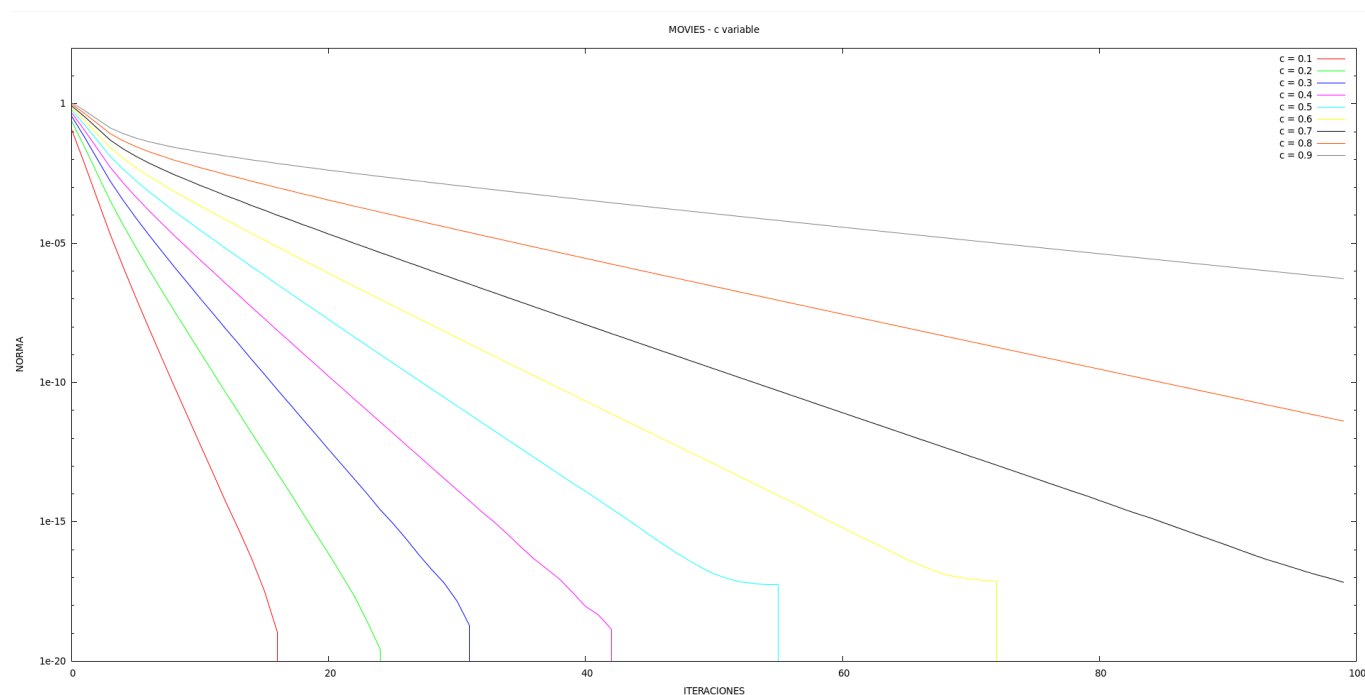
A continuación se listarán los casos utilizados y después se compararán los resultados.

- MOVIES: Este caso incluye 5797 páginas
- ABORTION: Este caso incluye 2293 páginas
- GENETIC: Este caso incluye 3468 páginas
- STANFORD: Este caso incluye 281903 páginas

4.2. Comparación de Normas

4.3. PageRank

Para evaluar el comportamiento de la norma manhattan variando la probabilidad del navegante aleatorio (a.k.a c)



Claramente podemos notar que a medida que el C crece, el algoritmo toma más iteraciones en achicar la norma. Esto se debe a que el grado de aleatoriedad elimina el peso de la unión entre los sitios e indica una uniformidad en el comportamiento, entonces la matriz si bien estocástica ahora se encuentra distribuida esa suma = 1 por columna en varias filas. Esto produce mayor cantidad de iteraciones en el método de la potencia[1]

4.4. Comparación de Tiempos

5. Discusión

6. Conclusiones

Referencias

- [1] [http : //personales.upv.es/ pedroche/inv_docs/fpedrochev4\(sema\).pdf](http://personales.upv.es/pedroche/inv_docs/fpedrochev4(sema).pdf)