

Calcular el pagerank de la Stanford Web Matrix

Los **objetivos** de esta práctica son: calcular el **pagerank de Stanford Web Matrix**, compuesta de 281903 páginas y 2.3 millones de links, y buscar en la red otras matrices web similares.

1. Calcular el pagerank de Stanford Web Matrix.

1. Bajar de AulaVirtual, descomprimir y guardar en el directorio de Matlab la Stanford Web Matrix.

2. Cargar la matriz A con los siguientes comandos (que no se os olvide el ;):

```
load stanford-web.dat;
A=spconvert(stanford_web); % make the matrix square
n=max(size(A)); A(n,n)=0; A=A';
```

3. Con el comando spy(A) obtenéis la gráfica de la dispersión de A.

4. Dar los comandos necesarios para conocer las características de la matriz A.

- Tipo de matriz (dispersa, completa).
- ¿Cuántos nodos tiene la red?.
- Tamaño en memoria de A.
- Número de elementos no nulos de A.
- La matriz A ¿qué tipo de matriz es de las descritas en las transparencias (C, A, S o G)? Indicar por qué.
- Calcular el índice de dispersión de A (número de elementos no nulos/número total de elementos).
- Calcular el número medio de outlinks (links de salida).
- Calcular los siguientes vectores:

```
Nj % Nj(k)=1 si la página k tiene nodos de salida, Nj(k)=0 en otro caso.
j_0 % Indices del vector Nj cuyo contenido es 0 (nodos sin salida)
j_1 % Indices del vector Nj cuyo contenido es 1 (nodos con salida)
```

- ¿Cuántos nodos sin salida tiene la red? ¿Cuántos nodos con salida tiene la red?.
- Comprobar que se verifica la siguiente relación:

número de páginas/nodos totales = número de nodos sin salida + número de nodos con salida.

5. Calcular el pagerank de Stanford Web Matrix.

Utilizar la función `[x,precision]=calculo_PR(A)` para calcular el pagerank de la Stanford Web Matrix, con una precisión del orden de la precisión de la máquina ($1e-13$). Completar la siguiente tabla.

	Tiempo [seg]	Memoria [MB]	Nº iteraciones	Precisión $\ Gx - x\ _1$
N=281x10 ³				

6. Visualizar y ordenar los resultados.

- Visualizar el pagerank obtenido con el comando `bar(x)`.
- Utilizar el comando `sort` para ordenar los elementos del vector de pagerank x.
- ¿Cuáles son los nodos que tienen los 20 mayores pagerank? Visualizarlos con el comando `bar`. Completar la siguiente tabla con esos 20 nodos ordenados de mayor a menor:

Orden	Nodo	Pagerank
1		
2		
.		
.		
20		

- ¿Cuáles son los nodos que tienen los 20 menores pagerank?. ¿Son iguales todos los pagerank? ¿Por qué?. Justificar la respuesta.

2. Buscar otras matrices web en la red

- Utilizar el buscador Google para encontrar otras matrices web, similares a la Stanford Web Matrix.
- La matriz debe ser del mayor tamaño posible.
- Indicar los criterios de búsqueda que habéis utilizado, y su localización.
- Descargar la matriz en Matlab.
- Calcular el pagerank de la matriz.
- Completar de nuevo la tabla anterior con los 20 nodos, ordenados de mayor a menor, que tengan el mayor pagerank.