

Aplicaciones numéricas de la Informática

Práctica 3: Calcular el pagerank de matrices de gran tamaño

Rubén Ibáñez Redondo

Mario Martín López

1. Calcular e interpretar el pagerank de Stanford Web Matrix

2. Cargar la matriz A

```
>> load stanford-web.dat;
```

```
>> A=spconvert(stanford_web);
```

```
N=size(A,1);A(N,N)=0;A=A';
```

```
spy(A);title('Gráfica de la dispersión de la matriz A')
```

```
>> Nj=sum(A);nnz(A)%Para ver si la matriz es dispersa o no. Contamos el numero de los elementos distintos de 0
```

```
ans =
```

```
2312497
```

```
>> spy(A);title('Gráfica de la dispersión de la matriz A')
```

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

La matriz es cuadrada es de 281903X281903. Tiene 281903 nodos.

La matriz es dispersa porque de los 281903X281903 elementos solo 2312497 son distintos de cero.

Volcado de datos: whos

A	281903x281903	39255184 double sparse
---	---------------	------------------------

N	1x1	8 double
---	-----	----------

Nj	1x281903	6762928 double sparse
----	----------	-----------------------

ans	1x1	8 double
-----	-----	----------

stanford_web	2382912x3	57189888 double
--------------	-----------	-----------------

La matriz A en memoria ocupa 39255184 Bytes.

Solo 2312497 son distintos de cero.

La matriz que nos dan es una matriz de tipo A debido a que como máximo sus elementos de Nj son 1 pero tiene algunos que son 0, por lo que no puede ser una matriz S, y además es dispersa.

```
>> dispersionA = (nnz(A)/ (281903*281903))
```

```
dispersionA =
```

```
2.9099e-05
```

```
>> r = sum(Nj) / length(Nj)
```

```
r =
```

```
(1,1)    0.9994
```

Cada página tiene de media 0.9994 links de salida

```
>> NodosSinSalida = length(find(~Nj))
```

```
NodosSinSalida =
```

```
172
```

```
>> NodosConSalida = nnz(Nj)
```

```
NodosConSalida =
```

```
281731
```

```
>> NodosTotales= NodosConSalida + NodosSinSalida
```

```
NodosTotales =
```

```
281903
```

Que es igual a:

```
>> length(A)
```

```
ans =
```

```
281903
```

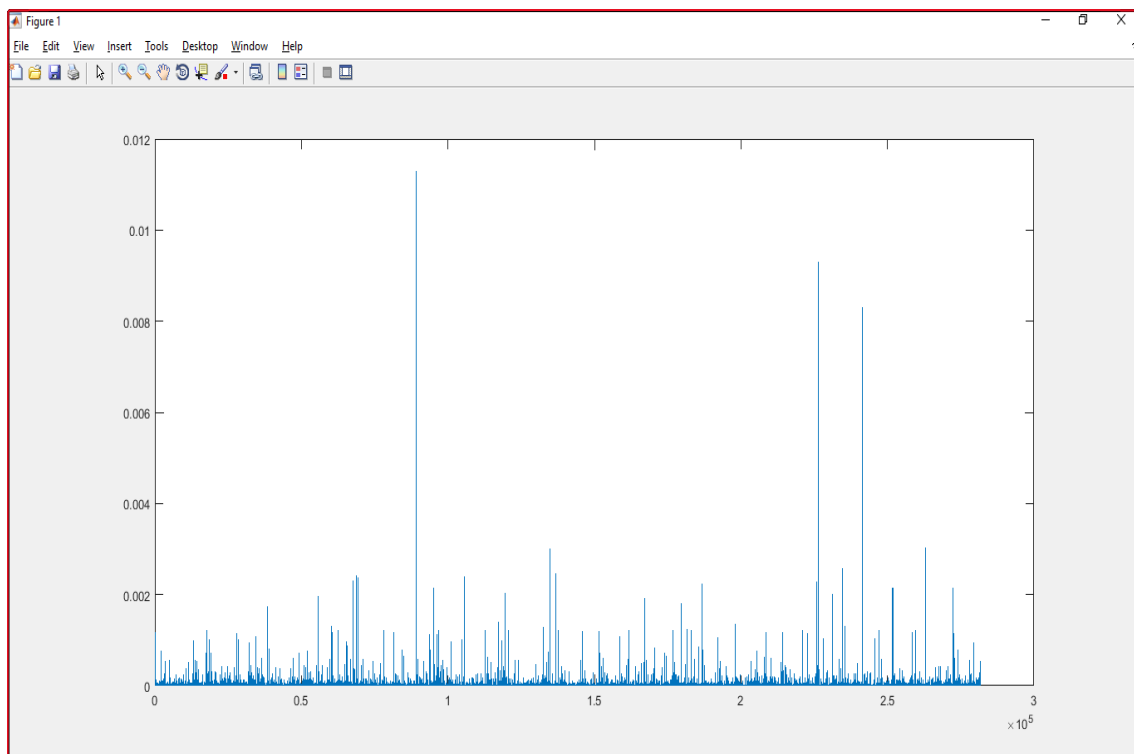
4. Calcular el pagerank de Stanford Web Matrix.

	Tiempo [seg]	Memoria [MB]	Nº iteraciones	Precisión Norm(Gx- x)
N=281x10 ³	20.1424	2255224 Bytes	150	1.0207e-12

6. Visualizar y ordenar los resultados.

```
>>bar(x)
```

```
>>plot(x)
```



```
>>[S,y] = sort(x,'descend'); -Creamos una función para imprimir los valores en una tabla:
```

```
function imprimir(S,y)
for k = 1:10
    fprintf('Orden %1.0f Nodo %4.0f Pagerank %1.4f \n',k,y(k),S(k));
end
return
```

```
>> imprimir(S,y)
```

Orden 1 Nodo 89073 Pagerank 0.0113

Orden 2 Nodo 226411 Pagerank 0.0093

Orden 3 Nodo 241454 Pagerank 0.0083

Orden 4 Nodo 262860 Pagerank 0.0030

Orden 5 Nodo 134832 Pagerank 0.0030

Orden 6 Nodo 234704 Pagerank 0.0026

Orden 7 Nodo 136821 Pagerank 0.0025

Orden 8 Nodo 68889 Pagerank 0.0024

Orden 9 Nodo 105607 Pagerank 0.0024

Orden 10 Nodo 69358 Pagerank 0.0024

```
>> [S,y] = sort(x);
```

Modificamos la función para que el pagerank imprima más decimales

```
function imprimir(S,y)
for k = 1:10
    fprintf('Orden %1.0f Nodo %4.0f Pagerank %1.12f \n',k,y(k),S(k));
end
return
```

```
>> imprimir(S,y)
```

```
Orden 1 Nodo  1 Pagerank 0.000000533366
```

```
Orden 2 Nodo  4 Pagerank 0.000000533366
```

```
Orden 3 Nodo  8 Pagerank 0.000000533366
```

```
Orden 4 Nodo 24 Pagerank 0.000000533366
```

```
Orden 5 Nodo 26 Pagerank 0.000000533366
```

```
Orden 6 Nodo 63 Pagerank 0.000000533366
```

```
Orden 7 Nodo 82 Pagerank 0.000000533366
```

```
Orden 8 Nodo 96 Pagerank 0.000000533366
```

```
Orden 9 Nodo 106 Pagerank 0.000000533366
```

```
Orden 10 Nodo 131 Pagerank 0.000000533366
```

Sí, los nodos que tienen el menor pagerank, tienen el mismo valor.

Identificamos el menor valor y modificamos nuestra función imprimir para que imprima todos estos nodos, que resultan ser 20315

```
function imprimir(S,y)
for k = 1: length(S)
    if S(k) == 5.333661624414200e-07
        fprintf('Orden %1.0f Nodo %4.0f Pagerank %1.12f \n',k,y(k),S(k));
    end
end
return
```

```
>> imprimir(S,y) %Para visualizar los 20315 nodos que tienen el menos valor
```

2. Calcular e interpretar el pagerank de SNAP/cit-Patents

2. Cargar la matriz B

```
>> load cit-Patents.mat
```

```
>> Problem
```

```
>> B=Problem.A;
```

```
>> Nj=sum(B); nnz(B) )%Para ver si la matriz es dispersa o no. Contamos el numero de los elementos distintos de 0
```

```
ans =
```

```
16518948
```

```
>> spy(B);title('Gráfica de la dispersión de la matriz SNAP/cit-Patents')
```

3. Determinar las características de la matriz B, de la misma forma que en el apartado 3 del ejercicio anterior

La matriz es cuadrada es de 3774768x3774768. Tiene 3774768 nodos.

La matriz es dispersa porque de los 281903X281903 elementos solo 16518948 son distintos de cero.

```
>> whos
```

Name	Size	Bytes	Class	Attributes
B	3774768x3774768	294501320	double	sparse
Problem	1x1	324711900	struct	

La matriz B ocupa en memoria 294501320 Bytes.

Solo 16518948 son distintos de cero.

La matriz que nos dan es una matriz de tipo B debido a que como máximo sus elementos de Nj son 1 pero tiene algunos que son 0, por lo que no puede ser una matriz S, y además es dispersa.

```
>> dispersionB = (nnz(B)/ (3774768*3774768))
```

```
dispersionB =
```

```
1.1593e-06
```

```
>> r = sum(Nj) / length(Nj)
```

```
r =
```

```
(1,1) 4.3761
```

Cada página tiene de media 4.3761 links de salida

```
>> NodosSinSalida = length(find(~Nj))

NodosSinSalida =

    515785

>> NodosConSalida = nnz(Nj)

NodosConSalida =

    3258983

>> NodosTotales= NodosConSalida + NodosSinSalida

NodosTotales =

    3774768

Que es igual a:

>> length(B)

ans =

    3774768
```

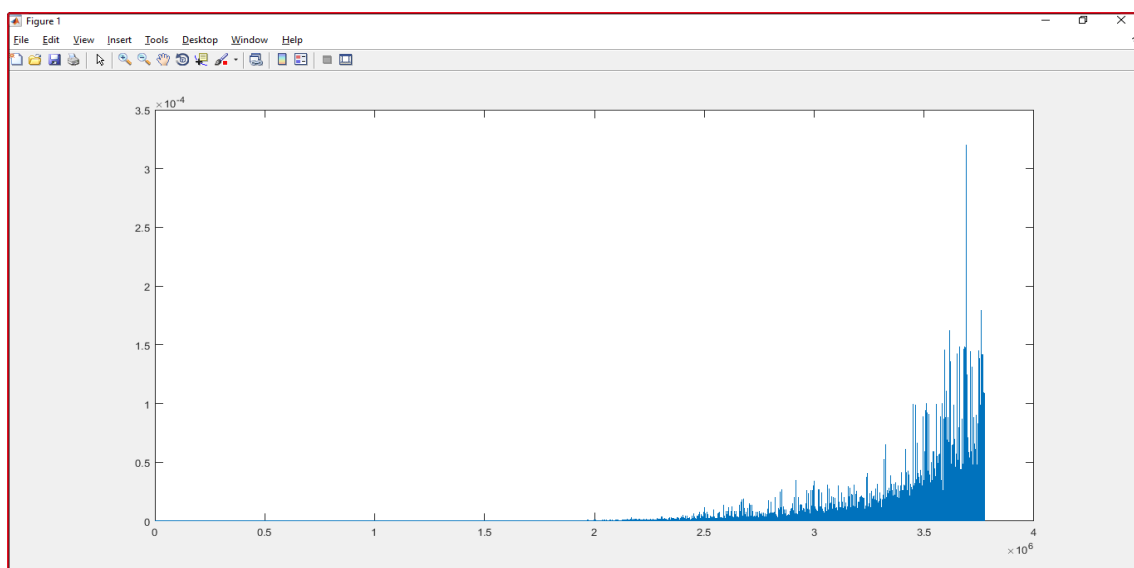
4. Calcular el pagerank de Stanford Web Matrix.

	Tiempo [seg]	Memoria [MB]	Nº iteraciones	Precisión Norm(Gx- x)
N=281x10 ³	230.1696	30198144 Bytes	150	9.5514

5. Visualizar y ordenar los resultados.

```
>> bar(x)
```

```
>> plot(x)
```



```
>>[S,y] = sort(x,'descend');
```

-Creamos una función para imprimir los valores en una tabla:

```
function imprimir(S,y)
for k = 1:10
    fprintf('Orden %1.0f Nodo %4.0f Pagerank %1.12f \n',k,y(k),S(k));
end
return
```

```
>> imprimir(S,y)
```

Orden 1 Nodo 3690055 Pagerank 0.000320375102

Orden 2 Nodo 3760370 Pagerank 0.000179049638

Orden 3 Nodo 3614420 Pagerank 0.000162747357

Orden 4 Nodo 3686115 Pagerank 0.000148628705

Orden 5 Nodo 3661750 Pagerank 0.000148202033

Orden 6 Nodo 3683317 Pagerank 0.000148037675

Orden 7 Nodo 3688969 Pagerank 0.000148037675

Orden 8 Nodo 3679419 Pagerank 0.000146632816

Orden 9 Nodo 3593454 Pagerank 0.000145592736

Orden 10 Nodo 3749040 Pagerank 0.000144720101

```
>> [S,y] = sort(x);
```

```
>> imprimir(S,y)
```

Orden 1 Nodo 1 Pagerank 0.000000035513

Orden 2 Nodo 2 Pagerank 0.000000035513

Orden 3 Nodo 3 Pagerank 0.000000035513

Orden 4 Nodo 4 Pagerank 0.000000035513

Orden 5 Nodo 5 Pagerank 0.000000035513

Orden 6 Nodo 6 Pagerank 0.000000035513

Orden 7 Nodo 7 Pagerank 0.000000035513

Orden 8 Nodo 8 Pagerank 0.000000035513

Orden 9 Nodo 9 Pagerank 0.000000035513

Orden 10 Nodo 10 Pagerank 0.000000035513

Sí, los nodos que tienen el menor pagerank, tienen el mismo valor.

Identificamos el menor valor y modificamos nuestra función imprimir para que imprima todos estos nodos, que resultan ser 1685423

```

function imprimir(S,y)
for k = 1: length(S)
    if S(k) == 3.551347717264515e-08
        fprintf('Orden %1.0f Nodo %4.0f Pagerank %1.12f \n',k,y(k),S(k));
    end
end
return

```

>> imprimir(S,y) %Para visualizar los 1685423 nodos que tienen el menos valor

6. Describir la matriz SNAP/cit-Patents. ¿Cuál es su contenido?. ¿Cuál es su finalidad?. Describir e interpretar las características de la matriz.

La matriz SNAP/cit-Patents contiene todas las patentes de utilidad concedidas en citaciones durante los años 1963 y 1999

Estadísticas:

Dataset statistics	
Nodes	3774768
Edges	16518948
Nodes in largest WCC	3764117 (0.997)
Edges in largest WCC	16511741 (1.000)
Nodes in largest SCC	1 (0.000)
Edges in largest SCC	0 (0.000)
Average clustering coefficient	0.0757
Number of triangles	7515023
Fraction of closed triangles	0.02343
Diameter (longest shortest path)	22
90-percentile effective diameter	9.4