

PageRank : Script e Test dell'Algoritmo

Il seguente script mostra l'utilizzo dell'Algoritmo di PageRank implementato leggendo la matrice di Adiacenze G e un array U di URL da un file esterno o adoperando la funzione "surfer".

Descrizione

Lo script calcola:

- Il PageRank della Matrice di Adiacenza G,
- Mostra la sparsità della stessa attraverso un grafico,
- Mostra i primi 15 risultati in ordine decrescente in termini di Outdegree e Indegree,
- Ricava un grafico a barre dei PageRank calcolati,
- Mostra un sottografo orientato dove i nodi rappresentano le pagine web con rank maggiore della media dei rank.

Inoltre vengono effettuati dei casi di test sulla correttezza dell'algoritmo implementato e sulla performance dello stesso prendendo di riferimento la funzione centrality del MATLAB. Infine vengono effettuati dei casi di test sulla robustezza dell'algoritmo.

Utilizzo

Per prima cosa bisogna pulire il workspace, caricare la matrice G e l'array U da un file esterno. Utilizziamo lo script surfer.

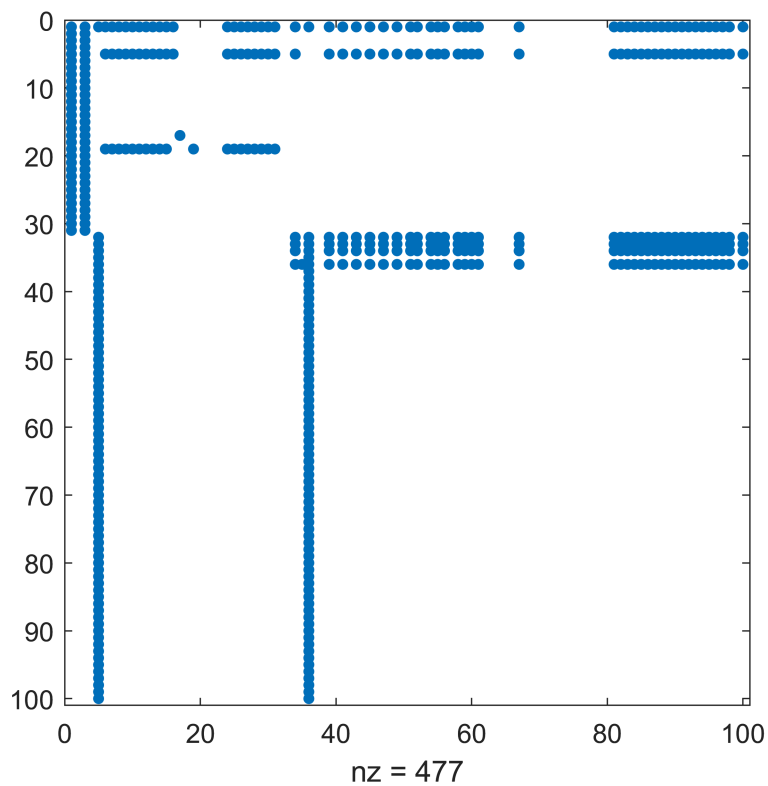
```
clear
[U,G] = surfer('http://www.unina.it',100);
%load mathwork200.mat
%load repubblica.mat
```

NB: In alternativa alla funzione surfer è possibile caricare uno tra i file "mathwork200.mat" e "repubblica.mat" che contengono già la matrice G e l'array U.

Struttura della Matrice G

Per visualizzare la struttura della matrice G viene utilizzato il comando spy che mostra un grafico della sparsità della matrice.

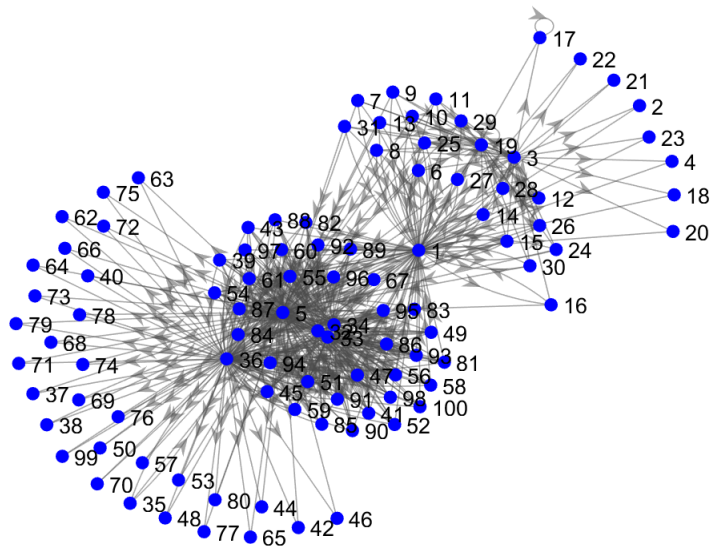
```
spy(G)
```



Grafo Associato alla Matrice G

In allegato alla struttura della matrice G, viene ricavato il grafo associato ad essa. Ogni nodo del grafo rappresenta l'identificativo della pagina e ogni arco i collegamenti (URL) tra le pagine.

```
grafo = digraph(G');
plot(grafo, 'NodeColor', 'b', 'EdgeColor', [.3 .3 .3]);
axis off
```



Uso dell'Algoritmo di PageRank

Una volta ottenuta la matrice G e l'array U è possibile ottenere dall'Algoritmo di PageRank, il vettore dei Rank delle pagine associate ai relativi Outdegree e Indegree, per poi rappresentarli sul grafo associato alla matrice G .

N.B : Per informazioni sull'Algoritmo di PageRank si faccia riferimento alla documentazione specificata nel file "Documentazione_PageRank.pdf".

```
[R,Outdeg,Indeg] = PageRank(G);
grafo.Nodes.PageRank = R;
grafo.Nodes.InDegree = Indeg;
grafo.Nodes.OutDegree = Outdeg;
```

Tabella degli Outdegree e Indegree

Mostriamo dunque di seguito i primi 15 Rank prodotti dall'algoritmo in ordine decrescente, con i rispettivi Identificativi, OutDegree, InDegree nel grafo.

```
[R,indice] = sort(R,'descend');
siti_web = U(indice(1:15));
id = indice(1:15);
ranks = R(1:15);
outdegrees = Outdeg(indice(1:15));
indegrees = Indeg(indice(1:15));
table(id,ranks,outdegrees,indegrees,'RowNames',siti_web)
```

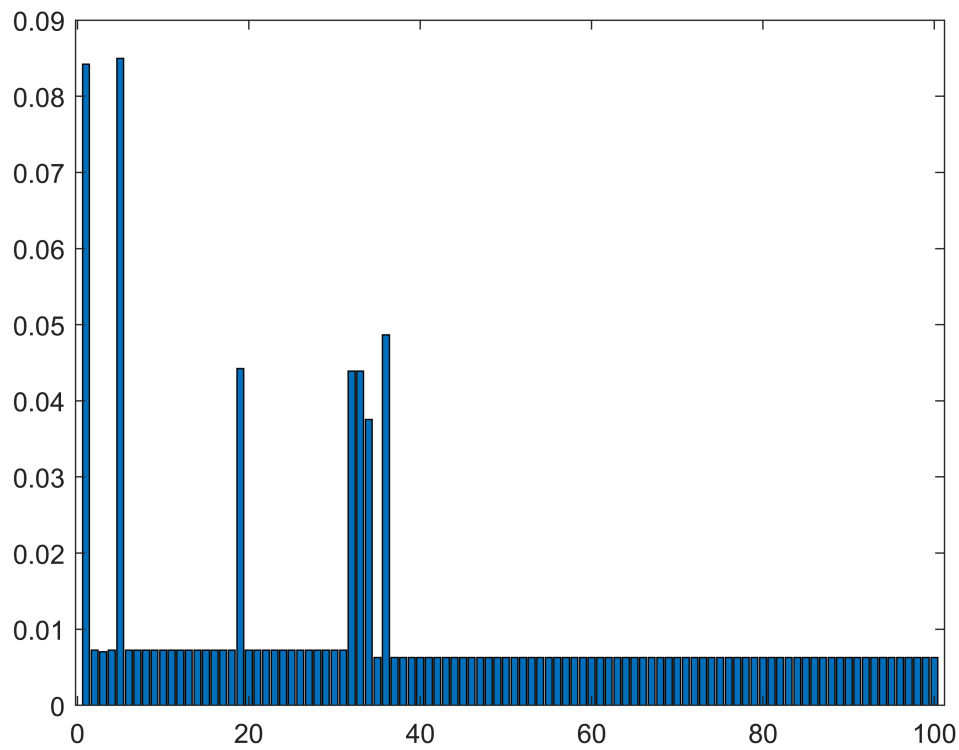
ans = 15x4 table

	id	ranks	outdegrees	indegrees
1 http://www.sinapsi.unina.it	5	0.0850	70	57
2 http://www.unina.it	1	0.0842	30	58
3 http://www.sinapsi.unina.it/h...	36	0.0487	69	38
4 http://www.radiof2.unina.it	19	0.0443	0	20
5 http://www.sinapsi.unina.it/f...	32	0.0439	0	38
6 http://www.sinapsi.unina.it/f...	33	0.0439	0	38
7 http://www.sinapsi.unina.it/f...	34	0.0375	5	37
8 http://www.unina.it/unina-hom...	2	0.0073	0	2
9 http://www.unina.it/unina-hom...	4	0.0073	0	2
10 http://www.unina.it/~188890...	6	0.0073	3	2
11 http://www.unina.it/~189654...	7	0.0073	3	2
12 http://www.unina.it/~189396...	8	0.0073	3	2
13 http://www.unina.it/~189692...	9	0.0073	3	2
14 http://www.unina.it/~189332...	10	0.0073	3	2
15 http://www.unina.it/~189320...	11	0.0073	3	2

Istogramma dei PageRank

Di seguito mostriamo un istogramma dei PageRank ottenuti dall'Algoritmo.

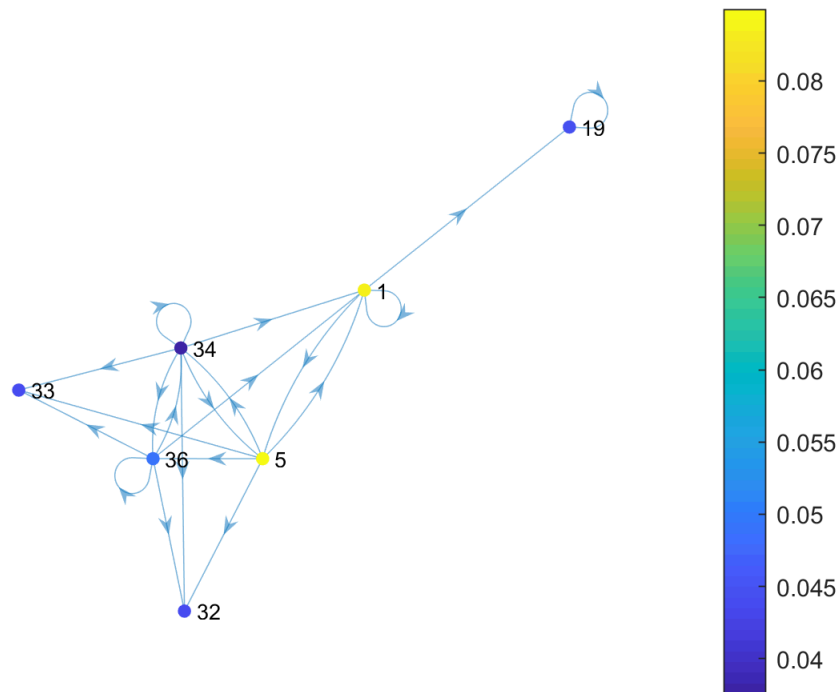
```
bar(grafo.Nodes.PageRank)
```



Sottografo dei PageRank

Associato all'istogramma, mostriamo il sottografo dei nodi con PageRank maggiore della media dei PageRank.

```
condition = find(grafo.Nodes.PageRank > mean(grafo.Nodes.PageRank));  
sotto_grafo = subgraph(grafo,condition);  
plot(sotto_grafo, 'NodeLabel',condition, 'NodeCData',sotto_grafo.Nodes.PageRank, 'Layout', 'force',  
colorbar;  
axis off;
```



Valutazione Performance e Test di Correttezza

Per valutare la correttezza dell'algoritmo implementato viene valutata la performance confrontandola alla funzione centrality del MATLAB, mostrando in una tabella i risultati ottenuti.

Valuta_Performance()

ans = 15x4 table

	ID	rank	outdegree	indegree
1 http://www.sinapsi.unina.it	5	0.0850	70	57
2 http://www.unina.it	1	0.0842	30	58
3 http://www.sinapsi.unina.it/h...	36	0.0487	69	38
4 http://www.radiof2.unina.it	19	0.0443	0	20
5 http://www.sinapsi.unina.it/f...	32	0.0439	0	38
6 http://www.sinapsi.unina.it/f...	33	0.0439	0	38
7 http://www.sinapsi.unina.it/f...	34	0.0375	5	37
8 http://www.unina.it/unina-hom...	2	0.0073	0	2
9 http://www.unina.it/unina-hom...	4	0.0073	0	2
10 http://www.unina.it/~188890...	6	0.0073	3	2

	ID	rank	outdegree	indegree
11 http://www.unina.it/~189654...	7	0.0073	3	2
12 http://www.unina.it/~189396...	8	0.0073	3	2
13 http://www.unina.it/~189692...	9	0.0073	3	2
14 http://www.unina.it/~189332...	10	0.0073	3	2
15 http://www.unina.it/~189320...	11	0.0073	3	2

Test di Robustezza

Per valutare la robustezza dell'algoritmo è stata implementata una test suite. I casi di test sono stati scelti a partire dalle condizioni di errore o warning che possono avere luogo. Ogni caso di test lo descriveremo brevemente a partire dai parametri di input e dalla funzionalità che viene testata. Il numero di casi di test è: 8, implementati in una classe definita dal matlab *matlab.unittest.TestCase*. In tal modo si è automatizzato il processo di esecuzione dei test. Le istruzioni per eseguire i test saranno mostrate successivamente.

Di seguito sono mostrati i casi di test più rilevanti che sono stati effettuati, rispettando la precedenza tra essi.

Test case 3

Verifica l'errore nel caso in cui la matrice G è non sparsa.

Input

```
G = rand(10)
```

Test case 5

Verifica l'errore nel caso in cui la matrice G è sparsa ma non quadrata.

Input

```
G = rand(10, 15)
```

Test case 6

Verifica l'errore nel caso in cui la matrice G non è logica.

Input

```
G = round(sprand(10, 10, 0.2) * 10)
```

Test case 7

Verifica il warning nel caso in cui la matrice G contenga dei selfloops.

Input

```
G = logical(G + speye(size(G)))
```

Esecuzione Test Suite

Vengono effettuati in modo automatizzato i casi di test sugli input definiti in precedenza.

```
results = runtests('CasiTest.m')
```

```
Running CasiTest
```

```
.....
```

```
Done CasiTest
```

```
results =
```

```
1x7 TestResult array with properties:
```

```
Name  
Passed  
Failed  
Incomplete  
Duration  
Details
```

```
Totals:
```

```
7 Passed, 0 Failed, 0 Incomplete.
```

```
1.69 seconds testing time.
```

Riportiamo su una tabella i casi di test effettuati.

```
table(results)
```

```
ans = 7x6 table
```

	Name	Passed	Failed	Incomplete	Duration	Details
1	'CasiTest...	1	0	0	0.6311	1x1 struct
2	'CasiTest...	1	0	0	0.0457	1x1 struct
3	'CasiTest...	1	0	0	0.0381	1x1 struct
4	'CasiTest...	1	0	0	0.0333	1x1 struct
5	'CasiTest...	1	0	0	0.1281	1x1 struct
6	'CasiTest...	1	0	0	0.0806	1x1 struct
7	'CasiTest...	1	0	0	0.7332	1x1 struct

In particolare l'ultimo caso di test effettuato verifica il funzionamento dell'algoritmo nel caso in cui sono presenti self loops, che indicano pagine che si autoreferenziano. Queste non devono essere considerate nel calcolo del rank, per tale motivo il caso di test genererà un warning. Riportiamo per completezza il comportamento di quest'ultimo caso di test.

```
results = runtests('CasiTest.m','ProcedureName','TestCase7')
```

```
Running CasiTest
```

```
.
```

```
Done CasiTest
```

```
results =
```

```
TestResult with properties:
```

```
Name: 'CasiTest/TestCase7'  
Passed: 1
```



```
Failed: 0
Incomplete: 0
Duration: 0.0456
Details: [1x1 struct]
```

Totals:

```
1 Passed, 0 Failed, 0 Incomplete.
0.045559 seconds testing time.
```

Riferimenti

- Testing in Matlab: <https://it.mathworks.com/help/matlab/matlab-unit-test-framework.html>
- Misura di Performance in Matlab <https://it.mathworks.com/help/matlab/performance-and-memory.html>
- PageRank - Wikipedia <https://it.wikipedia.org/wiki/PageRank>

Autori

Giuseppe Napolano M63000856 Raffaele Formisano M63000912 Giuseppe Romito M63000936