

PageRank : Script e Test dell'Algoritmo

Il seguente script mostra l'utilizzo dell'Algoritmo di PageRank implementato andando a leggere 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 G di Adiacenza
- 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 Grafo orientato dove i nodi rappresentano le pagine web con rank maggiore della media dei rank.

Oltre a ciò 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 e caricare la matrice G e l'array U da un file esterno.

Utilizziamo lo script *"surfer"*

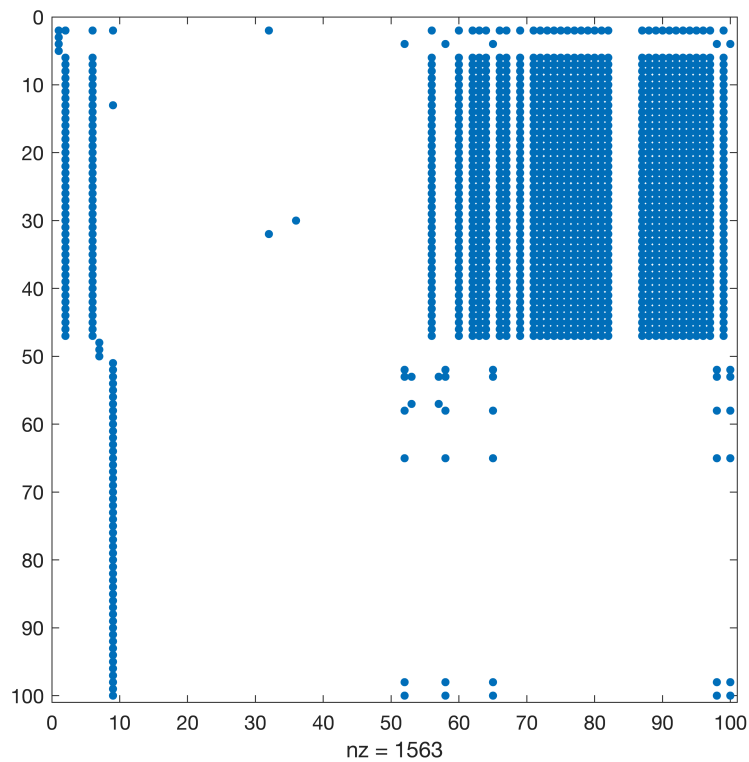
```
clear
[U,G] = surfer('https://www.apple.com/it/',100);
%load mathwork200.mat
%load repubblica.mat
```

In alternativa alla funzione *surfer* è possibile caricare uno tra i file *"mathwork200.mat"* e *"repubblica.mat"* che contengono già la matrice G e il vettore 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.

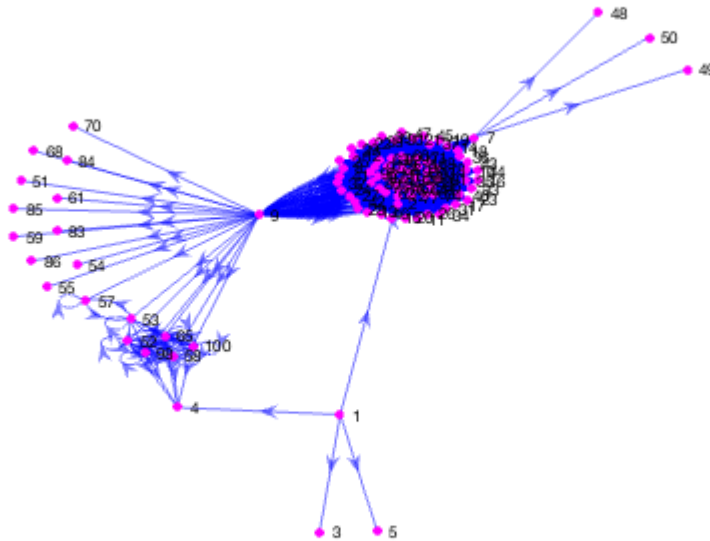
```
spy(G)
```



Grafo Associato della Matrice G

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

```
grafo = digraph(G');
plot(grafo, 'NodeColor', 'm', 'EdgeColor', 'b', 'Layout', 'force');
axis off
```



Uso dell'Algoritmo di PageRank

Una volta ottenuta la matrice G ed il vettore U è possibile ottenere dall'Algoritmo di PageRank, il vettore dei Rank delle pagine associati agli Outdegree e Indegree delle stesse per poi associarle al grafo in sovrapposizione.

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 senso decrescente, in accoppiata ai corrispettivi OutDegree, InDegree e agli Identificativi 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)
```

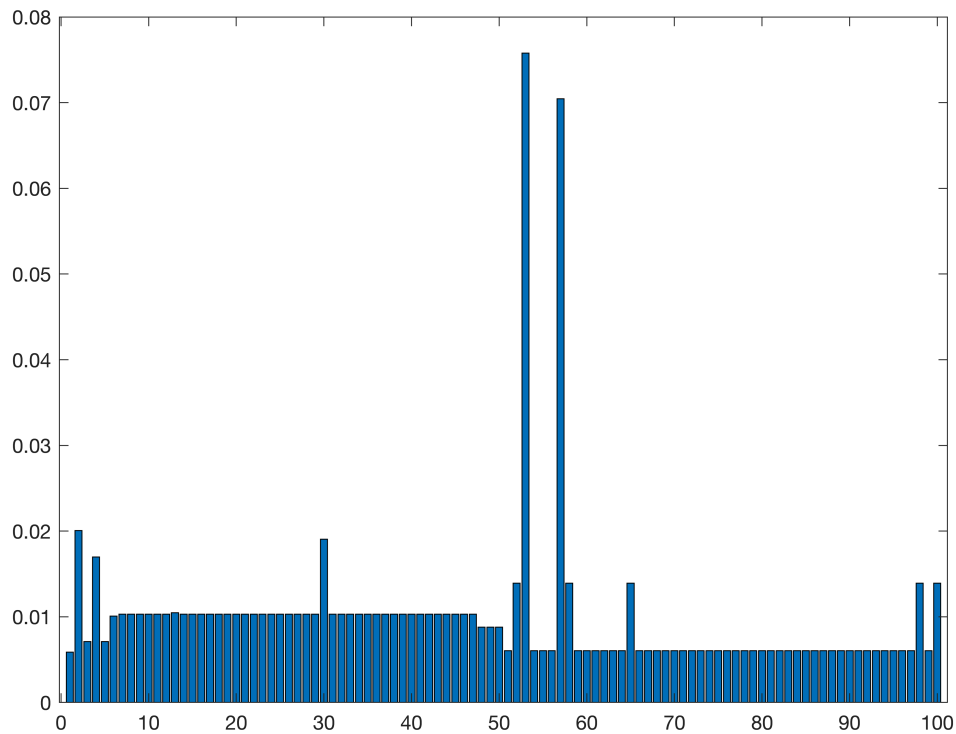
```
ans = 15x4 table
```

	id	ranks	outdegrees	indegrees
1 http://purl.org/dc/elements/1.1	53	0.0758	1	7
2 http://purl.org/dc/elements/1...	57	0.0705	1	2
3 http://ogp.me/ns#	2	0.0201	42	36
4 http://groups.google.com/grou...	30	0.0190	0	35
5 http://schema.org	4	0.0170	0	6
6 http://xmlns.com/foaf/0.1	52	0.0139	6	5
7 http://xmlns.com/foaf/0.1/hom...	58	0.0139	6	5
8 http://xmlns.com/foaf/0.1/dep...	65	0.0139	6	5
9 http://xmlns.com/foaf/0.1/mbox	98	0.0139	6	5
10 http://xmlns.com/foaf/0.1/ph...	100	0.0139	6	5
11 http://www.imdb.com/title/tt...	13	0.0105	0	35
12 http://ogp.me/logo.png	7	0.0103	3	34
13 http://ogp.me/ns/fb#	8	0.0103	0	34
14 http://ogp.me/ns/ogp.me.rdf	9	0.0103	52	34
15 http://ogp.me/ns/ogp.me.ttl	10	0.0103	0	34

Istogramma dei PageRank

Di seguito mostriamo un istogramma dei PageRank ottenuti dall'Algoritmo

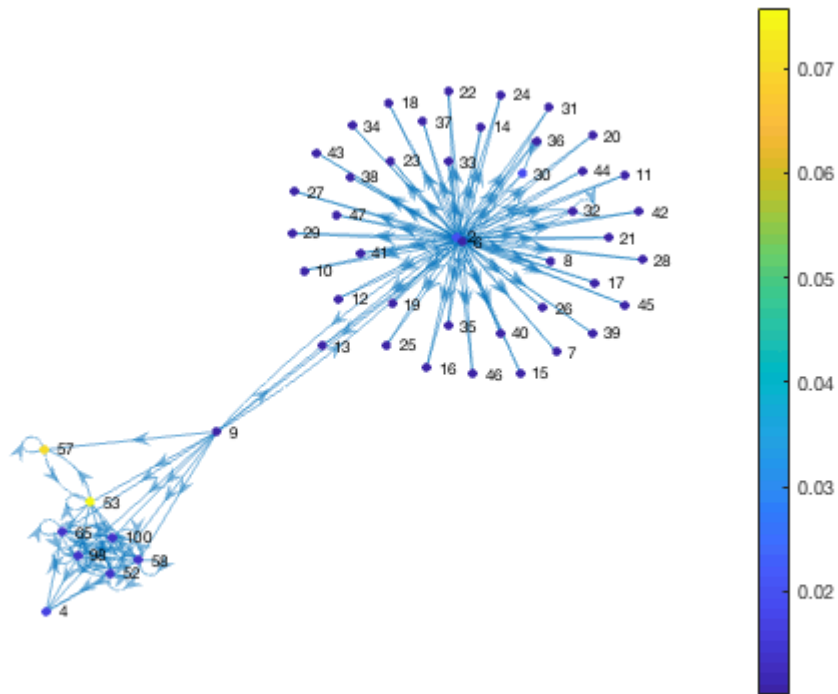
```
bar(grafo.Nodes.PageRank)
```



Sottografo dei PageRank

Associato all'istogramma, mostriamo il sottografo dei Nodi tali che essi abbiano 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 effettuata una valutazione delle performance rispetto alla funzione "centrality" del MATLAB, mostrando in una tabella i risultati ottenuti.

Valuta_Performance()

ans = 15x4 table

	ID	rank	outdegree	indegree
1 http://purl.org/dc/elements/1.1	53	0.0758	1	7
2 http://purl.org/dc/elements/1...	57	0.0705	1	2
3 http://ogp.me/ns#	2	0.0201	42	36
4 http://groups.google.com/grou...	30	0.0190	0	35
5 http://schema.org	4	0.0170	0	6
6 http://xmlns.com/foaf/0.1	52	0.0139	6	5
7 http://xmlns.com/foaf/0.1/hom...	58	0.0139	6	5
8 http://xmlns.com/foaf/0.1/dep...	65	0.0139	6	5
9 http://xmlns.com/foaf/0.1/mbox	98	0.0139	6	5
10 http://xmlns.com/foaf/0.1/ph...	100	0.0139	6	5

	ID	rank	outdegree	indegree
11 http://www.imdb.com/title/tt...	13	0.0105	0	35
12 http://ogp.me/logo.png	7	0.0103	3	34
13 http://ogp.me/ns/fb#	8	0.0103	0	34
14 http://ogp.me/ns/ogp.me.rdf	9	0.0103	52	34
15 http://ogp.me/ns/ogp.me.ttl	10	0.0103	0	34

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 1

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

Input

$G = \text{rand}(10)$

Test case 2

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

Input

$G = \text{rand}(10, 15)$

Test case 3

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

Input

$G = \text{round}(\text{sprand}(10, 10, 0.2) * 10)$

Test case 4

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

Input

$G = \text{logical}(G + \text{speye}(\text{size}(G)))$

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.
    0.064495 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	0.0084	1x1 struct
2	'CasiTest...	1	0	0	0.0100	1x1 struct
3	'CasiTest...	1	0	0	0.0064	1x1 struct
4	'CasiTest...	1	0	0	0.0069	1x1 struct
5	'CasiTest...	1	0	0	0.0101	1x1 struct
6	'CasiTest...	1	0	0	0.0061	1x1 struct
7	'CasiTest...	1	0	0	0.0166	1x1 struct

In particolare l'ultimo caso di test effettuato verifica il funzionamento dell'algorithmo nel caso in cui sono presenti self loops, che indicano pagine che si autoreferenziano. Queste non devono essere considerate nel calcolo del rank e appunto il caso di test genererà un warning apposito. Riportiamo per completezza il comportamento del suddetto 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: 0.0173  
Details: [1x1 struct]
```

Totals:

```
1 Passed, 0 Failed, 0 Incomplete.  
0.017324 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