

PageRank

Calcola il rank delle pagine web tramite l'algoritmo **PageRank** di Google, per poter realizzare un ordinamento univoco delle pagine Web in base alla loro importanza.

Sintassi

```
[R, OUT, IN] = PageRank(G)
```

Descrizione

- `[R, OUT, IN] = PageRank(G)` : calcola, a partire dalla matrice di adiacenza G che descrive il collegamento, mediante i link, tra le pagine web, il rank dei siti web mediante l'algoritmo **PageRank** di Google. Inoltre, la funzione, calcola per ogni pagina web gli outdegree (OUT) e gli indegree (IN), che sono rispettivamente il numero di link uscenti da quella pagina web e il numero di link che puntano a quella pagina web.

Esempi

```
i=[1 1 1 2 2 3 3 6 8 10 10 10 ];
j=[2 3 4 4 10 3 10 10 9 1 2 3 ];
value(1:size(j,2))=true;

G=sparse(i,j,value);

[R,OUT,IN] = PageRank(G)
```

```
R = 10x1
    0.1557
    0.1295
    0.1139
    0.0367
    0.0367
    0.1139
    0.0367
    0.0679
    0.0367
    0.2725
OUT = 10x1
     1
     2
     2
     2
     0
     0
     0
     0
     1
     3
IN = 10x1
     3
```

```
2
1
0
0
1
0
1
0
3
```

Argomenti di input

G - Matrice di adiacenza (Matrice di logical, sparsa, dimensione NxN)

Matrice di adiacenza relativa ad un insieme di pagine web che rappresenta il grafo di connettività tra le pagine web, che indica come le pagine sono collegate tra loro tramite i link presenti in esse. La matrice G deve essere una matrice quadrata sparsa contenente elementi di tipo logical .

Esempio :

```
i=[1 1 2 3 10 10 10 11 15 20];
j=[2 3 3 4 8 3 20 10 11 15];
value(1:size(j,2))=true;

G=sparse(i,j,value)
```

```
ans = 20x20 sparse logical array
```

```
(1,2)      1
(1,3)      1
(2,3)      1
(10,3)     1
(3,4)      1
(10,8)     1
(11,10)    1
(15,11)    1
(20,15)    1
(10,20)    1
```

Data Types : logical

Argomenti di output

R - Rank delle pagine (Array di double di dimensione N)

Vettore contenente il rank delle pagine calcolato utilizzando l' algoritmo del **PageRank** a partire da G. Il rank è calcolato in base al numero di link che puntano a quella pagina, tale parametro determina l'importanza della pagina, più alto è e più è importante. Il rank viene a sua volta distribuito a tutte le pagine a cui quella pagina punta. Inoltre tale valore non è influenzato dalla presenza dei self-loop[1] e dei dangling node[2].

Data Types : single | double

OUT - Outdegrees (Array di interi di dimension N)

Vettore degli outdegrees delle pagine web, numero di link uscenti dalla pagina ottenuto come somma degli elementi della colonna di G relativa alla pagina considerata.

Data Types : integer

IN - Indegrees (Array di interi di dimension N)

Vettore degli indegrees delle pagine web, numero di link che puntano alla pagina ottenuto come somma degli elementi della riga di G relativa alla pagina considerata.

Data Types : integer

Note

[1] self-loop : link che puntano alla medesima pagina in cui essi sono presenti.

[2] dangling node : Risorse che sono puntate da un link ma che non puntano a nessuna altra pagina web, come ad esempio immagini o video.

Autore

Gabriele Previtera