

# PageRank (*pagerank*)

---

La Professoressa Finocchi ha sfidato i tutor: devono riuscire a calcolare quanto sia facile trovare la pagina del Corso di preparazione alle Olimpiadi a partire da vari URL, come quello del sito della LUISS.

La LUISS ha fornito una serie di grafi non orientati che rappresentano i collegamenti tra vari siti web legati all'ateneo – per esempio, *sport.luiss.it* oppure *impresaemanagement.luiss.it* – tra cui è sempre presente il suddetto sito del nostro Corso.

Demetrio ha scoperto una cosa: all'interno di questi grafi, il sito più "raggiungibile" è proprio quello del Corso. Per questo motivo, ha chiesto a Davide di verificare questa osservazione tramite l'algoritmo PageRank.



## *I tutor nel Loft*

PageRank ( $\$PR\$$ ) assegna un valore numerico a ogni pagina all'interno del World Wide Web che ne quantifica l'importanza o il prestigio relativi all'interno di questa rete. Fu introdotto nel 1998 da Sergey Brin e Larry Page come ingrediente fondamentale di Google.

Nell'algoritmo in questione, bisogna descrivere il comportamento di un utente che naviga nelle reti rappresentate dai grafi forniti in input, cliccando in maniera randomica sui link che collegano le pagine. L'assunzione iniziale è che ogni pagina ha la stessa probabilità iniziale di essere raggiunta, ma le pagine con una centralità maggiore o presenti in aree più dense del grafo saranno quelle più facilmente raggiungibili.

In particolare, si procede come segue:

- ogni nodo viene inizializzato con un valore di  $\$PR\$$  iniziale  $\$R_0=1/N\$$ , dove  $\$N\$$  è il numero di nodi del grafo;

- a ogni iterazione  $t$ , si itera su  $N$  nodi  $i$  per aggiornare il loro valore di  $PR: R_t(i) = \frac{\alpha}{N + (1 - \alpha) \sum_{j \in \text{pred}(i)} \frac{R_{t-1}(j)}{k_{\text{out}}(j)}}$ 
  - $\alpha$  rappresenta il **fattore di teletrasporto** o di smorzamento, ossia la probabilità che l'utente interrompa la navigazione e ricominci un nuovo percorso partendo da una nuova pagina randomica.
  - $\text{pred}(i)$  è l'insieme dei predecessori  $j$  del nodo  $i$ : ossia l'insieme dei nodi (le pagine) direttamente collegati al nodo  $i$ .
  - $k_{\text{out}}(j)$  rappresenta il numero di collegamenti uscenti dal predecessore  $j$ , ossia il numero di  $\text{pred}(j)$ , poiché, in un grafo non orientato,  $\text{length}(\text{pred}(i)) = k_{\text{in}}(i) = k_{\text{out}}(i)$ .

In parole povere, il  $PR$  iniziale di ogni nodo viene aggiornato con ogni iterazione a un valore che dipende dal numero dei collegamenti del nodo, dal numero dei collegamenti dei nodi collegati e così via. In base al coefficiente  $\alpha$ , il  $PR$  convergerà a un valore costante.

## Esempio esplicativo

### Input:

```
1
3

5
0  1 3 4
1  0 2 3 4
2  1 4
3  0 1 4
4  0 1 2 3
```

Ovvero:

- 1, il numero di casi di test;
- 3, ossia la parte decimale del fattore di teletrasporto  $\alpha$  (per cui in questo caso  $\alpha = 0,3$ ), seguito da una riga vuota;
- 5, ossia il numero  $N$  di nodi del grafo;
- 0, ossia l'indice del primo nodo, seguito da due spazi e dagli indici dei nodi con cui è connesso, separati da uno spazio: 1, 3 e 4;
- 1, ossia l'indice del secondo nodo e così via.

### Output:

```
Case #1: 4
```

In questo caso, dopo aver ricostruito il grafo, applicato l'algoritmo con il corretto  $\alpha$  e rilevato il nodo con il valore di PageRank massimo, se ne ottiene l'indice, in questo caso 4.

*Aiuta Demetrio e Davide a scoprire la posizione della pagina con il PageRank più alto, ossia a verificare che sia quella del nostro Corso; altrimenti, la Professoressa Finocchi troverà presto due nuovi tutor per rimpiazzarli!*

---

## Dati di Input

Il tuo input sarà dato da:

- un valore  $\alpha < 10$ ,  $\alpha \in \mathbb{N}$  che rappresenta la parte decimale del fattore di teletrasporto;
- il numero  $N$  di nodi del grafo;
- L'indice  $i$  del  $(i+1)$ esimo nodo, seguito da due spazi e dagli indici dei nodi con cui è connesso, separati da uno spazio.

N.B. **tutti** i nodi del grafo sono rappresentati, quindi un nodo isolato sarà presente nell'input ma non sarà seguito da alcun numero. Inoltre, si ricorda che nel caso di più casi di test, essi sono separati da una riga vuota.

## Assunzioni

- $i$  intero,  $\alpha$  naturale;
- $\alpha < 10$ ;
- $0 < N \leq 10^3$ ;
- $0 \leq i < N$ .