

Sistemi Informativi

Laboratorio 7

Catalin Copil

Mattia de Stefani

Giulio Lovisotto

May 20, 2015

1 Descrizione

La nostra funzione di reperimento utilizzerà i primi N documenti prodotti da BM25 per costruire il grafo delle citazioni R_q , tale grafo verrà espanso per ottenere un grafo allargato B_q , e su quest'ultimo verrà calcolato HITS. La funzione combinerà poi gli score di BM25 (sc) con i punteggi di authority ($auth$) e hubbiness (hub) per i primi N documenti, e li riordinerà per $hits_{score}$ decrescente:

$$hits_{score} = \alpha \cdot sc + \beta \cdot auth + \gamma \cdot hub,$$

α varia in $[0, 1]$, mentre β, γ sono parametri che possono variare tra $[-1, 1]$, in quanto vogliamo cogliere possibili influenze negative sul di tali valori. Nella precedente funzione i valori di $sc, auth, hub$ vengono normalizzati in $[0, 1]$ prima del calcolo.

Vogliamo provare diverse combinazioni di tali valori per ottenere la massima precisione.

2 Risultati

Abbiamo studiato la relazione tra il grafo radice R_q e il grafo allargato B_q per capire le dinamiche dell'espansione. Per esempio Figura 1 riportano i tali grafi per la query 56. Abbiamo scelto per i parametri α, β, γ i valori 1.0, 0.1, 0.1, mentre per N abbiamo usato il valore 50. Abbiamo ottenuto la *map* riportata in Figura 2:

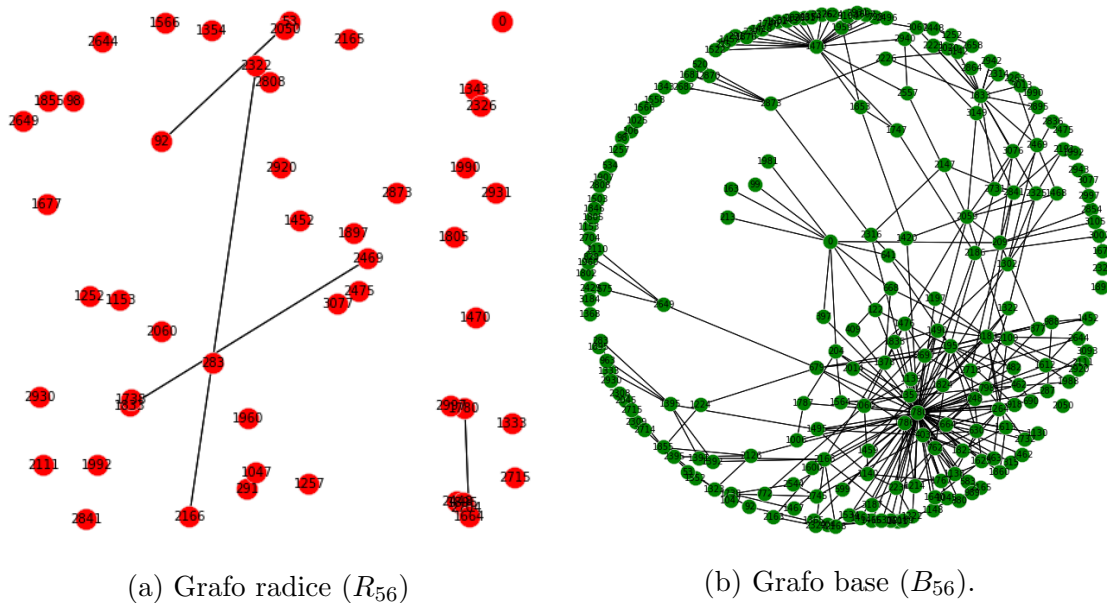


Figure 1: Grafi ottenuti applicando HITS sui top $N = 50$ documenti per la query 56.

| | | |
|-------------|-----|-----------|
| runid | all | G12R9HITS |
| num_q | all | 43 |
| num_ret | all | 38957 |
| num_rel | all | 719 |
| num_rel_ret | all | 601 |
| map | all | 0.3292 |

Figure 2: Mean Average Precision utilizzando HITS nella funzione di reperimento.