

Cognome e Nome _____

Matricola _____

Appello del 12 giugno 2025

Esercizio 1 [8 punti]

Siano date delle quintuple (**id_studente**, **id_insegnamento**, **CFU**, **voto**, **data**). Ciascuna quadrupla rappresenta il verbale di un esame svolto in data **data** dallo studente con matricola **id_studente** per l'insegnamento **id_insegnamento** da **CFU** CFU, con voto **voto**.

Si progettino (eventualmente utilizzando più step di computazione tra loro concatenati) le funzioni *Map* e *Reduce* per restituire in output **uno degli studenti che ha conseguito la media pesata (rispetto ai CFU) più alta dei voti degli esami sostenuti**, utilizzando spazio locale $O(n^{1/2})$, dove n è il numero di studenti. Si assuma che le matricole siano numeri interi da 0 a $n-1$, e che ogni studente ha svolto un numero costante di esami.

Suggerimento: si utilizzi un primo step di computazione per calcolare la media aritmetica dei voti di ciascuno studente; successivamente, si utilizzino gli step di computazione necessari a trovare uno studente con la media più alta.

Esempio di input:

- (0, 100, 20, 6, 10/04/2023)
- (1, 100, 30, 6, 07/09/2024)
- (0, 210, 28, 3, 05/07/2023)
- (1, 210, 28, 3, 10/07/2024)
- (0, 300, 24, 9, 01/07/2024)

In questo input abbiamo 5 esami.

Il suggerimento consiglia di calcolare prima le coppie (**id_studente**, **media**)

che nell'esempio risulteranno essere:

- (0, 23,33)
- (1, 29,33)

Per poi calcolare uno degli studenti che ha conseguito la media più alta (1 nell'esempio).

Esercizio alternativo [valutazione massima 4 punti]: assumere di avere in input le coppie (**id_studente**, **media**) (con **id_studente** che varia da 0 a $n-1$) e dare in output uno degli studenti con la media più alta, utilizzando spazio locale $O(n^{1/2})$.

Esercizio 2 [9 punti]

Si consideri il problema dei **Frequent Itemset**.

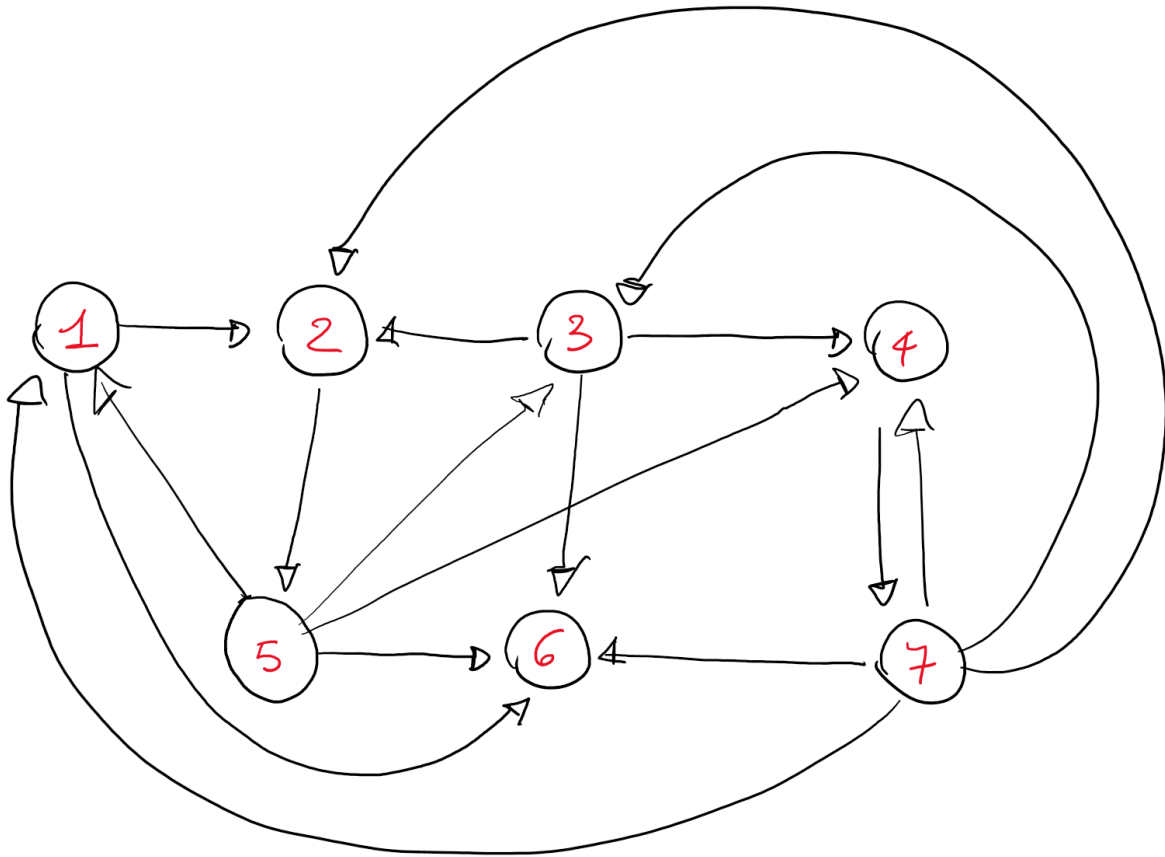
- Si definisca il **bordo negativo** e si indichi come è possibile calcolarlo
- Si descriva il funzionamento dell'**algoritmo di Tovoinen**
- Si definiscano i **falsi negativi** e i **falsi positivi**, e si discuta come vengono gestiti dall'algoritmo di Tovoinen

Esercizio 3 [8 punti]

- [4 punti]** Si dica in modo **sintetico e puntuale, giustificando la risposta**, che cosa è il PageRank e come è possibile calcolarlo in modo efficiente, spiegando in modo particolare il ruolo parametro β di teleport e cosa accadrebbe se, in assenza di dead-end, si scegliesse $\beta=1$.
- [4 punti]** Si faccia un esempio di grafo dei link del web con **7 nodi** e **12 archi**, in cui siano presenti esattamente un dead-end e uno spider trap, e si impostino le equazioni per il calcolo del **Topic-Specific PageRank** con parametro β generico, tenendo conto del fatto che l'insieme di teleport deve costituito da **3 nodi** a vostra scelta.

Esercizio 4 [9 punti]

Si consideri la rete sociale in figura.



Applicando l'algoritmo A-priori, **si individuino** i cluster di almeno 5 nodi indotti da soglia di supporto uguale almeno a 2. Si mostri ogni passo della computazione effettuata.

Regole per lo svolgimento della prova scritta:

- Per svolgere il compito si hanno a disposizione **90** minuti
- Scrivere **subito** nome, cognome, matricola su **OGNI FOGLIO (compreso questo)**.
- Durante la prova scritta **non** è possibile abbandonare l'aula.
- Non è ammesso **per nessun motivo** comunicare in qualsiasi modo con altre persone
- Non è possibile consultare appunti, libri e dispense.
- Qualsiasi strumento elettronico di calcolo o comunicazione (telefoni cellulari, calcolatrici, palmari, computer, etc...) deve essere **completamente disattivato** e **depositato in vista sulla cattedra**
- Mettere in vista sul banco un valido documento di identità.