Cognome e Nome _____

Matricola _____

Big data – Graph Mining
A.A. 2024/2025

Appello del 30 gennaio 2025

Esercizio 1 [8 punti]

Sia dato un grafo diretto di transazioni, pesato sia sui nodi che sugli archi. Un nodo corrisponde ad una persona e il suo peso al saldo iniziale della persona. Un arco diretto dal nodo i al nodo j avente peso w indica un trasferimento di w denaro dalla persona i alla persona j.

Si progettino (eventualmente utilizzando più step di computazione tra loro concatenati) le funzioni *Map* e *Reduce* per restituire in output ***il massimo saldo finale tra i saldi finali di tutte le persone del grafo*** dato un **grafo diretto pesato su nodi e archi**, utilizzando spazio locale $O(n^{1/4})$, dove n è il numero di nodi del grafo. Si assuma che i nodi siano rappresentati da numeri interi da 0 a n-1, che il nodo **i** del grafo, avente saldo iniziale **s**, sia codificato dalla coppia **(i, s)** e che un arco del grafo, che va dal nodo ***i*** al nodo ***j*** ed ha peso **w**, sia codificato dalla tripla **(i, j, w)**. Si assuma anche che ogni nodo sia coinvolto in un numero costante di transazioni.

Suggerimento: si utilizzi un primo step di computazione per calcolare il saldo finale di ogni nodo; successivamente, si utilizzino gli step di computrazione necessari a trovare il massimo saldo finale.

Esempio di input:

- (0, 10000)
- (1, 4000)
- (2, 5000)
- (0, 1, 3000)
- (2, 1, 2000)

In questo input abbiamo 3 nodi (uno per ogni coppia) e 2 archi (uno per ogni tripla).

Il suggerimento consiglia di calcolare prima le triple **(id_nodo, saldo_finale)**
che nell'esempio risulteranno essere:

- (0, 7000)
- (1, 9000)
- (2, 3000)

Per poi calcolare il massimo saldo finale (9000 nell'esempio).

Esercizio alternativo [valutazione massima 4 punti]: assumere di avere in input le coppie **(id_nodo, saldo_finale)** (con id_nodo che varia da 0 a n-1) e dare in output il saldo finale più grande, utilizzando spazio locale $O(n^{1/4})$.

Esercizio 2 (9 punti)

1. **[2 punti]** Si definisca cosa è il **supporto** di un itemset e quando un itemset può considerarsi **frequente**; si enunci la proprietà della **monotonia della frequenza**.
2. **[5 punti]** Far vedere sulla seguente istanza l'esecuzione passo passo dell'algoritmo A-priori per individuare tutti gli insiemi di item che abbiano supporto almeno uguale a **3**.
3. **[2 punti]** Si definisca il bordo negativo e lo si calcoli sulla stessa istanza

basket	item
1	1, 3, 4
2	5, 7, 8, 9, 10
3	3, 5, 6, 7, 8
4	2, 7, 8, 9
5	7, 9
6	1, 5, 6, 7
7	1, 3, 4, 5, 7, 9, 10
8	7, 8, 9
9	5, 7, 8
10	1, 2, 3, 4

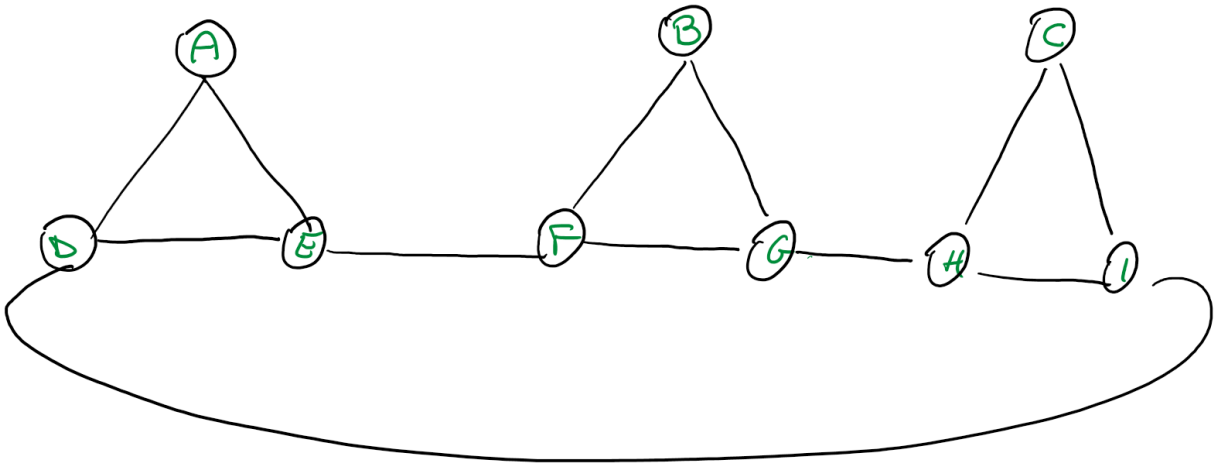
Esercizio 3 [9 punti]

a. **[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 quando, in assenza di dead-end, si sceglie **β=1**.

b. **[5 punti]** Si faccia un esempio di grafo dei link del web con **5** nodi e **10** archi, in cui sia presente esattamente un dead-end, 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 [8 punti]

Si consideri la rete sociale in figura.



Si approssimi la **edge betweeness** (con l'algoritmo di **Girvan–Newman**) assumendo di considerare solo i contributi dovuti alle visite radicate ai nodi **a**, **c, f** ed **h**.

Si discuta se tale calcolo può portare a dividere il grafo in due o più cluster, giustificando la risposta.

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.