

Cognome e Nome \_\_\_\_\_

Matricola \_\_\_\_\_

## Appello dell'8 luglio 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 suo saldo iniziale. Un arco diretto dal nodo  $i$  al nodo  $j$  avente peso  $w$  indica una transazione che comporta 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 **tutte le persone aventi come saldo finale 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  $O(n^{1/2})$ , 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 computazione 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 una persona che è associata al massimo saldo finale (1 nell'esempio, che è associata al saldo 9000).

**Esercizio alternativo [valutazione massima 6 punti]:** svolgere l'esercizio con l'unica modifica che bisogna dare in output **una qualsiasi persona avente come saldo finale il saldo finale massimo** (sempre utilizzando spazio locale  $O(n^{1/2})$ ).

**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 **una qualsiasi persona avente come saldo finale il saldo finale massimo**, utilizzando spazio locale  $O(n^{1/2})$ .

#### 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.

---

**Esercizio 2 [9 punti]**

Si considerino gli insiemi descritti nella tabella qui sotto.

- **[4 punti]** Si scriva la colonna dell'insieme  $S_2$  in modo che la Jaccard Similarity tra  $S_1$  e  $S_2$  sia  $1/2$  e la Jaccard Similarity tra  $S_2$  e  $S_4$  sia  $2/5$ .
- **[5 punti]** Si calcoli, applicando il metodo visto a lezione ed evidenziando tutti i passaggi fatti riga per riga, la matrice delle firme **minhash** degli insiemi  $S_1, S_2, S_3$  e  $S_4$  secondo le permutazioni indotte dalle 3 funzioni hash evidenziate in tabella.

Riga	$S_1$	$S_2$	$S_3$	$S_4$	$(x+5) \bmod 7$	$2x \bmod 7$	$(x-3) \bmod 7$
0	1		1	0			
1	0		1	1			
2	0		0	0			
3	1		0	1			
4	0		1	0			
5	0		0	1			
6	1		1	1			

---

**Esercizio 3 [8 punti]**

Si consideri l'algoritmo *CURE* per il clustering di punti in uno spazio euclideo con  $d$  dimensioni.

Si dica in modo  **sintetico e puntuale, giustificando le risposte**:

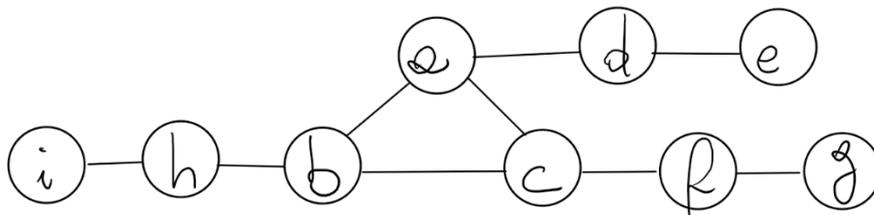
- Quali sono le assunzioni che tale algoritmo fa circa la conformazione dei cluster.
- Come viene identificato e memorizzato un cluster
- Se tale algoritmo è di tipo gerarchico e/o di tipo con assegnamento di punti

Infine, si descriva l'algoritmo evidenziando le sue varie fasi.

---

**Esercizio 4 [9 punti]**

Si consideri la rete sociale in figura.



Si calcoli la **edge betweenness** delle rete con l'algoritmo di **Girvan–Newman** (si noti la simmetria che si può sfruttare per velocizzare il calcolo).

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