

Cognome e Nome _____

Matricola _____

Appello dell'11 giugno 2026

Esercizio 1 [8 punti]

Si progettino (eventualmente utilizzando più step di computazione tra loro concatenati) le funzioni *Map* e *Reduce* per generare tutti i **cicli di lunghezza 4 con il loro relativo peso** in un **grafo diretto pesato** $G=(V,E,w)$. Si assuma che un arco del grafo di input, che va dal nodo i al nodo j ed ha peso x , sia codificato come la tripla in input (i, j, x) e che il peso di un ciclo sia dato **dal minimo peso** degli archi che lo compongono.

Si chiede di dare in output i cicli **senza ripetizioni**: ad esempio, il ciclo formato dai nodi $2 \rightarrow 4 \rightarrow 3 \rightarrow 10$ non dovrà comparire anche sottoforma di $4 \rightarrow 3 \rightarrow 10 \rightarrow 2$, $3 \rightarrow 10 \rightarrow 2 \rightarrow 4$ o $10 \rightarrow 2 \rightarrow 4 \rightarrow 3$.

Suggerimento: prima trovare i path di 2 archi (quindi con 3 nodi) e successivamente trovare i cicli di 4 archi analizzando come si possono concatenare i path di 2 archi trovati al primo step.

Esercizio alternativo [valutazione massima 4 punti]: generare, invece dei cicli di 4 nodi, tutti i path di 2 archi (e quindi con 3 nodi) con il relativo peso (ottenuto sempre come minimo peso degli archi che lo compongono).

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**.

basket	item
1	1, 2, 3, 4, 11
2	1, 2, 3, 5
3	1, 2, 4, 5
4	1, 2, 3, 4, 5
5	2, 3, 4, 6
6	1, 2, 3, 7, 12
7	1, 3, 4, 8
8	2, 3, 4, 9
9	1, 2, 3, 10, 11
10	1, 2, 4, 12

3. [2 punti] Si definisca il bordo negativo e lo si calcoli sulla stessa istanza

Esercizio 3 [8 punti]

Si consideri l'algoritmo di *Bradley, Fayyad e Reina* per il clustering di punti in uno spazio euclideo con d dimensioni.

Si dica in modo **sintetico e puntuale**, giustificando la risposta:

- L'assunzione di fondo che questo algoritmo fa sui punti di ogni cluster
- Cosa sono i *discard set*, i *compressed set* e i *retained set* e perché vengono chiamati così.
- Come è possibile ricostruire le informazioni su media e varianza dei cluster a partire dalle informazioni memorizzate per ogni cluster.
- Come è definita la distanza di Mahalanobis e che ruolo ha nell'algoritmo BFR

Esercizio 4 [9 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 quando, in assenza di dead-end, si sceglie $\beta=1$.
- [5 punti]** Si faccia un esempio di grafo dei link del web con **5** nodi e **11** 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.

Regole per lo svolgimento della prova scritta:

- Per svolgere il compito si hanno a disposizione **90** minuti (**70** minuti se non si svolge l'Esercizio 1)
- 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, tablet, computer, smartwatch, etc...) deve essere **completamente disattivato e depositato in vista sulla cattedra**
- Mettere in vista sul banco un valido documento di identità.