

Cognome e Nome _____

Matricola _____

Appello del 25 gennaio 2024

PRIMA PARTE

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 (i, j, x) e che il peso di un ciclo sia dato dalla somma dei pesi 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) con la tecnica vista a lezione e successivamente trovare i cicli di 4 archi analizzando come si possono concatenare i path di 2 archi trovati al primo step.

Esercizio 2 [9 punti]

Si consideri la variante multi-hash con due tabelle hash dell'algoritmo *Park Chen Yu*, nell'ambito del problema degli *itemset frequenti*, e se ne descriva in dettaglio il funzionamento.

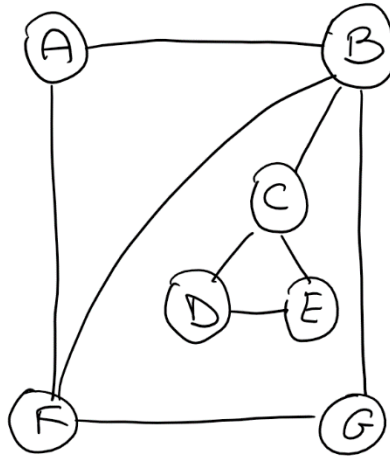
Si dica inoltre, in modo **sintetico e puntuale, giustificando la risposta**:

- cosa si può affermare quando un contatore di una delle due tabelle hash raggiunge la soglia di frequenza;
- cosa deve accadere perché una coppia venga considerata non candidata ad essere frequente;
- come avviene la compressione delle tabelle hash;
- quale vantaggio offre rispetto all'algoritmo PCY con singola tabella hash.

SECONDA PARTE

Esercizio 3 [8 punti]

Si consideri la rete sociale in figura.



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

Si discuta se e come tale calcolo può portare a dividere il grafo in cluster, giustificando la risposta.

Esercizio 4 [9 punti]

- Si dica in modo **sintetico e puntuale, giustificando la risposta**, che cosa è il PageRank e come è possibile calcolarlo, spiegando in modo particolare il ruolo parametro β di teleport e cosa accadrebbe quando $\beta=1$ (ovvero in assenza di teleport).
- Si faccia un esempio di grafo del web con **4** nodi e **7** archi, in cui sia presente un dead-end, e si impostino le equazioni per il calcolo del PageRank con parametro β generico.
- Si supponga ora di voler calcolare il Topic-Specific PageRank in cui l'insieme di teleport è costituito da due nodi a vostra scelta. Si dica quali sono i due nodi considerati e si riscrivano le equazioni per il calcolo del Topic-Specific PageRank con parametro β generico.

Regole per lo svolgimento della prova scritta:

- Per svolgere il compito si hanno a disposizione **45** minuti per la seconda parte e **90** minuti per l'intero compito. Chi non consegna entro i primi 45 minuti automaticamente rinuncia al voto dell'esonero.
- 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 lontano dalla propria postazione**.
- Mettere in vista sul banco un valido documento di identità.