DOMANDE TEORICHE

1. [**4 punti**] Definire che tipo di proprietà di anti-monotonicity per la confidenza è usata per generare efficientemente regole associative interessanti dai frequent itemset.
2. [**4 punti**] Descrivere brevemente come calcolare in MapReduce il prodotto di una matrice $n \times n$ per un vettore di lunghezza n in 2 round con local space $o(n)$
3. [**4 punti**] Dimostrare che l'algoritmo k-means termina sempre
4. [**4 punti**] Definire la nozione di *diametro* $D(G)$ di un grafo non diretto G connesso, e spiegare come ottenere una stima $\Delta \in [D(G), 2D(G)]$ in tempo lineare.

ESERCIZI

1. [**7 punti**] Sia T un dataset di transazioni sull'insieme di item I. Siano $X, Y \subseteq I$ due closed itemset e si definisca $Z = X \cap Y$. Siano $T_X, T_Y, T_Z \subseteq T$ gli insiemi di transazioni che contengono X, Y e Z, rispettivamente, e si ricordi che dalla chiusura di X e Y discende che $X = \cap_{t \in T_X} t$ e $Y = \cap_{t \in T_Y} t$. Si noti inoltre che $T_X, T_Y \subseteq T_Z$. Usando queste proprietà, dimostrare che Z è un closed itemset.

2. [**9 punti**] Sia P un insieme di N punti in $\Re^n$ e sia $\Phi^{\text{opt}}_{\text{kmeans}}(k)$ il valore ottimo della funzione obiettivo del k-means clustering. Sia A un algoritmo che, applicato a P, restituisce un k-clustering $\mathcal{C}_A$ per cui, con probabilità 1/2, vale

$$\Phi_{\text{kmeans}}(\mathcal{C}_A) \leq (\ln k)\Phi^{\text{opt}}_{\text{kmeans}}(k).$$

   Far vedere come utilizzando opportunamente A si può ottenere un'analoga qualità con probabilità almeno $1 - 1/N$.

   Suggerimento: Si utilizzi il Chernoff Bound che stabilisce che per una variabile Binomiale X di media μ, $\Pr(X \leq (1-\delta)\mu) \leq 2^{-\mu\delta^2/2}$.

TEMPO COMPLESSIVO A DISPOSIZIONE: 2 ore