

Simulazione esame – gene_small

Si consideri il database **genes_small.sql**, estratto da un database del genoma umano originariamente creato per una challenge internazionale di biological data mining. I geni codificano le proteine che si localizzano in diverse parti della cellula e interagiscono fra di loro per eseguire le funzioni vitali della cellula stessa. Il dataset contiene informazioni sulla localizzazione (tabella *classification*), sulle interazioni (tabella *interactions*) e sulle funzioni dei geni (tabella *genes*), ed è strutturato secondo il diagramma ER illustrato nella pagina seguente. Il database è scaricabile dal repository: https://github.com/TdP-datasets/genes_small

Si intende costruire un'applicazione FLET che permetta di svolgere le seguenti funzioni:

PUNTO 1

- All'avvio del programma, si crei un grafo semplice, pesato e orientato i cui vertici siano tutti i cromosomi (tabella *genes*, colonna *chromosome*, considerando solo i valori diversi da 0).
- Un arco collega due cromosomi diversi solo se i due cromosomi contengono due geni (uno per cromosoma) che compaiono (nello stesso ordine) nella tabella *interactions*. Si noti che, per ciascun cromosoma, possono esistere più geni, e ciascuno di essi potrebbe essere presente più volte (associato a *function* diverse).
- Il peso di ciascun arco dovrà essere calcolato come la somma algebrica della correlazione (tabella *interactions*, campo *Expression_Corr*), facendo attenzione a contare ogni coppia di geni una sola volta. Nell'esempio seguente, il peso dell'arco tra il cromosoma 5 ed il cromosoma 11 sarà pari a 2,468455212.

Chromosome1	Chromosome2	Gene1	Gene2	Expression_Corr
5	11	G234334	G234684	0,567936255
5	11	G235312	G234684	0,709874329
5	11	G237509	G235586	0,937991552
5	11	G238510	G239467	0,252653076

- Si visualizzi il numero di vertici ed archi del grafo, ed i valori minimo e massimo dei pesi degli archi.
- Permettere all'utente di inserire un valore soglia (*S*), verificando che tale valore sia compreso nell'intervallo minimo-massimo calcolato al punto d).
- Alla pressione del bottone "Conta archi" stampare il numero di archi il cui peso è <S, ed il numero di archi il cui peso è >S.

Esempio interfaccia grafica

Lab14-Simulazione esame gene_small

Crea Grafo

soglia

Conta Archi

Cerca cammino

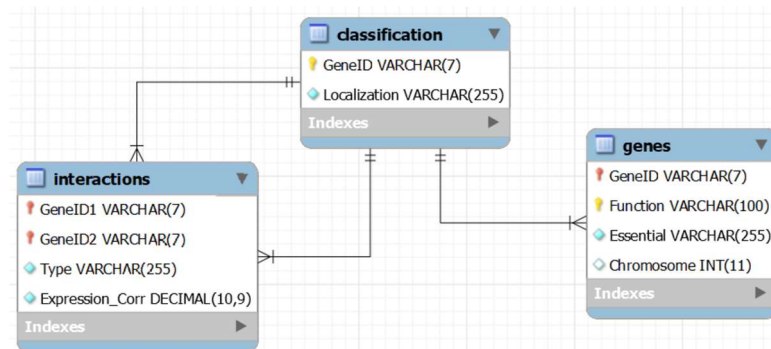
PUNTO 2

- A partire dal grafo calcolato al punto precedente, alla pressione del tasto "Ricerca Cammino", si avvia una procedura di ricerca ricorsiva per determinare il più lungo cammino di vertici (cromosomi) che sia composto esclusivamente da archi di peso >S. La lunghezza del cammino sarà valutata dalla somma dei pesi degli archi incontrati.
- Si stampi la sequenza di cromosomi di lunghezza massima così ottenuta.

Nella realizzazione del codice, si lavora a partire dalle classi e dal database contenuti nel progetto di base. È ovviamente permesso aggiungere o modificare classi e metodi.

Tutti i possibili errori di immissione, validazione dati, accesso al database, ed algoritmici devono essere gestiti, non sono ammesse eccezioni generate dal programma. Nelle pagine seguenti, sono disponibili due esempi di risultati per controllare la propria soluzione.

Le tabelle **classification**, **interactions** e **genes** contengono rispettivamente informazioni sulla localizzazione dei geni (e delle rispettive proteine), sul tipo e intensità del rapporto di interazione, e sulla funzione da loro espressa. Tutte le tabelle sono collegate tra loro tramite la chiave primaria che identifica i geni (GeneID).



ESEMPI DI RISULTATI PER CONTROLLARE LA PROPRIA SOLUZIONE:

Lab14-Simulazione esame gene_small

Crea Grafo

Numero di vertici: 16 Numero di archi: 197

Informazioni sui pesi degli archi - valore minimo: -1.26892411 e valore massimo: 7.6615387550000005

soglia

3

Conta Archi

Numero archi con peso maggiore della soglia: 21

Numero archi con peso minore della soglia: 176

Cerca cammino

Peso cammino massimo: 41.752609250000006

11 --> 2: 3.5929774639999996

2 --> 7: 3.2205624540000004

7 --> 15: 4.868331692

15 --> 4: 4.722425025