

TdP – Esame del 18/07/2024 – Traccia A

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

- Si permetta all'utente di selezionare due interi *Chromosoma min*, *Chromosoma max*, tramite i due menù a tendina corrispondenti già predisposti. I valori presenti nei menù a tendina devono rappresentare i possibili valori della variabile *Cromosoma* (tabella *genes*), e saranno usati nel punto successivo per filtrare i geni nella creazione del grafo.
Dal menù a tendina *Localization* si dovrà invece permettere all'utente di selezionare un valore tra quelli della colonna *Localization* (tabella *classification*).
Nota bene: al momento della pressione del pulsante *Crea Grafo* bisogna gestire le eccezioni nella selezione dei menu a tendina, anche il caso in cui *Chromosoma min sia maggiore di Chromosoma max*.
- Alla pressione del pulsante *Crea Grafo*, Si crei un grafo diretto e pesato con le seguenti specifiche:
 - I **nodi** sono tutti i geni contenuti nella tabella *genes* il cui Chromosoma sia tale che **Chromosoma min <= Chromosoma <= Chromosoma max**, dove Chromosoma min e Chromosoma max sono dati dall'utente tramite menù a tendina.
 - Un **arco** collega due geni diversi se e solo se i due geni hanno la **stessa Localizzazione** (tabella *classification*) e **GeneID diverso**.
 - Il **peso** dell'arco corrisponde all'indice di correlazione dell'interazione fra i due geni (tabella *interactions*).
 - Il **verso** dell'arco sarà uscente dal gene con Cromosoma minore ed entrante nel gene con Cromosoma maggiore. Il caso in cui due geni appartengono allo stesso cromosoma va gestito aggiungendo entrambi gli archi.
- Alla creazione del grafo, stampare automaticamente le seguenti informazioni (vedere screenshots di esempio):
 - I numero di vertici ed archi del grafo.
 - I 5 nodi col maggiore numero di archi uscenti ed il peso complessivo di questi archi (la somma dei loro pesi). I nodi devono essere stampati in ordine decrescente per numero di archi uscenti.
- Permettere all'utente di selezionare da un dropdown la lista delle localizzazioni disponibili nel database, e alla pressione del tasto *Dettagli* stampare tutti i geni che condividono la localizzazione scelta. (vedere screenshots di esempio).

PUNTO 2

- A partire dal grafo definito al punto 1, si vuole implementare una procedura ricorsiva che identifichi il cammino più lungo che minimizza la somma dei pesi del percorso, e con le seguenti caratteristiche:
 - Un nodo può essere attraversato una sola volta.
 - Gli archi possono essere attraversati solo nella loro direzione di percorrenza
 - Nel cammino, non ci possono essere due geni consecutivi con lo stesso valore del campo *Essential* (ovvero, si devono alternare geni *Essential* e *Non-essential*)
 - Si possono attraversare solo archi di peso crescente (non strettamente, ovvero \geq).

Nello specifico, la procedura dovrà identificare il cammino più lungo ammissibile. In seconda istanza, tra i diversi cammini di pari lunghezza, dovrà prediligere il percorso a peso totale minimo.

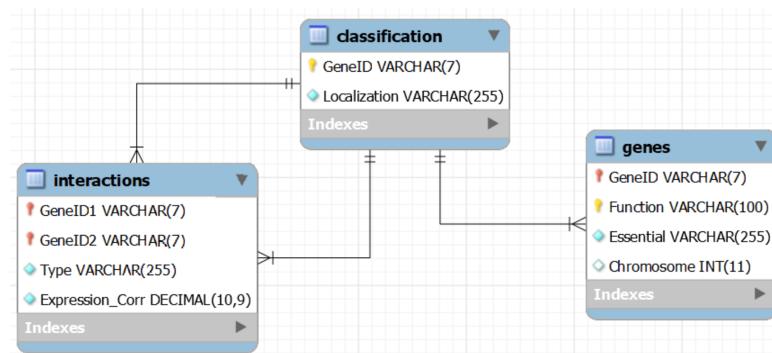
b. Si stampi:

- I. Il numero di nodi nel cammino trovato
- II. Il peso totale del cammino trovato
- III. La sequenza di nodi attraversati

Nella realizzazione del codice, si lavori 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:

TdP 2024 - Esame del 04/07/2024 - A

TdP 2024 - Esame del 04-07-2024 - A

Cromosoma min: 3 Cromosoma max: 7 Crea Grafo Localization: cytoplasm Dettagli Cammino

Created graph with 754 nodes and 768 edges

There are 525 components connected weakly

The 5 nodes with the highest number of edges are:

- G235287 - CELLULAR ORGANIZATION (proteins are localized to the corresponding organelle) | Ess.: Non-Essential, Chrom.: 4 | num. edges: 21 | total weight: 6.923854268
- G235287 - TRANSCRIPTION | Ess.: Non-Essential, Chrom.: 4 | num. edges: 21 | total weight: 6.923854268
- G235247 - CELLULAR ORGANIZATION (proteins are localized to the corresponding organelle) | Ess.: Essential, Chrom.: 4 | num. edges: 18 | total weight: 7.536789386000002
- G235247 - TRANSCRIPTION | Ess.: Essential, Chrom.: 4 | num. edges: 18 | total weight: 7.536789386000002
- G235244 - TRANSCRIPTION | Ess.: Essential, Chrom.: 4 | num. edges: 15 | total weight: 6.3706065060000014

Results point 2

TdP 2024 - Esame del 04/07/2024 - A

TdP 2024 - Esame del 04-07-2024 - A

Cromosoma min: 3 Cromosoma max: 7 Crea Grafo Localization: cytoplasm Dettagli Cammino

The genes with location cytoplasm are:

- G234212 - CELL GROWTH, CELL DIVISION AND DNA SYNTHESIS | Ess.: Non-Essential, Chrom.: 4
- G234212 - CELLULAR COMMUNICATION/SIGNAL TRANSDUCTION | Ess.: Non-Essential, Chrom.: 4
- G234212 - CELLULAR ORGANIZATION (proteins are localized to the corresponding organelle) | Ess.: Non-Essential, Chrom.: 4
- G234232 - CELLULAR ORGANIZATION (proteins are localized to the corresponding organelle) | Ess.: Non-Essential, Chrom.: 4
- G234232 - CELLULAR ORGANIZATION (proteins are localized to the corresponding organelle) | Ess.: Essential, Chrom.: 4
- G234234 - CELLULAR TRANSPORT AND TRANSPORT MECHANISMS | Ess.: Essential, Chrom.: 4
- G234234 - TRANSCRIPTION | Ess.: Essential, Chrom.: 4
- G234248 - CELL GROWTH, CELL DIVISION AND DNA SYNTHESIS | Ess.: Non-Essential, Chrom.: 4
- G234248 - CELL RESCUE, DEFENSE, CELL DEATH AND AGEING | Ess.: Non-Essential, Chrom.: 4
- G234248 - CELLULAR ORGANIZATION (proteins are localized to the corresponding organelle) | Ess.: Non-Essential, Chrom.: 4

Results point 2

TdP 2024 - Esame del 04/07/2024 - A

TdP 2024 - Esame del 04-07-2024 - A

Cromosoma min: 9 Cromosoma max: 10 Crea Grafo Localization: cytoskeleton Dettagli Cammino

Creato grafo con 196 nodi e 60 archi

Ci sono 178 componenti debolmente connesse

I 5 nodi col maggior numero di archi uscenti sono:

- G234610 - CELL RESCUE, DEFENSE, CELL DEATH AND AGEING | Ess.: Essential, Chrom.: 10 | num. archi uscenti: 4 | peso tot.: -2.919477116
- G234610 - CELLULAR ORGANIZATION (proteins are localized to the corresponding organelle) | Ess.: Essential, Chrom.: 10 | num. archi uscenti: 4 | peso tot.: -2.919477116
- G234610 - PROTEIN DESTINATION | Ess.: Essential, Chrom.: 10 | num. archi uscenti: 4 | peso tot.: -2.919477116
- G234610 - TRANSCRIPTION | Ess.: Essential, Chrom.: 10 | num. archi uscenti: 4 | peso tot.: -2.919477116
- G235518 - CELLULAR ORGANIZATION (proteins are localized to the corresponding organelle) | Ess.: Non-Essential, Chrom.: 10 | num. archi uscenti: 4 | peso tot.: -1.74975746

Risultati punto2

TdP 2024 - Esame del 04/07/2024 - A

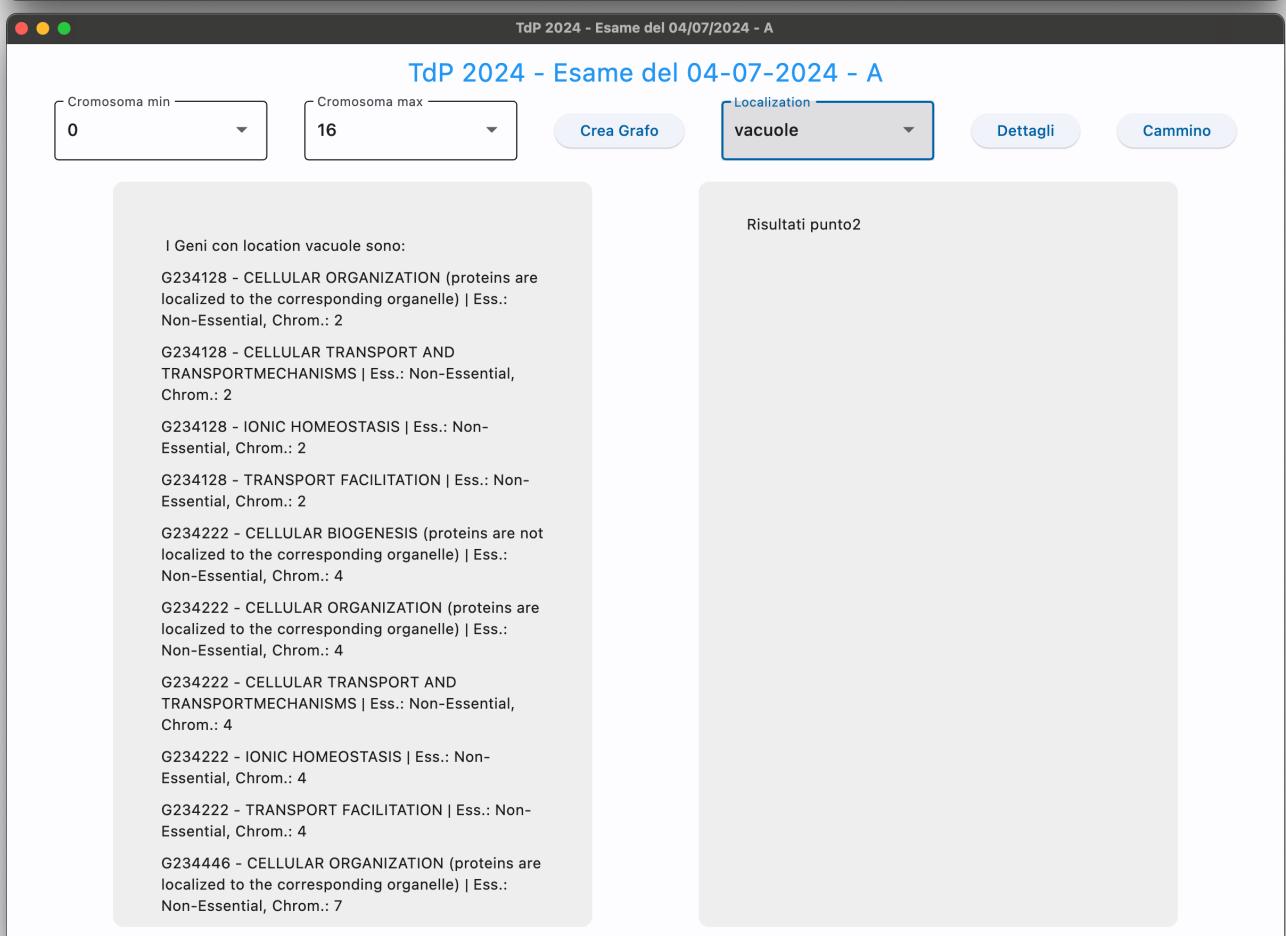
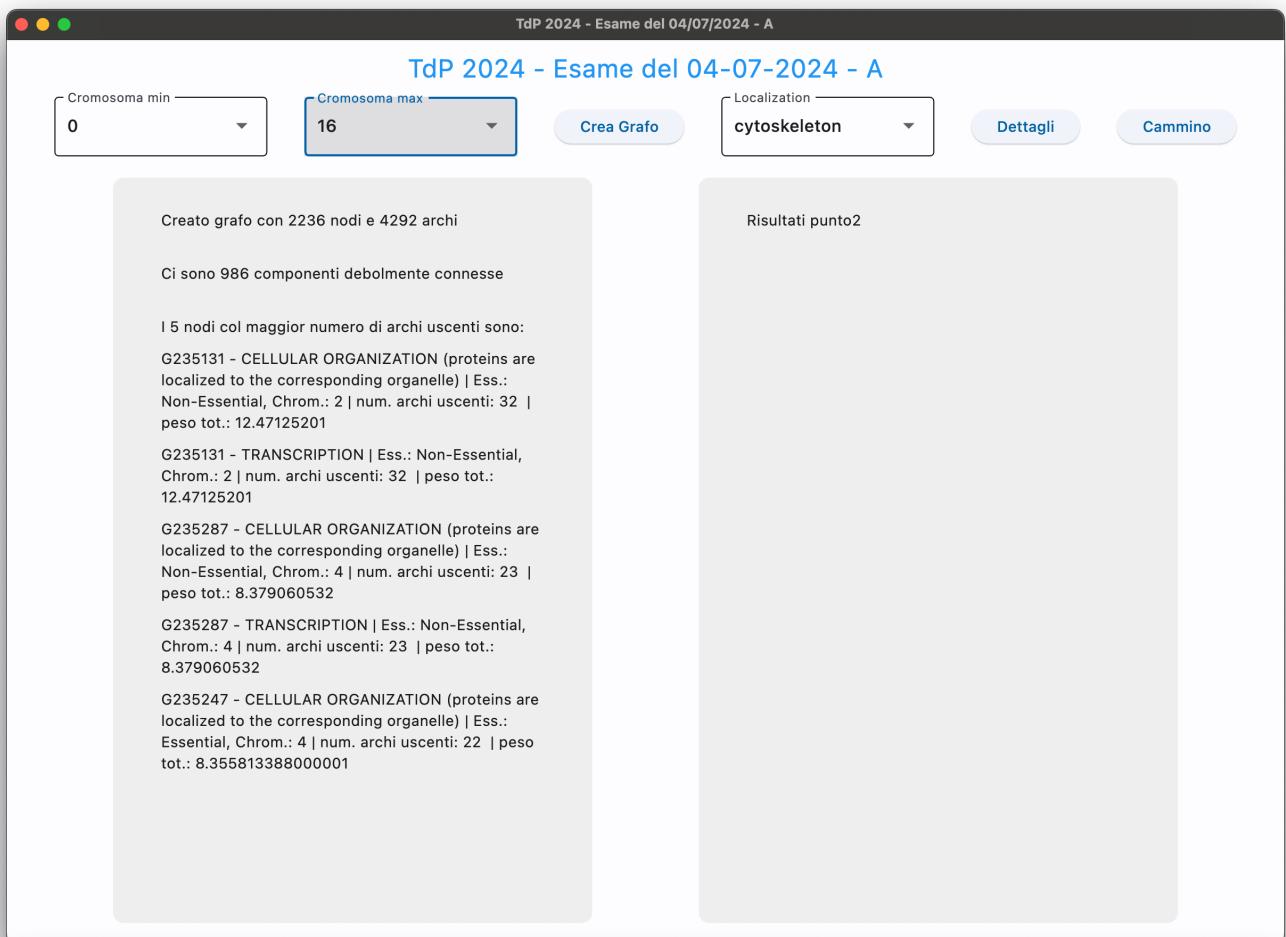
TdP 2024 - Esame del 04-07-2024 - A

Cromosoma min: 9 Cromosoma max: 10 Crea Grafo Localization: cytoskeleton Dettagli Cammino

I Geni con location cytoskeleton sono:

- G235486 - CELL GROWTH, CELL DIVISION AND DNA SYNTHESIS | Ess.: Non-Essential, Chrom.: 9
- G235486 - CELLULAR ORGANIZATION (proteins are localized to the corresponding organelle) | Ess.: Non-Essential, Chrom.: 9
- G235506 - CELLULAR ORGANIZATION (proteins are localized to the corresponding organelle) | Ess.: Essential, Chrom.: 9
- G237203 - CELL GROWTH, CELL DIVISION AND DNA SYNTHESIS | Ess.: Non-Essential, Chrom.: 9
- G237203 - CELLULAR ORGANIZATION (proteins are localized to the corresponding organelle) | Ess.: Non-Essential, Chrom.: 9

Risultati punto2



TdP 2024 - Esame del 04/07/2024 - A

TdP 2024 - Esame del 04-07-2024 - A

Cromosoma min — 10 — Cromosoma max — 10 — Localization — vacuole —

Crea Grafo **Dettagli** **Cammino**

Creata grafo con 128 nodi e 48 archi

Ci sono 116 componenti debolmente connesse

I 5 nodi col maggior numero di archi uscenti sono:

- G234610 - CELL RESCUE, DEFENSE, CELL DEATH AND AGEING | Ess.: Essential, Chrom.: 10 | num. archi uscenti: 4 | peso tot.: -2.919477116
- G234610 - CELLULAR ORGANIZATION (proteins are localized to the corresponding organelle) | Ess.: Essential, Chrom.: 10 | num. archi uscenti: 4 | peso tot.: -2.919477116
- G234610 - PROTEIN DESTINATION | Ess.: Essential, Chrom.: 10 | num. archi uscenti: 4 | peso tot.: -2.919477116
- G234610 - TRANSCRIPTION | Ess.: Essential, Chrom.: 10 | num. archi uscenti: 4 | peso tot.: -2.919477116
- G235518 - CELLULAR ORGANIZATION (proteins are localized to the corresponding organelle) | Ess.: Non-Essential, Chrom.: 10 | num. archi uscenti: 4 | peso tot.: -1.74975746

Risultati punto2

TdP 2024 - Esame del 04/07/2024 - A

TdP 2024 - Esame del 04-07-2024 - A

Cromosoma min — 10 — Cromosoma max — 10 — Localization — cell wall —

Crea Grafo **Dettagli** **Cammino**

I Geni con location cell wall sono:

- G236116 - CELL GROWTH, CELL DIVISION AND DNA SYNTHESIS | Ess.: Non-Essential, Chrom.: 10
- G236116 - CELL RESCUE, DEFENSE, CELL DEATH AND AGEING | Ess.: Non-Essential, Chrom.: 10
- G236116 - CELLULAR ORGANIZATION (proteins are localized to the corresponding organelle) | Ess.: Non-Essential, Chrom.: 10

Risultati punto2