

# TdP – Esame del 13/01/2025 – 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](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, bisogna permettere all'utente di selezionare un tipo di localizzazione tra tutti quelli presenti nel database (colonna **Localization** della tabella **classification**). I tipi di localizzazione possibili devono essere letti dal database tramite **query** e devono essere inseriti nel menu a tendina in **ordine alfabetico decrescente (Z->A)**.
- Alla pressione del pulsante **Crea Grafo**, dopo aver controllato che l'utente abbia selezionato una Localizzazione dal menu a tendina, si crei un grafo semplice (non orientato) e pesato con le seguenti caratteristiche:

- I nodi sono tutte le *Classificazioni* contenute nel database (**tabella classification**). (NB. Per il punto 2, può essere utile recuperare anche il campo sulla essenzialità (**tabella genes**)).
- Due nodi sono connessi se e solo se tra di loro esiste una interazione (**tabella interaction**).
- Il **peso** dell'arco è dato dalla **somma** di tutti i cromosomi (**colonna Chromosome della tabella genes**) associati ai *GeneID* dei due *vertici* (nodi *Classificazioni*). Nella somma, bisogna considerare i **cromosomi distinti** (ovverosia vanno sommati una sola volta).

### ESEMPIO 1:

- Vertice 1: {"G235468", "nucleus"} -> Cromosoma associato: 8
- Vertice 2: {"G237345", "nucleus"} -> Cromosoma associato: 8
- il peso tra questi due vertici è 8

### ESEMPIO 2:

- Vertice 1: {"G235677", "nucleus"} -> Cromosoma associato: 13
- Vertice 2: {"G238677", "nucleus"} -> Cromosoma associato: 4
- il peso tra questi due vertici è 17

- Stampare:
  - Il numero di nodi ed archi nel grafo
  - Gli archi, con il corrispettivo peso, **ordinati in senso crescente di peso (vedere screenshots sotto)**
- Alla pressione del pulsante **Analizza Grafo**, si stampino tutte le componenti connesse del grafo di dimensione maggiore di 1, in ordine decrescente di dimensione (**vedere screenshots sotto**). Per ogni componente connessa stampare il GeneID dei nodi nella componente e la dimensione della componente.

## PUNTO 2

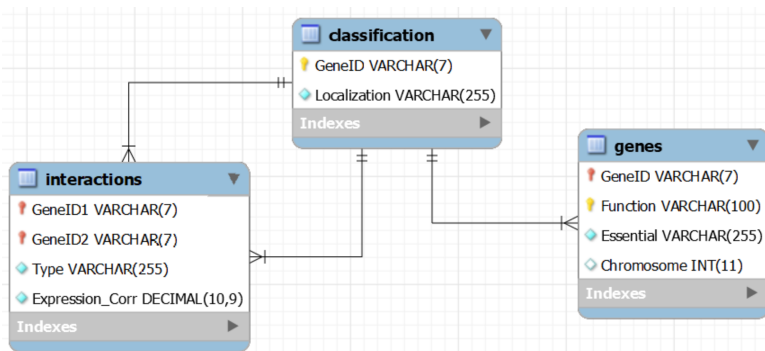
- Si implementi una procedura ricorsiva che sia in grado di identificare una lista di nodi appartenenti al grafo di cui al punto 1, tale per cui siano verificate le seguenti richieste:
  - I nodi devono essere presentati nella lista ordinati in senso crescente di GeneID
  - I nodi devono essere o tutti "Essenziali" oppure tutti "Non-Essenziali". Vanno esclusi i nodi per i quali l'essenzialità non è nota (ovvero in cui il campo è "?")

- III. La soluzione proposta deve massimizzare il numero di elementi nella lista.
  - IV. A parità di lunghezza della soluzione, il numero di componenti connesse del sottografo ottenuto considerando solo i nodi del set trovato deve essere minimo (hint: usare il metodo `Graph.subgraph()`)
- b. Si stampi la sequenza di cromosomi di lunghezza massima così ottenuta.

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 13/01/2025 - A

**TdP 2024 - Esame del 13-01-2025 - A**

Localization  
vacuole ▼

Crea Grafo   Analizza Grafo   Cammino

Creto grafo con 18 nodi e 5 archi

G235200 <-> G235234: peso 4  
G235518 <-> G235530: peso 10  
G237857 <-> G237865: peso 18  
G235625 <-> G235733: peso 25  
G235617 <-> G235832: peso 27

Le componenti connesse sono:

G235200, G235234, | dimensione componente = 2  
G235518, G235530, | dimensione componente = 2  
G235617, G235832, | dimensione componente = 2  
G235733, G235625, | dimensione componente = 2  
G237857, G237865, | dimensione componente = 2

TdP 2024 - Esame del 13/01/2025 - A

**TdP 2024 - Esame del 13-01-2025 - A**

Localization  
peroxisome ▼

Crea Grafo   Analizza Grafo   Cammino

Creto grafo con 10 nodi e 5 archi

G237892 <-> G238621: peso 11  
G237892 <-> G239393: peso 13  
G236068 <-> G238621: peso 19  
G238621 <-> G239523: peso 21  
G236068 <-> G239523: peso 26

Le componenti connesse sono:

G237892, G239393, G239523, G236068, G238621, | dimensione componente = 5

TdP 2024 - Esame del 13/01/2025 - A

### TdP 2024 - Esame del 13-01-2025 - A

Localization  
cytoplasm

Crea Grafo   Analizza Grafo   Cammino

Creto grafo con 191 nodi e 76 archi

G234065 <-> G234073: peso 1  
G234064 <-> G234126: peso 3  
G234073 <-> G234084: peso 3  
G234065 <-> G234371: peso 6  
G234073 <-> G238190: peso 6  
G234123 <-> G234248: peso 6  
G235127 <-> G235228: peso 6  
G235127 <-> G235288: peso 6  
G235143 <-> G236607: peso 6  
G234103 <-> G235336: peso 7  
G235972 <-> G237778: peso 7  
G234103 <-> G235435: peso 9  
G237509 <-> G240048: peso 9  
G234103 <-> G235474: peso 10  
G239673 <-> G240400: peso 10  
G234103 <-> G234581: peso 12  
G235301 <-> G235407: peso 12  
G235336 <-> G235363: peso 12

TdP 2024 - Esame del 13/01/2025 - A

### TdP 2024 - Esame del 13-01-2025 - A

Localization  
cytoplasm

Crea Grafo   Analizza Grafo   Cammino

Le componenti connesse sono:

G237843, G234073, G234823, G234733, G234084, G235042, G234371, G234824, G238190, G235044, G234065, G234700, | dimensione componente = 12  
G235474, G235435, G234103, G234581, G235363, G235817, G235535, G235336, | dimensione componente = 8  
G235749, G235782, G235715, G235866, G235130, G235059, G235756, | dimensione componente = 7  
G235687, G235301, G234572, G235407, G234553, G235632, | dimensione componente = 6  
G235912, G235246, G235309, G235626, G235391, | dimensione componente = 5  
G239673, G237916, G234716, G240400, | dimensione componente = 4  
G235143, G235741, G236607, G235690, | dimensione componente = 4  
G234064, G235065, G234126, | dimensione componente = 3  
G235485, G234594, G234659, | dimensione componente = 3  
G235228, G235127, G235288, | dimensione componente = 3  
G235561, G235380, G235204, | dimensione componente = 3  
G234248, G234123, | dimensione componente = 2  
G234964, G234232, | dimensione componente = 2  
G234296, G234949, | dimensione componente = 2  
G234959, G234368, | dimensione componente = 2  
G234400, G238968, | dimensione componente = 2  
G234687, G234876, | dimensione componente = 2

TdP 2024 - Esame del 13/01/2025 - A

### TdP 2024 - Esame del 13-01-2025 - A

Localization  
nucleus

Crea Grafo   Analizza Grafo   Cammino

Creto grafo con 367 nodi e 342 archi

G234106 <-> G234107: peso 2  
 G235131 <-> G235157: peso 2  
 G235133 <-> G235134: peso 2  
 G234178 <-> G235182: peso 3  
 G234194 <-> G235213: peso 4  
 G234194 <-> G235229: peso 4  
 G234240 <-> G235294: peso 4  
 G234275 <-> G235287: peso 4  
 G235189 <-> G235191: peso 4  
 G235191 <-> G237123: peso 4  
 G235244 <-> G235247: peso 4  
 G235244 <-> G235250: peso 4  
 G235244 <-> G235278: peso 4  
 G235244 <-> G235287: peso 4  
 G235247 <-> G235250: peso 4  
 G235247 <-> G235278: peso 4  
 G235247 <-> G235287: peso 4  
 G235250 <-> G235278: peso 4

TdP 2024 - Esame del 13/01/2025 - A

### TdP 2024 - Esame del 13-01-2025 - A

Localization  
nucleus

Crea Grafo   Analizza Grafo   Cammino

--- Componente ---

G235386, G235712, G235711, G235207, G236895, G235846, G235225, G235692, G235417, G235763, G237021, G238510, G235332, G239010, G236701, G235265, G234867, G238704, G236745, G234330, G235129, G235335, G236351, G234486, G234850, G235163, G234549, G235793, G239467, G239123, G235853, G238677, G235231, G235170, G235344, G235329, G239877, G235649, G234525, G235105, G234894, G235438, G235754, G234106, G234880, G235735, G238564, G235191, G234107, G239008, G235294, G239804, G234535, G235566, G235652, G235181, G235189, G234270, G235785, G235815, G235229, G235677, G234980, G235475, G236022, G235400, G234935, G239178, G235371, G234275, G235142, G235619, G238540, G234788, G235243, G234812, G235128, G234644, G235568, G237123, G235157, G235278, G235500, G235439, G235137, G235182, G235780, G235802, G234613, G238760, G234093, G235803, G234185, G235833, G235257, G235381, G235532, G235172, G235244, G235131, G235471, G235701, G236479, G235086, G236462, G235287, G238415, G239653, G238637, G234413, G239976, G235776, G235442, G235639, G235830, G234469, G235542, G234895, G234265, G235930, G234786, G235697, G235327, G234715, G235247, G234746, G235252, G235516, G239628, G235213, G235012, G234849, G235624, G238117, G234914, G234684, G234936, G235138, G235550, G234194, G234505, G238793, G235250, G235559, G235689, G235574, G235636, G235159, G234178, G237624, G235350, G235539, G235544, G234334, G234240, G235490, G235801, G235874, G235621, G234445, G235453, |  
 dimensione componente = 161

G234141, G234191, G234533, G234953, G235358, G235462, G234863, G234744, G234906, G234458, G236167, G234341, | dimensione componente = 12

G234636, G240017, G235426, G235025, G234562, G234202, | dimensione componente = 6

G235450, G238391, G235468, G238740, G237345, G236509, | dimensione componente = 6

G234085, G235865, G235573, G234444, | dimensione componente = 4

G235772, G235241, G235123, G235349, | dimensione componente = 4

G235717, G235762, G235810, G235184, | dimensione componente = 4