

## Attività – Graph Analytics per studenti della LM in Informatica

**Obiettivo** dell'attività è creare e analizzare un data graph attraverso la definizione di research question che possano essere risolte attraverso tecniche di graph analytics.

L'**attività** consisterà

- nella definizione dello schema del data graph e suo popolamento attraverso l'importazione dei dati
- nello studio di *almeno due research question* attraverso la progettazione, l'implementazione e l'esecuzione di *almeno 4 tecniche distinte e una loro interpretazione nel contesto applicativo scelto*.

Alcune precisazioni riguardo l'attività richiesta:

- Sarà necessario individuare un dataset che abbia una sua rappresentazione a grafo e che sia adeguato all'applicazione di tecniche di graph analytics. Ad esempio uno dei dataset disponibili in <https://networkrepository.com/>
- Sarà necessario definire lo schema del grafo
- Sarà necessario importare il dataset per popolare il grafo
- L'attività successiva dovrà essere sviluppata nelle seguenti fasi
  - FASE 1: Exploratory Data Analysis del data graph ovvero studio delle caratteristiche del grafo in termini di nodi e connettività. Per lo svolgimento di questa fase sarà possibile usare sia Cypher e la GDS sia codice Python attraverso pacchetti adeguati.
  - FASE 2: Per ogni research question
    - Attività di progettazione: (i) nella definizione delle proiezioni *Almeno una proiezione* dovrà essere una *Cypher Projection*; (ii) nella scelta degli algoritmi che siano adatti a risolvere la research question. Si consiglia di considerare più algoritmi. In questa fase, si potrà far uso delle funzioni di *memory estimation*.
    - Attività di implementazione: Creazione delle proiezioni, implementazione ed esecuzione degli algoritmi. Gli algoritmi potranno essere implementati sia usando gli algoritmi di GDS messi a disposizione da Neo4J sia attraverso l'esecuzione di query Cypher.
    - Interpretazione dei risultati ottenuti.

ATTENZIONE: è possibile individuare tutte le research question in partenza o partire da una research question e farsi guidare dai risultati per l'individuazione delle successive research question.

Il **risultato** dell'attività sarà un **notebook Python** contenente

- Una introduzione che descriva il data graph, lo schema, e il contesto di analisi scelto
- Il codice per la creazione del data graph e l'importazione dei dati
- Il codice relativo all'analisi esplorativa del data graph scelto e una discussione dei risultati ottenuti (output FASE 1) ;
- Per ogni research question
  - Una descrizione della research question
  - Una descrizione della soluzione proposta: graph projection e progettazione delle tecniche che dovranno essere *adeguatamente giustificata* (ovvero è necessario giustificare l'adeguatezza delle tecniche scelte per la soluzione della research question basandosi sulle caratteristiche delle tecniche stesse)
  - il codice Neo4J incluso in Python
  - i risultati ottenuti e una sua interpretazione che rappresenterà la risposta alla research question.

Le attività verranno valutate sulla base dei seguenti criteri:

- Motivazione, comprensione e story telling: L'elaborato fa credere al lettore che l'argomento sia rilevante? La parte in prosa dell'elaborato è convincente?
- Complessità del data graph e adeguatezza dello schema
- Progettazione della graph analytics e analisi dei risultati: Le proiezioni e gli algoritmi individuati sono adeguati in termini di correttezza e completezza a rispondere alle research question? L'interpretazione dei risultati risponde alle research question?
- Complessità dell'implementazione

Consegna: file zip contenente il notebook sviluppato, i file dei dati importati, un README con le istruzioni per l'esecuzione del notebook. Il file dovrà essere caricato dalla pagina di Moodle del corso seguendo il link dedicato.

Scadenza premio partecipazione: **31 maggio 2024**