

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

**Obiettivo** dell'attività è analizzare un data graph messo a disposizione da Neo4j attraverso la definizione di definizione research question che possano essere risolte attraverso tecniche di graph analytics.

L'**attività** consisterà 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 scegliere un *data graph* tra quelli elencati in <https://neo4j.com/docs/getting-started/appendix/example-data/> e non presente come Sandbox GDS
- Sarà necessario partire dalla Blank Sandbox - Graph Data Science o installare Neo4J desktop sul proprio PC e accedere al data graph scelto in remoto o creare il data graph direttamente nella Sandbox seguente le indicazioni presenti nella pagina (ad esempio per i dataset presenti nel Neo4j dataset demo server sarà possibile collegarsi in remoto al server `neo4j+s://demo.neo4jlabs.com` fornendo il nome e le credenziali per il dataset scelto)
- L'attività 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. In questa fase, si potrà far uso delle funzioni di *memory estimation*.
  - 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.
    - 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 documento contenente

- Una introduzione che descriva il data graph e il contesto di analisi scelto
- Un'analisi esplorativa del data graph scelto attraverso misure statistiche adeguate (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 delle query eseguite sulla Sandbox
  - 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?
- 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 pdf dell'attività svolta e eventuale codice prodotto in un unico zip. Il file dovrà essere caricato dalla pagina di Moodle del corso seguendo il link dedicato.

Scadenza premio partecipazione: **31 maggio 2024**