

Laboratorio 7 - Analítica de Aeropuertos

Profesor: Sebastián Ferrada

Para este laboratorio usted no solo deberá responder a las preguntas, sino que también deberá instalar Neo4j, configurarlo e importar los datos. Para lograr esto, siga los siguientes pasos:

1. Instale Neo4j Desktop, el cual puede descargar desde acá: <https://neo4j.com/deployment-center/?desktop-gdb>.
2. Cree una instancia local e instálele el plugin Graph Data Science. Instrucciones detalladas acá: <https://neo4j.com/docs/graph-data-science/current/installation/neo4j-desktop/>.
3. Descargue los dos CSV de este laboratorio desde material docente.
4. En Neo4j, haga click en “Import” y seleccione ambos archivos a la vez.
5. Cree un nodo con etiqueta “Airport”. En File, seleccione “airports.csv”. Luego, en Properties, haga click en el botón “Map from File” y seleccione todas las columnas. Marque la columna “IATA” como llave.
6. Con el mouse, dibuje una arista que salga y entre a Airport. En relationship type, agregue el nombre “vuelo”. En File, elija “routes.csv”. Luego, en Node ID Mapping, seleccione “Source Airport” en From y “Destination Airport” en To.
7. Haga click en el botón “Run Import”. Se debieron haber creado 6.072 nodos y 37.042 aristas.

Ahora responda las siguientes preguntas:

- P1.** 8 PUNTOS Encuentre el aeropuerto con mayor grado en la base de datos.
- P2.** 13 PUNTOS Proyecte un grafo que mantenga solo los aeropuertos que están en Estados Unidos y las rutas que existen entre ellos (ojo, los datos están en Inglés). Luego, calcule la centralidad Closeness. En este cálculo, retorne el nombre del aeropuerto y su ciudad. Ordene los resultados descendientemente según closeness. Opcionalmente mire el mapa y vea si sus resultados tienen sentido. Luego, calcule page rank y proyecte los resultados de la misma manera.
- P3.** 13 PUNTOS Ahora proyecte el grafo completo para hacer detección de comunidades con el algoritmo de Leiden (<https://neo4j.com/docs/graph-data-science/current/algorithms/leiden/>). Recuerde que este algoritmo solo considera grafos no dirigidos. Proyecte las ciudades, países y número de comunidad de cada aeropuerto. Ordene sus resultados por nombre de país, para ver que los resultados tengan sentido.
- P4.** 13 PUNTOS Usando búsqueda por amplitud (<https://neo4j.com/docs/graph-data-science/current/algorithms/bfs/>), retorne los países que se pueden alcanzar desde el aeropuerto JFK, con a lo más 4 vuelos.

- P5.** 13 PUNTOS Encuentre los 10 aeropuertos que tengan la mayor similitud topológica al aeropuerto de Santiago (SCL). Para esto, utilice similitud con filtros (<https://neo4j.com/docs/graph-data-science/current/algorithms/filtered-node-similarity/>), con los parámetros “sourceNodeFilter” y “topK”.