

## EXAMEN PARCIAL DE MINERÍA DE GRAFOS

Nombre	Cordero Hernández, Marco R.	Expediente	727272	Carrera	Ingeniería en Sistemas Computacionales	Fecha	2 de octubre del 2023
--------	--------------------------------	------------	--------	---------	---	-------	--------------------------

Yo, Marco Ricardo Cordero Hernández, al participar en este examen me comprometo a que lo entregado sea fruto solamente de mi trabajo, como una oportunidad de demostrar lo que he aprendido hasta el día de hoy. Además, reconozco que la copia es una falta a la honestidad académica y personal. Por lo que me comprometo a no incurrir en acciones que impliquen copiar o dejar que alguien me copie, ni participar en cualquier tipo de intercambio de información por cualquier medio con los estudiantes inscritos a este curso o personas ajenas al mismo.

Lee todo el documento, antes de iniciar a responderlo. A partir del momento en el que inicias el examen tienes 2 horas para completarlo. Al subir a Canvas los entregables se registrará la hora. ¡Mucha suerte!

El archivo **Country-nodes.csv** contiene la información de algunos de los principales países del mundo. El **Country-edges.csv** contiene la información de la relación ficticia entre los países, incluyendo el peso. Utilizando la información de los dos archivos, crea un grafo dirigido con peso en una base de datos en NEO4J.

El nombre de la base de datos y la contraseña deben contener las iniciales de tu carrera y tu número de expediente.

Puedes utilizar cualquiera de las herramientas vistas en clase para completar las actividades, pero cada respuesta debe tener una evidencia, por ejemplo, código cypher, tablas, imágenes de grafos, etc.

### 1. Crea el grafo

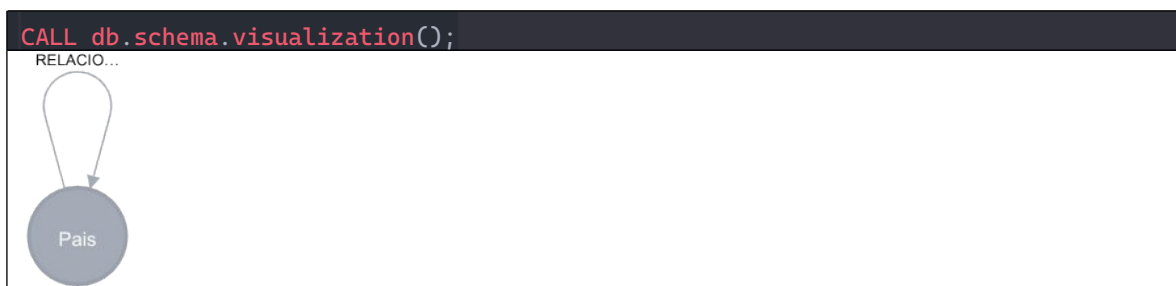
#### 1.1. Código cypher que te permita crear nodos atributos y sus relaciones dirigidas con peso [8 puntos]

```
// Importar nodos
LOAD CSV WITH HEADERS FROM "file:///Country-nodes.csv" AS row
CREATE (p:País)
SET p = row;

// Importar relaciones y pesos
LOAD CSV FROM "file:///Country-edges.csv" AS row
MATCH (p:País), (q:País)
WHERE p.Id = row[0] and q.Id = row[1]
MERGE (p)-[:RELACION_CON {Weight: toInteger(row[2])}]->(q);
```

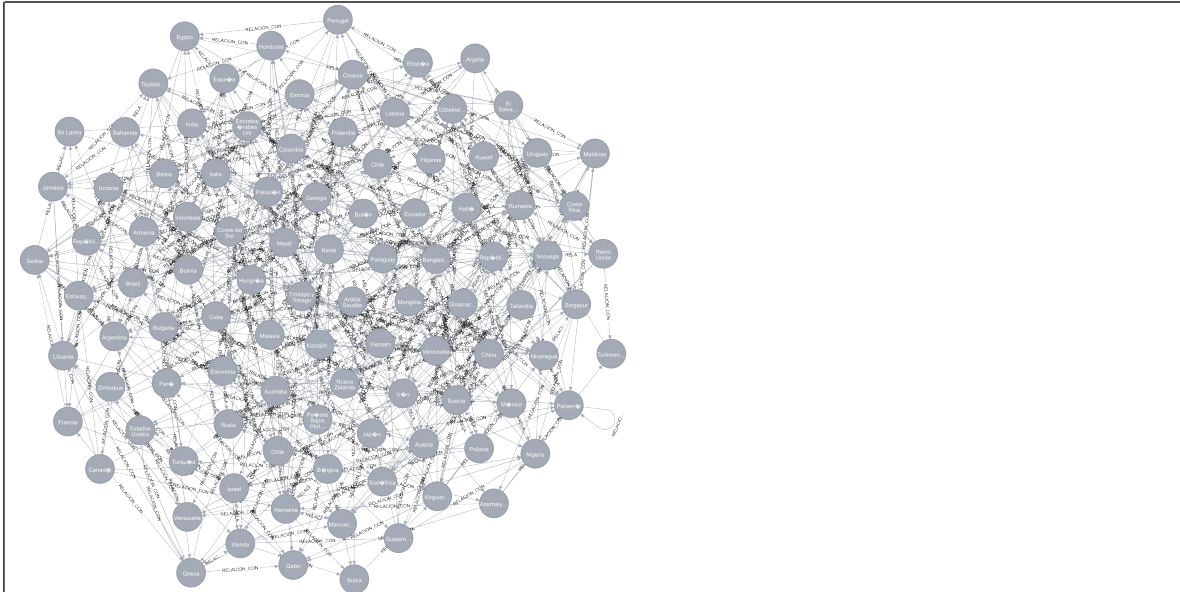
### 2. Describe el grafo

#### 2.1. Una imagen del esquema general [3 puntos]



#### 2.2. Una imagen del grafo completo [3 puntos]

```
MATCH p=()-[]-() RETURN p;
```



2.3. El número de nodos y atributos de cada nodo [2 puntos]

```
MATCH (p:Pais) RETURN COUNT(p) AS totalNodos;
CALL db.schema.nodeTypeProperties()
YIELD nodeType AS Nodo, propertyName AS Propiedad, propertyTypes AS Tipo;
```

	Nodo	Propiedad	Tipo
totalNodos	":Pais"	"Id"	["String"]
100	":Pais"	"Label"	["String"]

2.4. El número de relaciones y atributos de estas [2 puntos]

```
MATCH ()-[r]->() RETURN COUNT(r) as totalRelaciones;
CALL db.schema.relTypeProperties()
YIELD relType AS Relacion, propertyName AS Propiedad, propertyTypes AS Tipo;
```

	Relacion	Propiedad	Tipo
totalRelaciones	":RELACION_CON"	"Weight"	["Long"]
549	":RELACION_CON"	"Weight"	["Long"]

2.5. ¿Cuál es el diámetro del grafo y que nos indica? [4 puntos]

```
MATCH (start), (end)
WHERE start < end
WITH shortestPath((start)-[*]->(end)) AS p
WITH LENGTH(p) AS d
WHERE d IS NOT NULL
RETURN d AS diametroDirigido
ORDER BY d DESC
LIMIT 1;
```

diametroDirigido
6

El diámetro de este grafo indica que habría que recorrer por lo menos 6 países desde un país inicial hacia todos los demás para

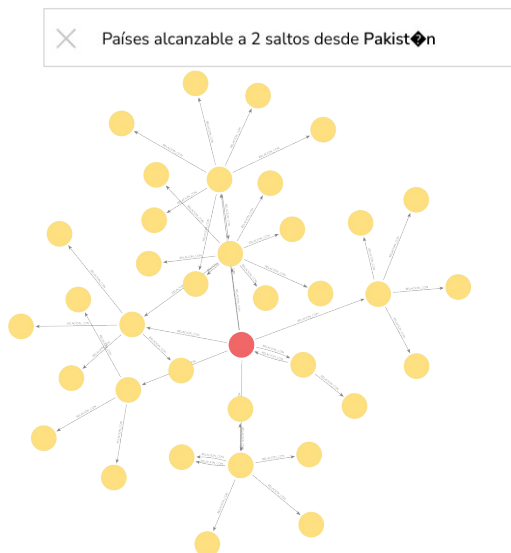
lograr obtener todas las relaciones entre ellos.

3. Crea un tablero que contenga
  - 3.1. El top 3 *degree* [6 puntos]
  - 3.2. El top 3 *in\_degree*, *out\_degree* [6 puntos]
  - 3.3. El top 3 *Closeness Centrality* [6 puntos]
  - 3.4. El top 3 *PageRank* [6 puntos]
  - 3.5. El top 3 *Betweenness Centrality* [6 puntos]

**Medidas de centralidad de países**

Top Degree		Top InDeg		Top OutDeg		Top Closeness		Top PageRank		Top Betweenness	
Pais	DEGREE	Pais	IN_DEGREE	Pais	OUT_DEGREE	Pais	Closeness	Pais	PageRank	Pais	Betweenness
Croacia	20	Emiratos Arabes Unidos	11	Bélgica	11	Australia	0.414	Pakistán	2.125	Italia	567.053
Rumania	17	Noruega	10	Croacia	11	Pakistán	0.409	Emiratos Arabes Unidos	2.082	Croacia	503.621
Noruega	17	Irán	10	Bolivia	10	Hungría	0.406	Lituania	1.84	Emiratos Arabes Unidos	502.761

4. Utilizando Bloom
  - 4.1. Crea una consulta personalizada para que reciba el nombre o identificador de un nodo y regrese el grafo de sus vecinos a dos saltos. Prueba tu consulta con el nodo que tenga mayor *PageRank* [8 puntos]



5. Con base en lo que sabes de los algoritmos de centralidad, responde las siguientes preguntas
  - 5.1. Una consulta con cypher que muestre el nodo con mayor *out\_degree*, el número de relaciones que tiene a un salto y como una colección todos los nodos que están conectados a él [8 puntos]

```
MATCH (m)-[r]->(n)
RETURN m.Label AS Nodo,
       COUNT(r) AS OUT_DEGREE,
       COLLECT(n.Label) AS vecinos1Salto
```

```
ORDER BY OUT_DEGREE DESC
LIMIT 1;
```

Nodo	OUT_DEGREE	vecinos1Salto
"Venezuela"	12	["Tailandia", "Estados Unidos", "Armenia", "Australia", "China", "Argentina", "México", "Rumania", "Nigeria", "Perú", "Vietnam", "Bulgaria"]

5.2. ¿Cuáles serían los 5 principales nodos que tienen más calidad conexiones con otros y de qué te servirían? [8 puntos]

```
// PageRank
MATCH (p:Pais)
RETURN p.Label AS Pais
ORDER BY p.PageRank DESC
LIMIT 5;
// Por medio de estos países, tratados internacionales serían posibles y las relaciones políticas serían fructíferas
```

Pais	
"Pakistan"	
"Emiratos Árabes Unidos"	
"Lituania"	
"Australia"	
"Austria"	

5.3. ¿Cuáles serían los 5 principales nodos que sirven como puente esencial para que continúe el flujo de información en el grafo y por qué? [8 puntos]

```
// Betweenness
MATCH (p:Pais)
RETURN p.Label AS Pais
ORDER BY p.Betweenness DESC
LIMIT 5;
// La posición geopolítica planeada estratégicamente para este ejercicio provee un panorama en donde los países mostrados conectan varias naciones, contando con varios de los idiomas principales del mundo, impacto cultural y social
```

Pais	
"Italia"	
"Croacia"	
"Emiratos Árabes Unidos"	
"Pakistan"	
"Alemania"	

5.4. ¿Cuáles serían los 5 principales nodos que propondrías para distribuir un mensaje y por qué? [8 puntos]

```
// Closeness
MATCH (p:Pais)
RETURN p.Label AS Pais
ORDER BY p.Closeness DESC
LIMIT 5;
// La densidad de las poblaciones mostradas, así como su espacio geográfico
indica, hacen posible una rápida distribución de mensajes/sucesos a través del
mundo. De igual manera, esta información podría resultar relevante en el contexto
de las telecomunicaciones.
```

Pais	
"Australia"	
"Pakist🎱n"	
"Hungre🎱a"	
"Emiratos 🎱rabes Unidos"	
"India"	

6. Menciona dos ejemplos de tipos de problemas que podrían solucionar con grafos (base de datos basadas en grafos) y dos que sería mejor solucionarlos usando otra tecnología [8 puntos]
- 6.1. **Ideal para grafos:** Comportamiento de masas relacionadas entre sí (sociedades, sindicatos, alumnos, habitantes) sobre las cuales se deseen realizar análisis complejos en base a sus relaciones; Conexiones en general, en donde un punto A se conecte con B, y a la vez su conexión sea relevante en el análisis. Como ejemplo concreto, las redes de comunicaciones alámbricas y la telemetría que su uso arroja como latencia y cargas puntuales pueden ser representadas a través de nodos y relaciones para determinar si su topología es correcta o podría mejorar.
  - 6.2. **No ideal para grafos:** Almacenamiento masivo de amplios volúmenes de datos no estructurados como textos o imágenes, de hecho, cualquier problema cuya solución se encuentre en la eficiencia de recuperación de recursos sobre su relación con los demás datos dentro de la base; Problemas relacionados con la estructuración rígida de datos relacionados por mecanismos ajenos a relaciones desde punto A hacia B, y que además se requiera la búsqueda dentro de múltiples campos de las colecciones para retornar resultados.