**Projecte de Neo4j: Padrons**

Bases de Dades no Relacionals

Grau en Matemàtica Computacional i Analítica de Dades

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Víctor Benito Segura (1597165)

Mireia Majó i Cornet (1597716)

Àlex Martín González (1605489)

Albert Roca Llevadot (1603375)

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

****

**Índex**

[Introducció 2](#c768yejhxza2)

[Treball en equip 2](#uvxrqavx5v6z)

[Treball previ 3](#rd08j7v6azq6)

Punt de partida [3](#vp09avcbq93j)

[Creació del repositori de Github 3](#bywo88ojh97k)

Exercicis  [4](#th4c5tu30mn2)

Exercici 1 4

Exercici 2 5

Exercici 3 5

**Introducció**

TODO

**Treball en equip**

TODO

* Albert: 2, 6, 10
* Mireia: 3, 7, 11
* Àlex: 4, 8, 12
* Víctor: 5, 9, 13

**Treball previ**

TODO

**Punt de partida**

TODO

**Creació del repositori de GitHub**

Per concloure aquest apartat de treball previ, crearem un repositori a GitHub per tal d’anar actualitzant l’estat del projecte. En aquest repositori, realitzarem els commits de les diferents tasques que se’ns proposen:

TODO

[Enllaç per accedir al repositori de GitHub](https://github.com/Victorbs18/Neo4j)

**Exercicis**

**Exercici 1. Importa les dades en la BD de Neo4j del projecte. Genera un script en cypher que carregui totes les dades, generi tots els nodes, relacions i afegeixi les característiques allà on toqui.**

TODO

**Exercici 2. Resoleu les següents consultes:**

1. Del padró de 1866 de Castellví de Rosanes (CR), retorna el número d'habitants i la llista de cognoms, sense eliminar duplicats.

| MATCH (i:Individu)-[v:viu {Any\_Padro: 1866, Municipi: "CR"}]->(h:Habitatge)  WITH count(i) AS Recompte, collect(i.Cognom1) AS Cognom  RETURN Recompte, Cognom |
| --- |

1. Per a cada padró de Sant Feliu de Llobregat (SFLL), retorna l’any de padró, el número d'habitants, i la llista de cognoms. Elimina duplicats i “nan”.

| MATCH (i:Individu)-[v:viu]->(h:Habitatge {Municipi:"SFLL"})  WITH count(i) AS N, h AS h, collect(distinct i.Cognom1) AS Cognom  RETURN h.Any\_Padro, N, Cognom |
| --- |

1. Dels padrons de Sant Feliu de Llobregat (SFLL) d’entre 1800 i 1845 (no inclosos), retorna la població, l'any del padró i la llista d'identificadors dels habitatges de cada padró. Ordena els resultats per l'any de padró.

| MATCH (i:Individu)-[v:viu]->(h:Habitatge)  WHERE h.Municipi = 'SFLL' AND i.Any\_Padro > 1800 AND i.Any\_Padro < 1845  RETURN COUNT(DISTINCT i) AS Poblacio, i.Any\_Padro AS Any\_Padro, COLLECT(DISTINCT h.Id\_Llar) AS Habitatges  ORDER BY i.Any\_Padro |
| --- |

1. Retorna el nom de les persones que vivien al mateix habitatge que "rafel marti" (no té segon cognom) segons el padró de 1838 de Sant Feliu de Llobregat (SFLL). Retorna la informació en mode graf i mode llista.

|  |
| --- |

1. Retorna totes les aparicions de "miguel estape bofill". Fes servir la relació SAME\_AS per poder retornar totes les instancies, independentment de si hi ha variacions lèxiques (ex. diferents formes d'escriure el seu nom/cognoms). Mostra la informació en forma de subgraf.

| MATCH (i:Individu)-[s:SAME\_AS]-(r:Individu)  WHERE i.Nom = "miguel" AND i.Cognom1 = "estape" AND i.Cognom2 = "bofill"  RETURN i, s, r |
| --- |

1. De la consulta anterior, retorna la informació en forma de taula: el nom, la llista de cognoms i la llista de segon cognom (elimina duplicats).

| MATCH (i:Individu)-[s:SAME\_AS]-(r:Individu)  WHERE i.Nom = "miguel" AND i.Cognom1 = "estape" AND i.Cognom2 = "bofill"  RETURN i.Nom AS Nom, COLLECT(DISTINCT r.Cognom1) AS Cognom1, COLLECT(DISTINCT r.Cognom2) AS Cognom2 |
| --- |

1. Mostra totes les persones relacionades amb "benito julivert". Mostra la informació en forma de taula: el nom, cognom1, cognom2, i tipus de relació.

| MATCH (i:Individu {Nom: 'benito', Cognom1: 'julivert'})  OPTIONAL MATCH (i)-[f:familia]->(p:Individu)  RETURN p.Nom AS Nom, p.Cognom1 AS Cognom1, p.Cognom2 AS Cognom2, f.Relacio AS tipus\_relació |
| --- |

1. De la consulta anterior, mostra ara només els fills o filles de "benito julivert". Ordena els resultats alfabèticament per nom.

|  |
| --- |

1. Llisteu totes les relacions familiars que hi ha.

| MATCH (I:Individu)-[f:familia]->(s:Individu)  RETURN I.Nom AS individu\_1, f.Relacio AS relacio\_familiar, s.Nom AS individu\_2 |
| --- |

1. Identifiqueu els nodes que representen el mateix habitatge (carrer i numero) al llarg dels padrons de Sant Feliu del Llobregat (SFLL). Seleccioneu només els habitatges que tinguin totes dues informacions (carrer i numero). Per a cada habitatge, retorneu el carrer i número, el nombre total de padrons on apareix, el llistat d’anys dels padrons i el llistat de les Ids de les llars (eviteu duplicats). Ordeneu de més a menys segons el total de padrons i mostreu-ne els 15 primers.

| MATCH (h:Habitatge {Municipi:"SFLL"})  RETURN h.Carrer AS carrer, h.Numero AS numero, count(h) AS n\_padrons, COLLECT (h.Any\_Padro) as any  ORDER BY n\_padrons DESC |
| --- |

1. Mostreu les famílies de Castellví de Rosanes amb més de 3 fills. Mostreu el nom i cognoms del cap de família i el nombre de fills. Ordeneu-les pel nombre de fills fins a un límit de 20, de més a menys.

| MATCH (i:Individu)-[f:familia {Relacio: 'jefe'}]->(i)  WHERE EXISTS((i)-[:viu]->(:Habitatge {Municipi: 'CR'}))  WITH i, size([(i)-[:familia {Relacio: 'fill'}]->() | 1]) + size([(i)-[:familia {Relacio: 'filla'}]->() | 1]) AS num\_fills  WHERE num\_fills > 3  RETURN i.Nom AS Nom, i.Cognom1 AS Cognom1, i.Cognom2 AS Cognom2, num\_fills  ORDER BY num\_fills DESC  LIMIT 20 |
| --- |

1. Mitja de fills a Sant Feliu del Llobregat l’any 1881 per família. Mostreu el total de fills, el nombre d’habitatges i la mitja de fills per habitatge. Fes servir CALL per obtenir el nombre de llars.

|  |
| --- |

1. Per cada padró/any de Sant Feliu de Llobregat, mostra el carrer amb menys habitants i el nombre d’habitants en aquell carrer. Fes servir la funció min() i CALL per obtenir el nombre mínim d’habitants. Ordena els resultats per any de forma ascendent.

|  |
| --- |

**Exercici 3. En aquest exercici analitzarem les dades del graf per entendre millor l’estructura de les dades.**

1. **Estudi de les components connexes (cc) i de l’estructura de les component en funció de la seva mida.**
2. **Semblança entre els nodes. Ens interessa saber quins nodes són semblants com a pas previ a identificar els individus que són el mateix (i unirem amb una aresta de tipus SAME\_AS).**