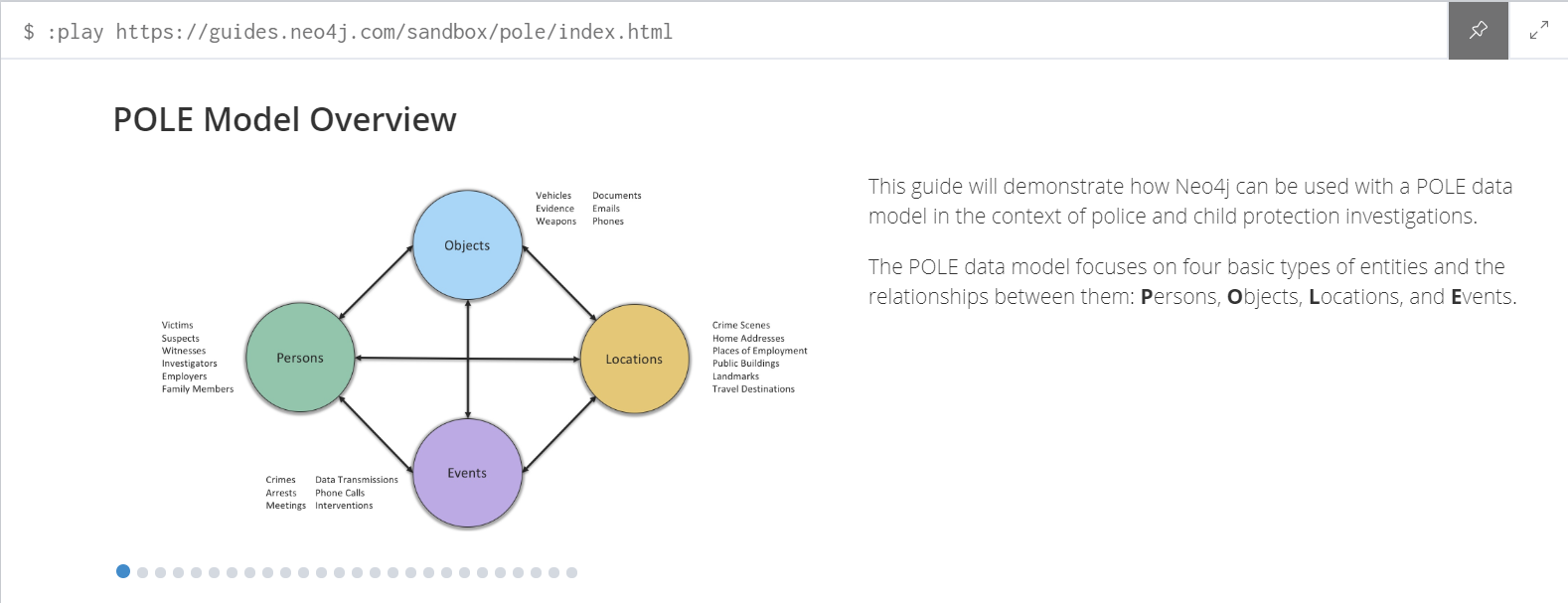
**Neo4j Guided Study**

Individual Assignment: 25 points

This is an individual assignment. It is a series of questions that can be answered using the Neo4j Crime Investigation sandbox. [(https://sandbox.neo4j.com/)](https://sandbox.neo4j.com/)

The Crime Investigation sandbox has 26 pages (unfortunately not numbered, represented by dots at the bottom left of the screen.) Read about and execute all the queries on each page. Pin the tutorial to the top of the screen using the thumbtack icon. If you don’t pin it, it will be pushed down in the viewing area when you are executing queries.



**Part 1: Investigating Crimes**

1.What do each of the letters stand for:

P: Persons

O: Objects

L: Locations

E: Events

2. Give two examples of each (e.g., a person can be a witness):

P: Victims, Suspects

O: Documents, Vehicles

L: Crime Scenes, Landmarks

E: Phone Calls, Interventions

3. Besides policing, investigation, and security, name three use cases for the POLE model:

a. Border Control

b. Missing Persons

c. Insurance Fraud Investigations

4. What aspects of the POLE dataset are real?

The L and E are real. (Location and Events)

5. What aspects of the POLE dataset are fictitious?

P and O (Persons and Objects).

* Execute the query on the 4th page. Drag the nodes around until you can see all of them in the viewing area.

6. Execute the node profiling query from the lecture notes (repeated here):

MATCH (n)

RETURN labels(n) AS labels, keys(n) AS keys, count(\*) AS total

ORDER BY total DESC;

a. Modify this query to return only *Crime* nodes. Give the modified query here:

MATCH (n: Crime)

RETURN labels(n) AS labels, keys(n) AS keys, count(\*) AS total

ORDER BY total DESC;

b. Execute the query. Each row is tells you the different properties that *Crime* nodes have and how many nodes of that type exist in the database. How many different kinds of *Crime* nodes are there?

9 Kinds

c. What are the properties of the *Crime* node that occur most frequently? How many instances are in the dataset for this type of node?

["date", "id", "type", "last\_outcome"] - 12,079 instances

7. Execute the relationship profiling query from the lecture notes (repeated here):

MATCH (m)-[r]->(n)

RETURN labels(m), type(r), keys(r), labels(n), count(\*) AS total

ORDER BY total DESC

List the pairs of relationships that have the same total number of occurrences (and how many occurrences). In other words, the total column has the same value for some relationships: give those relationships and what the count is.

[Crime OCCURRED\_AT Location] and [Crime INVESTIGATED\_BY Officer] – 28762 total

[Location HAS\_POSTCODE PostCode] and [Location LOCATION\_IN\_AREA Area] – 14904 total

[PhoneCall CALLED Phone] and [PhoneCall CALLER Phone] – 534 total

[Person HAS\_PHONE Phone] and [Person HAS\_EMAIL Email] - 328 total

8. What are the relationships between *Person* nodes?

KNOWS\_SN

FAMILY\_REL

KNOWS\_PHONE

KNOWS\_LW

9. [p. 5] What are the top 5 crimes in the dataset?

Violence and Sexual Offences

10. Execute the query on p. 6.

a. Modify it to give the top locations for "Robbery" type crimes (give the query):

MATCH (c:Crime)-[:OCCURRED\_AT]->(L:Location)

WHERE c.type = 'Robbery'

RETURN L.address, COUNT(L.address) as clocations

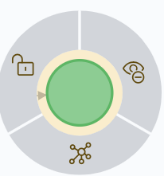
ORDER BY clocations DESC

b. What are the top 3 locations for robberies?

1. Parking Area
2. Piccadilly
3. Shopping Area

11. Execute the query on p. 7. Looking at the query answer, which crime occurs most frequently in the *crime\_type* arrays? How many times? (just scan it visually and count it up manually)

Public Order – 5



12. Execute the query on p. 8. Click on a green *Crime* node. A gray circle will appear surrounding the node. Click on the graph icon at the bottom. Which relationship do you see connected to the crime node, and what kind of node does it connect to?

OCURRED AT

Location Node

13. Continue to read and execute the queries up to p. 10. After executing the social group query, modify it to find another social group that is under investigation, but do not limit the *type* to drugs or a particular *Officer* (remove two of the filtering conditions).

a. Give your modified query here.

MATCH (c:Crime {last\_outcome: 'Under investigation'}),

(c)<-[:PARTY\_TO]-(p:Person)

WITH COLLECT(p) AS persons

UNWIND persons AS p1

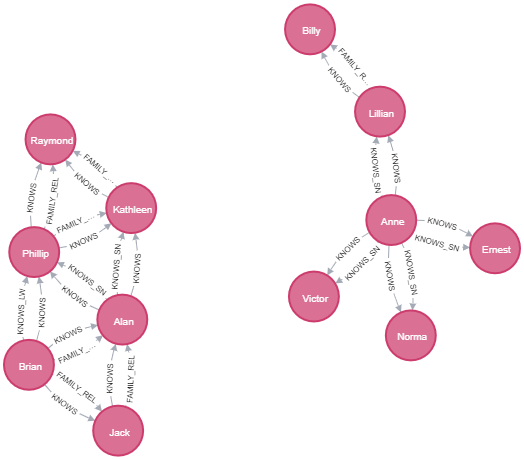
UNWIND persons AS p2

WITH \* WHERE id(p1) < id(p2)

MATCH path = allshortestpaths((p1)-[:KNOWS|KNOWS\_LW|KNOWS\_SN|FAMILY\_REL|KNOWS\_PHONE\*..3]-(p2))

RETURN path

b. Show a screen capture of the graph that is different from the one obtained by executing the unmodified query.



**Part 2: Vulnerable Persons**

1. [p. 12] Give the names of the top two persons with the most dangerous friends (and how many).

Anne Freeman - 8

Bonnie Gilbert - 7

2. Execute the query on p. 13. Modify it to include friends of friends of friends, i.e., change the maximum path length.

a. Give your query here:

MATCH (p:Person)-[:KNOWS\*1..3]-(friend)-[:PARTY\_TO]->(:Crime)

WHERE NOT (p:Person)-[:PARTY\_TO]->(:Crime)

RETURN p.name AS name, p.surname AS surname, p.nhs\_no AS id, count(distinct friend) AS dangerousFriends

ORDER BY dangerousFriends DESC

LIMIT 5

b. Execute your query. Who are the top 3 persons and what is the count associated with each?

Walter James – 18

Annie Duncan – 17

Kelly Peterson – 17

3. Execute the queries up to p. 15 and observe the results. You will be combining elements of these queries to answer the data demand in the next question. (This item does not require an answer to be recorded.)

4. Find persons who know someone who is party to a crime (or they have a friend who knows someone who is a party to a crime) that live their same area. There should be 4 instances: 3 distinct people and 4 friends.

a. Give the query.

MATCH (p:Person)-[k:KNOWS\*1..2]-(friend)-[pt:PARTY\_TO]->(c:Crime),

(p)-[ca1:CURRENT\_ADDRESS]->(aAddress)-[lia1:LOCATION\_IN\_AREA]->(area),

(friend)-[ca2:CURRENT\_ADDRESS]->(fAddress)-[lia2:LOCATION\_IN\_AREA]->(area)

RETURN p.name, friend.name

b. Give the 3 people and their friends/friends of a friend involved in a crime who live in the same area.

Antonio -> Norma

Richard -> Andrea

Anne -> Craig

Anne -> Craig

c. How do Richard and Andrea *KNOW* each other (who is the person who *KNOW*s both of them)? Give your query to find this information and the person's name.

MATCH (friend1)-[:KNOWS]-(p:Person {name: 'Andrea'}), (friend1)-[:KNOWS]-(friend2 {name: 'Richard'})

WHERE (friend1.name <> 'Andrea') AND (friend1.name <> 'Richard')

RETURN friend1.name

Harry

5. Who has the most dangerous family friends and how many? [p. 17]

Kimberly Alexander – 7

6. How many criminal friends does Bonnie have? [p. 19]

8

**Part 3: Triangles and Centrality**

1. Who participates in the most triangles? How many? [p. 20]

Deborah Ford – 10

2. Who has the most criminal activity triangles? How many? [p. 22]

Phillip Williamson - 4

3. Which person has the highest centrality score and what is it? [p. 24]

Annie Duncan

4. How many persons are in the *KNOWS* graph for Annie Duncan with up to 3 hops? [p. 25]

121 (Including Annie)