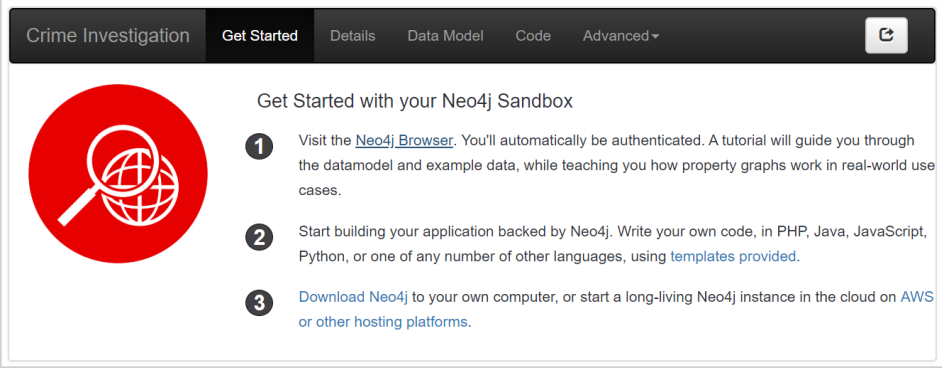
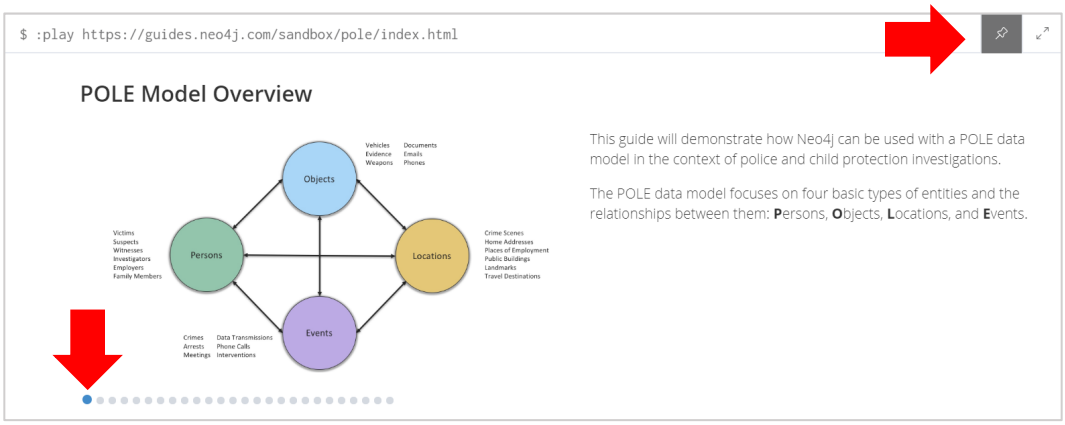
Neo4j Guided Study
Individual Assignment: 25 points

This is an individual assignment due on Wednesday May 8 at midnight. It is a series of questions that can be answered using the Neo4j Crime Investigation sandbox. (https://neo4j.com/sandbox-v2/)



The Crime Investigation sandbox has 27 pages (unfortunately not numbered, represented by dots at the bottom left of the screen.) At the upper right, you can select the thumbtack symbol to pin this screen to the top of your browser feed. When you start executing queries, the most recent query result will be at the top of the feed. If you don't pin this screen to the top, you will spend a lot of time scrolling down to find it after you execute a few queries. If you pin it, the most recent query result will appear under the guide window.



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: a person can be victims, suspects

O: vehicles, evidence

L: crime scene, land marks

E: crimes, arrests

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

a. Counter Terrorism

b. Missing Persons

c. Offender Rehabilitation

4. What aspects of the POLE dataset are real?

Only the L (location) and E (events) are provided realistic.

5. What aspects of the POLE dataset are fictitious?

All scenarios and persons portrayed in this demo are fictitious. Any similarity to real persons or events is coincidental and unintentional.

6. Execute the node profiling query from the lecture nodes (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. Excecute the query. How many different kinds of *Crime* nodes are there?

7 records.

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

["Crime"]	["date", "id", "type", "last_outcome"]	12079

7. Execute the relationship profiling query from the lecture nodes (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).

["Person"]	"REVIEWED"	["rating", "summary"]	["Movie"]	7

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

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

"Violence and sexual offences", "Public order", "Criminal damage and arson", "Burglary", "Vehicle crime"

10. Execute the query on p. 6.

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

```
MATCH (l:Location)<-[:OCCURRED_AT]-(:Crime{type:"Robbery"})
RETURN l.address AS address, l.postcode AS postcode, count(l) AS total
ORDER BY count(l) DESC
LIMIT 15
```

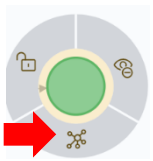
b. What are the top 3 locations for robberies?

"Piccadilly", "Shopping Area", "Nightclub"

11. Execute the query on p. 7. In the query answer, which crime occurs the most in the *crime_type* arrays? How many times?

["Bicycle theft", "Violence and sexual offences", "Public order"], occurred 3 times

12. Execute the query on p. 8. Click on the green *Crime* node. In the gray circle 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?



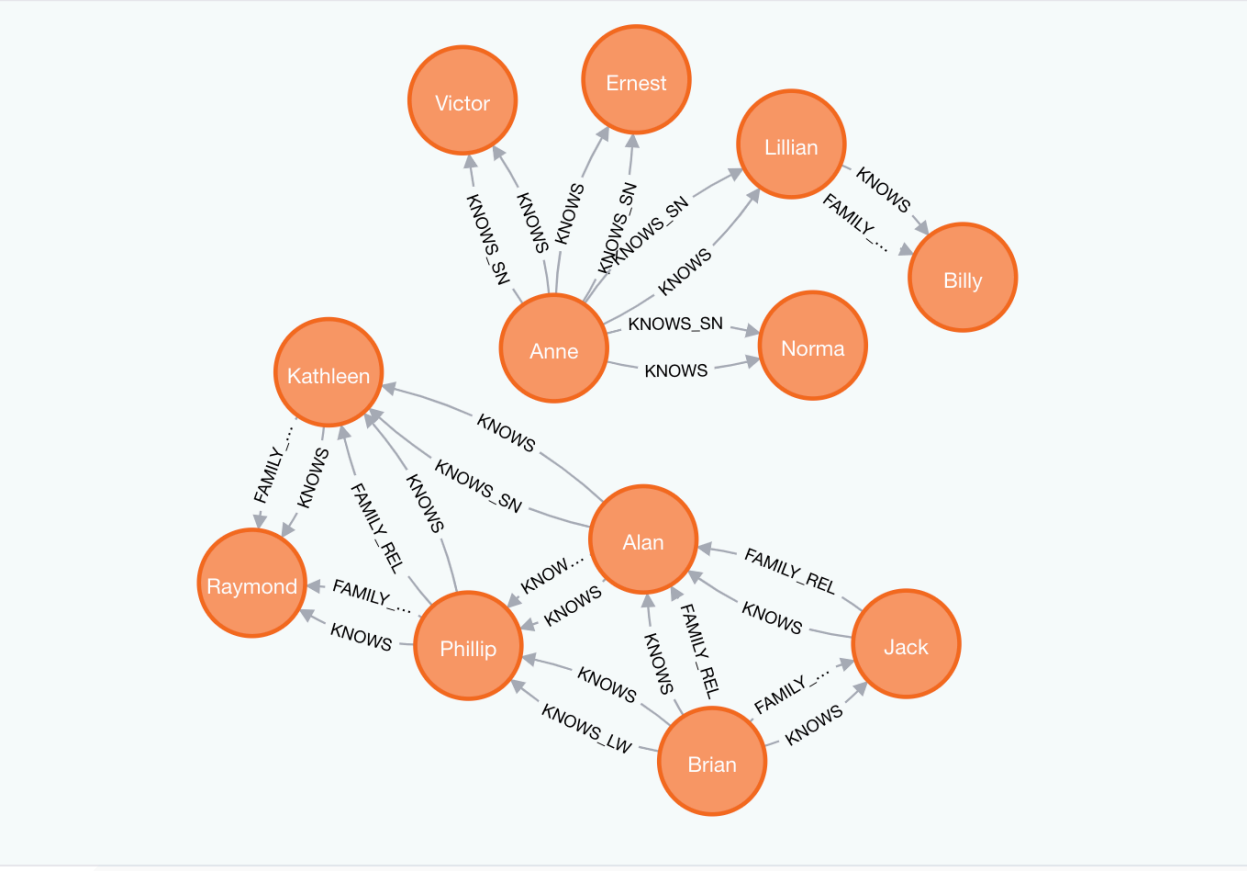
It connects to the location where it is occurred at.

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 it 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'})-[:INVESTIGATED_BY]->(:Officer),
(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 has 8 dangerous friends, and Bonnie has 7 dangerous friends.

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, Kelly, Annie are the top 3 people and the count means how many friends of friends they have are associated in the crime.

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.)

It shows the result that craig is involved in two crimes and a friend of him, anne is currently living in the same area with him.

Shows the local dangerous friends.

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)-[: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)
where not  (p: Person)-[pt:PARTY_TO]->(c:Crime)
RETURN *
```

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

Anne's friend Crig is involved in crime and is living in the same area as her
Antonio's friend Norma is involved in crime is living in the same area as him
Richard's friend Andrea is involved in cime and is living in the same are as him

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

```
MATCH (p: Person{name : "Richard"})-[:KNOWS*1..2]-(friend1)-[:KNOWS*1..2]-(friend2)-
[pt:PARTY_TO]->(c:Crime),
(p)-[ca1:CURRENT_ADDRESS]->(aAddress)-[lia1:LOCATION_IN_AREA]->(area),
(friend2)-[ca2:CURRENT_ADDRESS]->(fAddress)-[lia2:LOCATION_IN_AREA]->(area)
where not  (p: Person)-[pt:PARTY_TO]->(c:Crime)
RETURN *
```

Harry is the connection between Richard and Andrea.

5. Who has the most dangerous family friends and how many? [p. 17]
Kimberly Alexander has the most dangerous family friends, for 7.

6. How many criminal friends does Bonnie have? [p. 19]
She has 8 primal friends

Part 3: Triangles and Centrality

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

Deborah Ford participates the most, for 10 triangles.

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

Phillip Williamson has the most criminal activity triangles, for 4 triangles.

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

Annie Duncan has the highest centrality score, and its 5275.

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