

Kryminalne zagadki

Uniwersytet Gdański
Grafowe Bazy Danych

MICHAŁ REDKWA, MACIEJ MARZEC

12.01.2025

1 Wstęp

1.1 Założenia projektu

W ramach projektu zaliczeniowego stworzyliśmy grafową bazę danych, której celem jest pokazanie, w jaki sposób za pomocą algorytmów i zapytań można zidentyfikować osobę odpowiedzialną za popełnienie przestępstwa. Na bazie zostały przeprowadzone zapytania, które pozwalają na wskazanie osób, które mogą być bezpośrednio związane z dowodami przestępstwa lub pełnić rolę potencjalnych świadków. Dzięki zastosowaniu zaawansowanych algorytmów, jesteśmy w stanie ocenić ich znaczenie w kontekście śledztwa oraz wytypować osoby o najmocniejszych powiązaniach z podejrzanymi.

1.2 Opis wierzchołków oraz relacji pomiędzy nimi

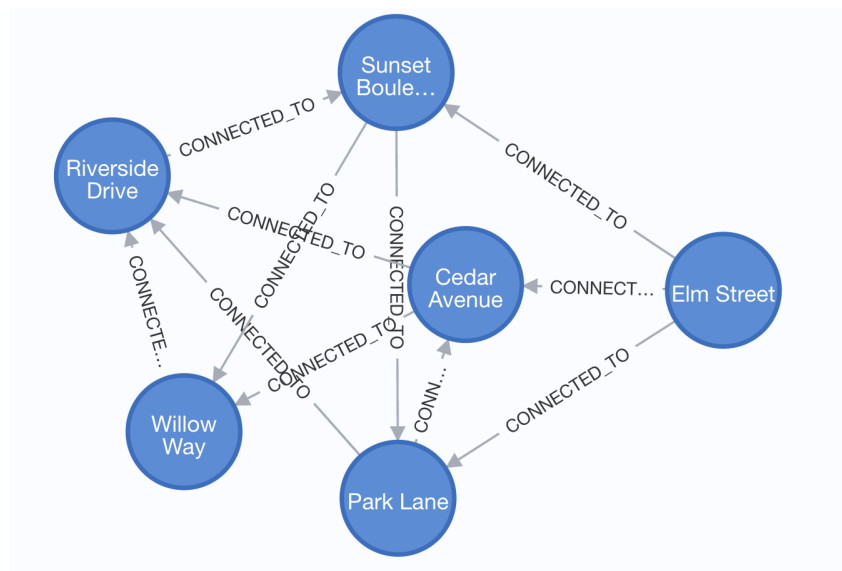
Person

Wierzchołek **Person** reprezentuje osobę, posiada on właściwości: imię, nazwisko, wiek, długość oraz kolor włosów, kolor oczu, numer telefonu, numer konta bankowego oraz zawód jaki wykonuje. Jest on powiązany z następującymi wierzchołkami:

- **Street** – relacja **LIVES_ON** pomiędzy Person a Street.
- **Shop** – relacja **BOUGHT_AT** lub **VISITED** pomiędzy Person a Shop. Relacja posiada dodatkowo właściwości, takie jak: typ wizyty, ilość zakupów, ilość wizyt.

Street

Wierzchołek **Street** reprezentuje ulicę, zawiera takie właściwości jak: nazwa, miasto, kod, typ. Ulice są powiązane ze sobą za pomocą relacji **CONNECTED_TO**.



Institution

Wierzchołek **Institution** reprezentuje instytucję, jego właściwościami jest nazwa oraz jaki jest to typ instytucji. Wierzchołek ten powiązany jest z ulicami za pomocą relacji **LOCATED_AT**.

Shop

Wierzchołek **Shop** reprezentuje sklep. Posiada następujące właściwości: nazwa, typ oraz godziny otwarcia. Jest powiązany z ulicą za pomocą relacji **LOCATED_AT**.

Crime

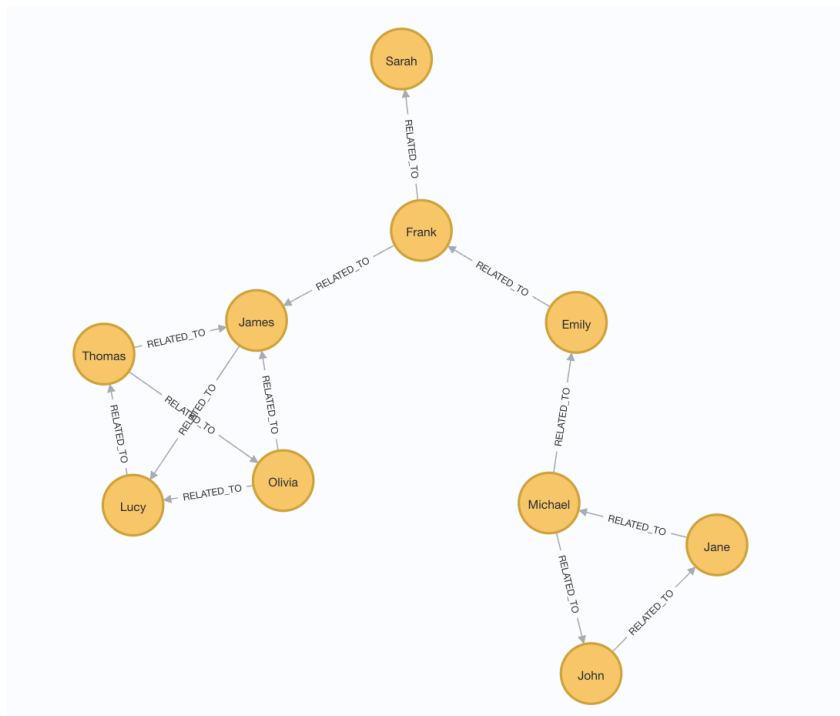
Wierzchołek **Crime** reprezentuje zbrodnie. Posiada właściwości: typ przestępstwa, data, wyrok. Jest powiązany z daną ulicą, sklepem czy instytucją za pomocą relacji o nazwie **COMMITTED_AT**.

Evidence

Wierzchołek **Evidence** reprezentuje dowód, który posiada właściwości takie jak: typ, opis oraz data złożenia lub jego znalezienia. Jest powiązany z wierzchołkiem zbrodni za pomocą relacji **EVIDENCE_IN**.

Criminal

Wierzchołek **Criminal** reprezentuje znanego przestępcę i zawiera następujące właściwości: imię, nazwisko, wiek, długość oraz kolor włosów, kolor oczu, pseudonim oraz kartotekę przestępczą (lista przestępstw). Węzeł ten jest wykorzystywany do reprezentowania osób zaangażowanych w działania przestępcze, umożliwiając analizę powiązań między nimi za pomocą relacji **RELATED_TO** która reprezentuje znane przypadki współpracy przy popełnianiu przestępstw, znajomości lub inne powiązania.

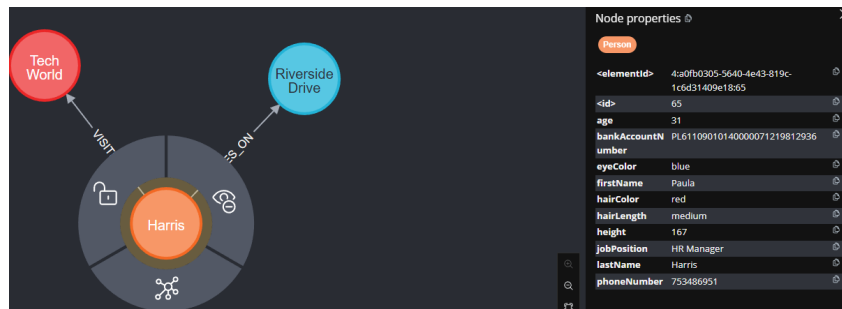


2 Zapytania Cypher

2.1 Proste rekomendacje - włamanie

W ramach demonstracji przeszukiwania grafowej bazy danych naszym podejrzanym w przypadku włamania do sklepu będzie Paula Harris. Na początek wypiszmy wszystkie informacje jakie mamy o tej osobie.

```
MATCH (p:Person {firstName: 'Paula', lastName: 'Harris'
    })-[r]->(related)
RETURN p, r, related
```



Teraz prześledźmy wszystkie dowody związane ze sprawą włamania i spróbujmy na podstawie ich znaleźć osobę, która właśnie jemu odpowiada.

```
MATCH (n)-[r]->(m)
WHERE n.crimeType = 'Robbery' AND n.date = '2023-05-12'
RETURN r, n, m;
```



Na podstawie zeznań świadków:

- A witness reported seeing a person with medium-length red hair leaving the scene of the robbery at around 12:05 AM.
- A witness reported that the suspect had a height between 160 and 170 cm.
- A witness saw a person with red hair near the scene of the robbery at approximately 11:50 PM.

oraz nagrania z kamery, które wg opisu mówi nam: CCTV footage shows a person with medium-length red hair and blue eyes near the robbery location at 11:55 PM.

Wyszukajmy osoby pasujące do tego rysopisu:

```
MATCH (person: Person)
WHERE person.hairColor = 'red' AND person.height >= 160
      AND person.height <= 170 AND person.eyeColor = '
      blue'
RETURN person.firstName, person.lastName, person.height
      , person.hairColor, person.eyeColor
LIMIT 5;
```

First Name	Last Name	Height (cm)	Hair Color	Eye Color
Daisy	Anderson	160	red	blue
Paula	Harris	167	red	blue
Wanda	Holmes	170	red	blue
Grace	Reed	164	red	blue
Maya	Xavier	163	red	blue

Tabela 1: Tabela przedstawiająca dane osób z kolorem włosów "red" i kolorem oczu "blue".

Jak widać, mamy na naszej liście osobę, która popełniła to przestępstwo, jednak nie daje nam to ostatecznego potwierdzenia, że jest to osoba, która powinna zostać uznana za winną popełnionego przestępstwa. Został jeszcze jeden dowód z centrali telefonicznej w obrębie obszaru, gdzie popełniono przestępstwo.

"A phone with the number which ends on ...951 was found to have been frequently used near the crime scene, with activity under the alias "AngryVoice" reported around the time of the robbery."

```
MATCH (person: Person)
WHERE toString(person.phoneNumber) ENDS WITH '951'
RETURN person.firstName, person.lastName, person.
      phoneNumber, person.hairColor, person.eyeColor
LIMIT 5;
```

First Name	Last Name	Phone Number	Hair Color	Eye Color
Paula	Harris	753486951	red	blue
Henry	Clark	852147951	grey	brown
Emma	Parker	852147951	blonde	green
Brian	Mills	456789951	brown	blue
David	Owens	753456951	black	brown

Tabela 2: Tabela przedstawiająca dane osobowe, numery telefonów oraz kolory włosów i oczu.

Jak widać, osoba którą podejrzewaliśmy znajduje się na tej liście - sprawa została rozwiązana.

2.2 Proste rekomendacje - dodatkowe

Zapytanie o osoby pasujące do opisu z monitoringu odnośnie zbrodni morderstwo:

```

MATCH (person : Person)
WHERE person.hairColor = 'black' AND person.height >=
    155 AND person.height <= 165
RETURN person.firstName , person.lastName , person.height
    , person.hairColor
LIMIT 5;

```

First Name	Last Name	Height (cm)	Hair Color
Olive	Hunt	164	black
Bella	Hughes	164	black
Isabel	Stevens	163	black
Amy	Lewis	160	black
Isla	Taylor	158	black

Tabela 3: Tabela przedstawiająca osoby z kolorem włosów "black" i wzrostem między 155 a 165 cm.

Zapytanie o osoby, których inicjały mogą pasować do "IT", na znalezionym przedmiocie zbrodni:

```

MATCH (person : Person)
WHERE substring(person.firstName , 0 , 1) + substring(
    person.lastName , 0 , 1) = 'IT'
RETURN person.firstName , person.lastName
LIMIT 5;

```

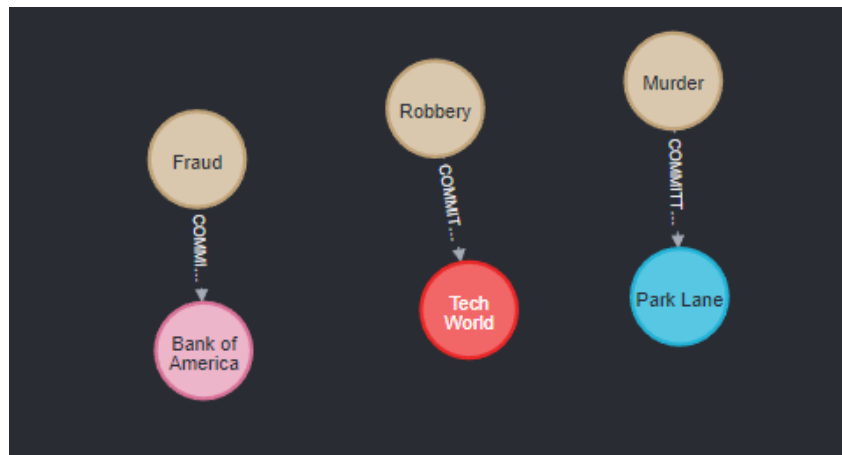
First Name	Last Name
Ivy	Taylor
Isla	Taylor
Isabel	Taylor
Isaac	Turner

Tabela 4: Tabela przedstawiająca osoby, których inicjały pasują do wyrytych na narzędziu zbrodni.

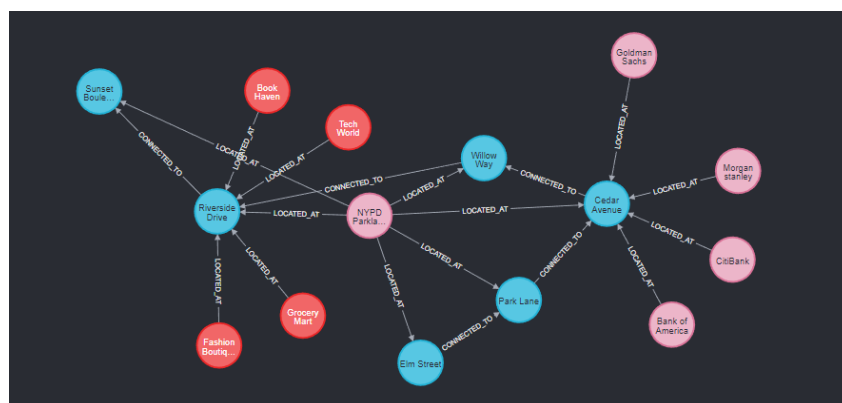
Kolejna sprawa rozwiązana - naszym mordercą w tym przypadku jest Isla Taylor.

2.3 Złożone zapytania

2.3.1 Grafowe przedstawienie powiązań do wykonywanych zapytań



Rys. 1: Graf przedstawiający połączenia dokonanych zbrodni z miejscem.



Rys. 2: Graf przedstawiający połączenia instytucji oraz sklepów z ulicami.

2.3.2 Przykładowanie zapytania oraz wyniki

1. Zapytanie, które ma na celu wskazanie potencjalnych świadków w odległości 2 od miejsca zbrodni włamanie.

```

MATCH ( crime : Crime ) -[:COMMITTED_AT]->( location ) -[:
    LOCATED_AT]->( street : Street ) ,
    path = ( street ) -[:CONNECTED_TO*1..2]->( neighbor :
        Street ) <-[:LIVES_ON]-( person : Person )
RETURN DISTINCT person . firstName , person . lastName ,
    neighbor . name AS potentialWitnessLocation , street .
    name AS crimeLocation ;

```

No.	First Name	Last Name	Potential Witness Location	Crime Location
1	Jack	Lee	Sunset Boulevard	Riverside Drive
2	Paul	Evans	Sunset Boulevard	Riverside Drive
3	Steve	Morris	Sunset Boulevard	Riverside Drive
4	Yara	Collins	Sunset Boulevard	Riverside Drive
.....				
33	Victor	King	Sunset Boulevard	Riverside Drive

Tabela 5: Tabela przedstawiająca dane świadków i lokalizacje zdarzeń.

2. Zapytanie, które zwróci w tabeli liczbę zbrodni w danej lokalizacji, jej ilość oraz typ.

```

MATCH ( crime : Crime ) -[:COMMITTED_AT]->( location )
OPTIONAL MATCH ( location ) -[:LOCATED_AT]->( street : Street
    )
OPTIONAL MATCH ( crime ) -[:COMMITTED_AT]->( institution :
    Institution )
RETURN
    COALESCE( street . name , location . name ) AS StreetName ,
    COUNT( DISTINCT crime ) AS CrimeCount ,
    COLLECT( DISTINCT CASE
        WHEN institution IS NOT NULL THEN institution .
            institutionName
        ELSE location . name
    END ) AS CrimeLocations ,
    COLLECT( DISTINCT crime . crimeType ) AS CrimeTypes ;

```

No.	StreetName	CrimeCount	CrimeLocations	CrimeTypes
1	Riverside Drive	1	Tech World	Robbery
2	Park Lane	1	Park Lane	Murder
3	Cedar Avenue	1	Bank of America	Fraud

Tabela 6: Tabela przedstawiająca dane dotyczące przestępstw, lokalizacji i typów przestępstw.

3. Podsumowanie wszystkich popełnionych zbrodni, zliczenie ilości dowodów i wymienienie ich typów.

```

MATCH (crime: Crime) -[:COMMITTED_AT]->(location)
OPTIONAL MATCH (evidence: Evidence) -[:EVIDENCE_IN]->(
    crime)
RETURN
    crime.crimeType AS CrimeType,
    COUNT(evidence) AS EvidenceCount,
    COLLECT(evidence.type) AS EvidenceTypes
ORDER BY EvidenceCount DESC;

```

No.	CrimeType	EvidenceCount	EvidenceTypes
1	Robbery	5	Witness Statement, Witness Statement, Phone Data, CCTV, Witness Statement
2	Fraud	4	Fake Invoice, Bank Record, Witness Testimony, Email Record
3	Murder	2	Security Footage, Murder Weapon

Tabela 7: Tabela przedstawiająca dane dotyczące przestępstw oraz zebranych dowodów powiązanych z nimi

4. Zebranie w tabelę ile ludzi żyje na danej ulicy, jakie są sklepy oraz instytucje.

```

MATCH (s: Street)
OPTIONAL MATCH (i: Institution) -[:LOCATED_AT]->(s)
OPTIONAL MATCH (sh: Shop) -[:LOCATED_AT]->(s)
OPTIONAL MATCH (p: Person) -[:LIVES_ON]->(s)
WITH s,
    collect(DISTINCT i.institutionName) AS
        Institutions,
    collect(DISTINCT sh.name) AS Shops,
    count(DISTINCT p) AS PeopleLiving
RETURN s.name AS Street, Institutions, Shops,
    PeopleLiving;

```

5. Zebranie w tabelę ile osób wg. dowodów pasuje do popełnionej zbrodni (osobno).

No.	Street	Institutions	Shops	PeopleLiving
1	Elm Street	NYPD Parklane, NYPD Elm-Street	-	41
2	Park Lane	NYPD Parklane	-	33
3	Cedar Avenue	Bank of America, CitiBank, Goldman Sachs, Morgan Stanley, NYPD Parklane, NYPD Cedar-Avenue	-	36
4	Willow Way	NYPD Parklane, NYPD WillowWay	-	36
5	Riverside Drive	NYPD Parklane, NYPD Riverside	Grocery Mart, Tech World, Fashion Boutique, Book Haven	38
6	Sunset Boulevard	NYPD Parklane, NYPD Sunset Boulevard, Hospital Sunset Boulevard	-	37

Tabela 8: Tabela przedstawiająca dane dotyczące ulic, instytucji, sklepów oraz liczby mieszkańców.

MATCH (p: Person)

WITH

COUNT(CASE **WHEN** p.height >= 160 AND p.height <= 170 **THEN** 1 **END**) AS HeightBetween160And170 ,
COUNT(CASE **WHEN** p.hairColor = 'red' **THEN** 1 **END**) AS RedHairColor ,
COUNT(CASE **WHEN** **TOSTRING**(p.phoneNumber) **ENDS WITH** '951' **THEN** 1 **END**) AS PhoneNumberEndsWith951 ,
COUNT(CASE **WHEN** p.eyeColor = 'blue' **THEN** 1 **END**) AS BlueEyeColor

RETURN

HeightBetween160And170 AS 'Height Between 160 and 170',
RedHairColor AS 'Red Hair Color',
PhoneNumberEndsWith951 AS 'Phone Number Ends With 951',
BlueEyeColor AS 'Blue Eye Color'

Height Between 160 and 170	Red Hair Color	Phone Number Ends With 951	Blue Eye Color
101	41	14	80

Tabela 9: Tabela przedstawiająca liczbę osób pasujących per dowód w sprawie włamania.

3 Algorytmy

Logarytm Betweenness Centrality dla ulic

To zapytanie wykorzystuje algorytm *Betweenness Centrality*, który identyfikuje najbardziej strategiczne węzły (wg. liczby najkrótszych ścieżek) w grafie ulic. Wynikiem jest lista ulic posortowanych według ich znaczenia w sieci (im wyższy wynik, tym większe znaczenie).

```
CALL gds.graph.project(
  'streetGraph',
  'Street',
  {
    CONNECTED_TO: {
      type: 'CONNECTED_TO',
      orientation: 'UNDIRECTED'
    }
  }
)
YIELD graphName, nodeCount, relationshipCount;

CALL gds.betweenness.stream('streetGraph')
YIELD nodeId, score
RETURN gds.util.asNode(nodeId).name AS Street, score
ORDER BY score DESC;
```

Street	Score
Cedar Avenue	1.1(6)
Willow Way	1.1(6)
Park Lane	0.58(3)
Riverside Drive	0.58(3)
Elm Street	0.25
Sunset Boulevard	0.25

Tabela 10: Betweenness Centrality dla ulic

Spółeczności wśród ulic (Community Detection)

Zapytanie stosuje algorytm *Louvain* do wykrywania społeczności w grafie ulic. Wynikiem jest podział ulic na społeczności, co pozwala zidentyfikować grupy wzajemnie silnie powiązanych węzłów.

```
CALL gds.louvain.stream(
  'streetGraph', // Nazwa grafu
  {
    relationshipTypes: ['CONNECTED_TO']
  }
)
YIELD nodeId, communityId
```

```
RETURN gds.util.asNode(nodeId).name AS Street, communityId
ORDER BY communityId;
```

Street	communityId
Elm Street	1
Park Lane	1
Cedar Avenue	2
Willow Way	2
Riverside Drive	2
Sunset Boulevard	2

Tabela 11: Wykryte społeczności ulic

Znalezienie najkrótszych ścieżek pomiędzy miejscem zbrodni a potencjalnymi sprawcami

Zapytanie wykorzystuje algorytm Dijkstry do znalezienia najkrótszej ścieżki między miejscem przestępstwa a miejscem zamieszkania potencjalnego sprawcy. Wynikiem jest ścieżka między węzłami oraz jej koszt.

```
MATCH (suspect:Person)
WHERE (suspect.hairColor = 'red' AND suspect.height >= 160 AND suspect.height <= 170)
      OR (suspect.phoneNumber CONTAINS '951')
WITH suspect
MATCH (crime:Crime {crimeType: 'Robbery'})-[:COMMITTED_AT]->(shop:Shop)-[:LOCATED_AT]
      (suspect)-[:LIVES_ON]->(suspectStreet:Street)
WITH suspect, crimeLocation, suspectStreet
CALL apoc.algo.dijkstra(suspectStreet, crimeLocation, 'CONNECTED_TO', 'distance_properties')
YIELD path, weight
RETURN suspect.firstName + ' ' + suspect.lastName AS Suspect,
       suspectStreet.name AS 'Lives on',
       weight AS TotalDistance
ORDER BY TotalDistance ASC;
```

No.	Suspect	Lives on	TotalDistance
1	Wendy Turner	Riverside Drive	0.0
2	Daisy Anderson	Riverside Drive	0.0
3	Paula Harris	Riverside Drive	0.0
4	Ruby Henderson	Riverside Drive	0.0
.....			
22	Ursula Jones	Elm Street	9.0

Tabela 12: Tabela podejrzanych i najkrótszych ścieżek do miejsca zbrodni

Closeness Centrality dla siatki przestępców

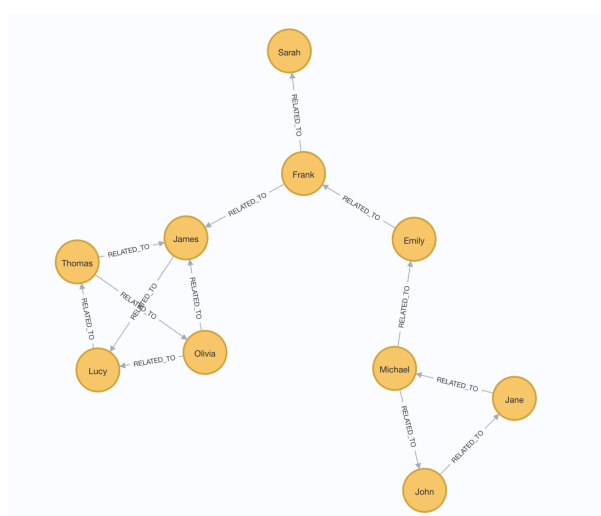
Algorytm Closeness Centrality służy do oceny, jak "blisko" dany wierzchołek jest od innych wierzchołków. Im mniejsza suma odległości do innych wierzchołków, tym wyższa centralność danego wierzchołka. Przy jego użyciu identyfikujemy, którzy przestępcy są najważniejszymi węzłami w sieci powiązań między innymi przestępcami.

```
CALL gds.graph.project( 'relationsGraph', 'Criminal',
{
  RELATED_TO: { type: 'RELATED_TO', orientation: 'UNDIRECTED' }
}
)
YIELD graphName, nodeCount, relationshipCount;

CALL gds.beta.closeness.stream('relationsGraph')
YIELD nodeId, score
RETURN gds.util.asNode(nodeId).firstName + ' ' + gds.util.asNode(nodeId).lastName AS criminal
ORDER BY score DESC;
```

No.	Criminal	Score
1	Frank Carter	0.53
2	Emily White	0.47
3	James King	0.47
4	Michael Stone	0.39
.....		
10	Jane Black	0.3

Tabela 13: Tabela przestępców i ich wyników Closeness Centrality



Rys. 3: Siatka przestępców