

Pokročilé databázové technológie

Zadanie č. 6 – Neo4j

Do kedy?

Keďže všetci robíte radi s predstihom, deadline je 31.12.2021 o 23:59.

Čo odovzdať?

Dotazy v separátnom textovom editore + dotaz a screenshot výsledku v pdf.

Hodnotenie

Úlohy spolu za 10 bodov rozdelené nasledovne:

- 1,2,3,4,5 a 6 sú za jeden bod,
- 7, 8 za 2 body.
- Bonus za 2 extra body.

Zadanie – Neo4j

1. Zoberte si nami vytvorený dataset z vašich tweetov:

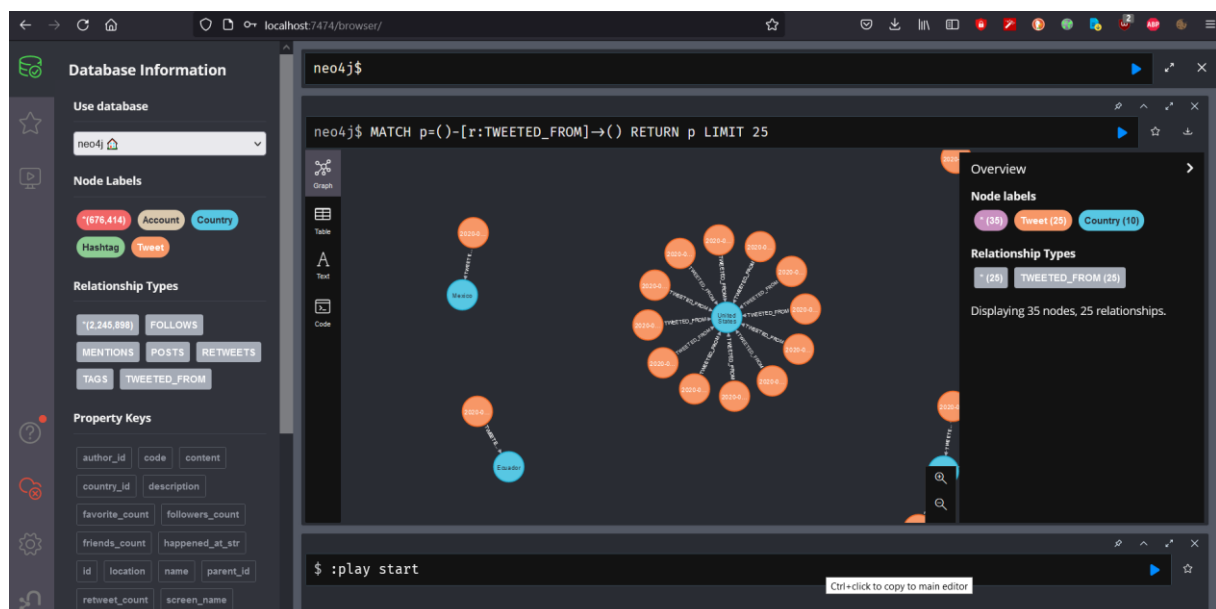
https://drive.google.com/file/d/1kufTAFP5DH8hiCT-BRLHqAicoulJ_0Qi/view?usp=sharing

a importujte ho cez neo4j-admin

Riešenie:

```
neo4j-admin load --from=./import/tweets2021-440.dump --database=neo4j --force
```

```
D:\PDT\zadanie6\neo4j-community-4.4.2\bin>neo4j-admin load --from=./import/tweets2021-440.dump --database=neo4j --force
Selecting JVM - Version:11+28, Name:OpenJDK 64-Bit Server VM, Vendor:Oracle Corporation
Done: 68 files, 279.3MiB processed.
```



2. Vypíšte 5 Accountov s najvyšším množstvom followerov. Nezoradujte Accounty podľa poľa followers_count, hodnota je prevzatá z Twitteru a nezodpovedá našim vzťahom v datasete. Zaujímajú nás followujúce Accounty v našom datasete cez vzťah FOLLOWS.

Riešenie:

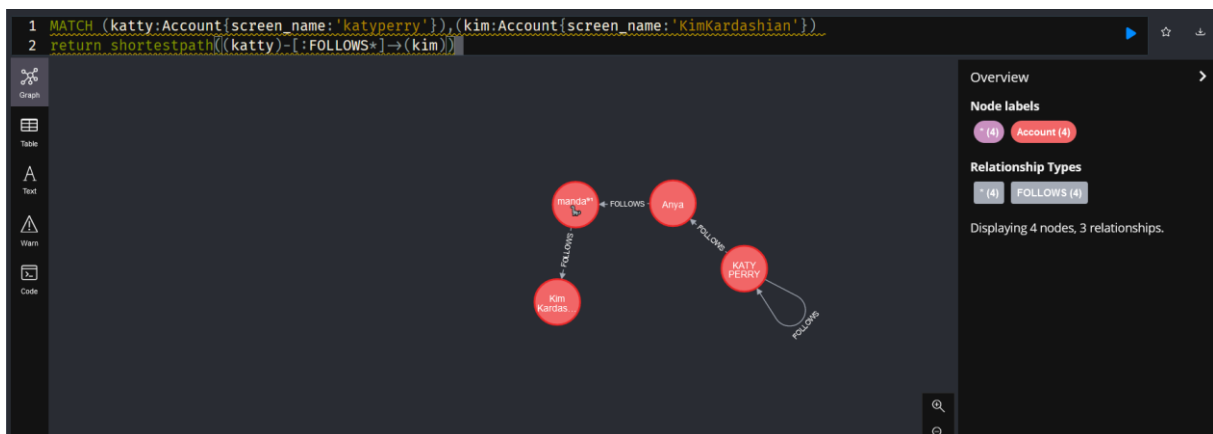
```
1 MATCH (ac1)-[rel:FOLLOWS]→(ac2)
2 return ac2.name as name, count(rel) as follower_count order by count(rel) desc limit 5
```

	name	follower_count
1	"Barack Obama"	12725
2	"KATY PERRY"	11462
3	"Donald J. Trump"	6735
4	"Kim Kardashian West"	6680
5	"Narendra Modi"	6425

Started streaming 5 records after 6 ms and completed after 2213 ms.

3. Nájdite najkratšie cesty medzi Katy Perry {screen_name: 'katyperry'} a Kim Kardashian {screen_name: 'KimKardashian'} cez vzťah FOLLOWS. Všetky cesty, kde Kim Kardashian followuje niekoho, kto followuje niekoho, kto..., kto followuje Katy Perry.

Riešenie:



4. Vyhľadajte neúspešné tweety influencerov. Vyhľadajte 10 najmenej retweetovaných tweetov od Accountov, ktoré sú na prvých 10 miestach v celkovom počte retweetov.

Riešenie:

The screenshot shows a Neo4j interface with a Cypher query and its results in table view. The query is as follows:

```
1 match (a:Account)-[posts:POSTS]-(tweets:Tweet)-[retweets:RETWEETS]-(other1:Tweet)
2 with a as accounts, count(retweets) as retweet_count order by retweet_count desc limit 10
3 match (a)-[:POSTS]-(bad_tweets:Tweet)-[bad_retweets:RETWEETS]-(other2:Tweet)
4 return a as accounts,bad_tweets as tweet, count(bad_retweets) as bad_retweet_count order by bad_retweet_count asc limit 10
```

The results are displayed in two columns: 'accounts' and 'tweet'. The 'accounts' column shows details for an account with identity 28543, screen name 'ResisterDude', and a name 'Pissed Off 😡 Until Further Notice'. The 'tweet' column shows details for a tweet with identity 313660, author '1170304086497005568', and content 'Today's tweet storm by @realDonaldTrump means one thing: He's trying to distract from GDP down 32%'. The interface also shows a sidebar with 'Graph', 'Table', 'Text', and 'Code' views, and a status bar at the bottom indicating 'Started streaming 10 records after 1 ms and completed after 11122 ms.'

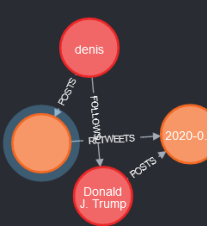
Aj s vizualizáciou vzťahov:

The screenshot shows a Neo4j interface with the same Cypher query as above, but the results are displayed in graph view. The graph shows a network of nodes and relationships. The nodes are labeled with their identities and names, and the relationships are labeled with their types. The graph view shows a complex network of nodes and relationships, with a sidebar on the right showing 'Overview', 'Node labels', and 'Relationship Types'. The sidebar also shows the number of nodes and relationships for each label and type. The status bar at the bottom indicates 'Displaying 20 nodes, 10 relationships'.

5. Vytvorte volaním iba jednej query nový Account s Vaším menom, ktorý bude followovať Donalda Trumpa {screen_name:"realDonaldTrump"} a v tom istom volaní vytvorte tweet, ktorý bude retweetom Donaldovho najretweetovanejšieho tweetu.

Riešenie:

```
1 match (trump:Account{screen_name:'realDonaldTrump'})-[:POSTS]-(tweets:Tweet)
  <-[retweets:RETWEETS]-(:Tweet)
2 with trump, tweets as najretweet_trumpa, count(retweets) as retweet_count order by
  retweet_count desc limit 1
3 create(trump)<-[:FOLLOWS]-(denis:Account{screen_name:'DeDe', name:'denis'})-[:POSTS]->
  (moj_tweet:Tweet{content:'epstein did not kill himself!'})-[:RETWEETS]->(najretweet_trumpa)
4 return trump, najretweet_trumpa, denis, moj_tweet
```



Node Properties

Tweet

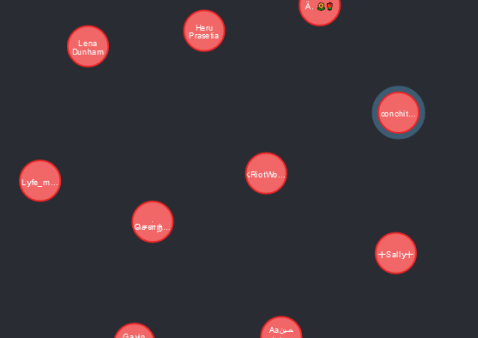
<id> 676697

content epstein did not kill himself!

6. Odporučte používateľovi {screen_name:"777stl"} followovanie ďalších Accountov, na základe followovania rovnakých Accountov: Vyhľadajte 10 Accountov, ktoré followujú najviac rovnakých Accountov ako náš používateľ, ale náš používateľ ich ešte nefollowuje.

Riešenie:

```
1 match (acc:Account{screen_name:'777stl'})-[:FOLLOWS]-(followed_acc_by_acc)
  <-[same_follows:FOLLOWS]-(other_acc:Account)
2 where not (other_acc:Account)-[:FOLLOWS]-(followed_acc_by_acc)
3 return other_acc, count(same_follows) as follows order by follows desc limit 10
```



Node Properties

Account

<id> 72355

description #Cubana 100%,
Pinareña, Fidelista,
Chavista, Revolucionaria,
amante de la música y de
las buenas acciones. Por
#Cuba lo mas grande.🇨🇺

followers_c 3147
ount

friends_cou 3596
nt

id 1089951252681760773

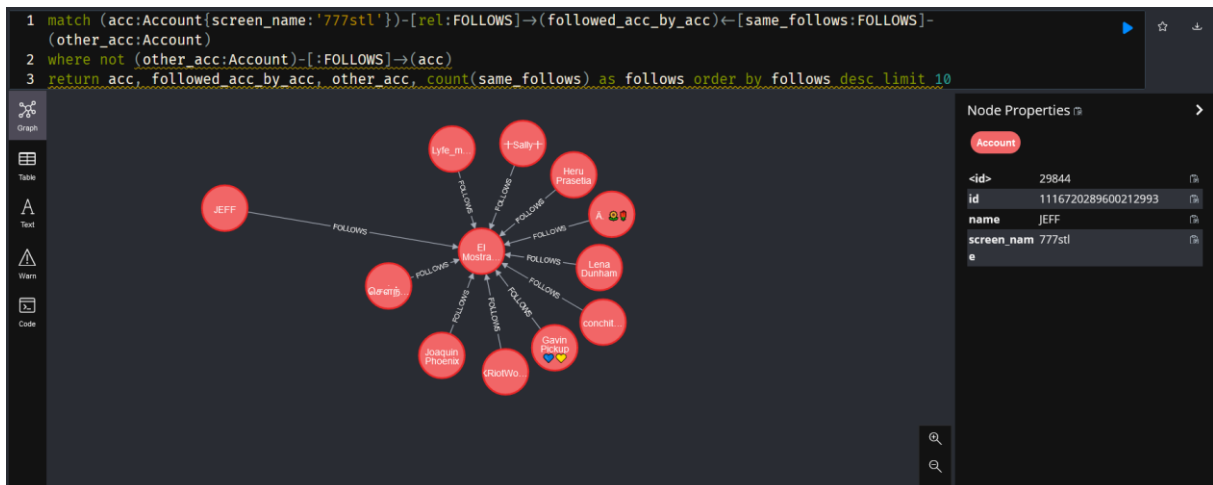
name conchita pre ❤️

screen_na conchitapinar

me

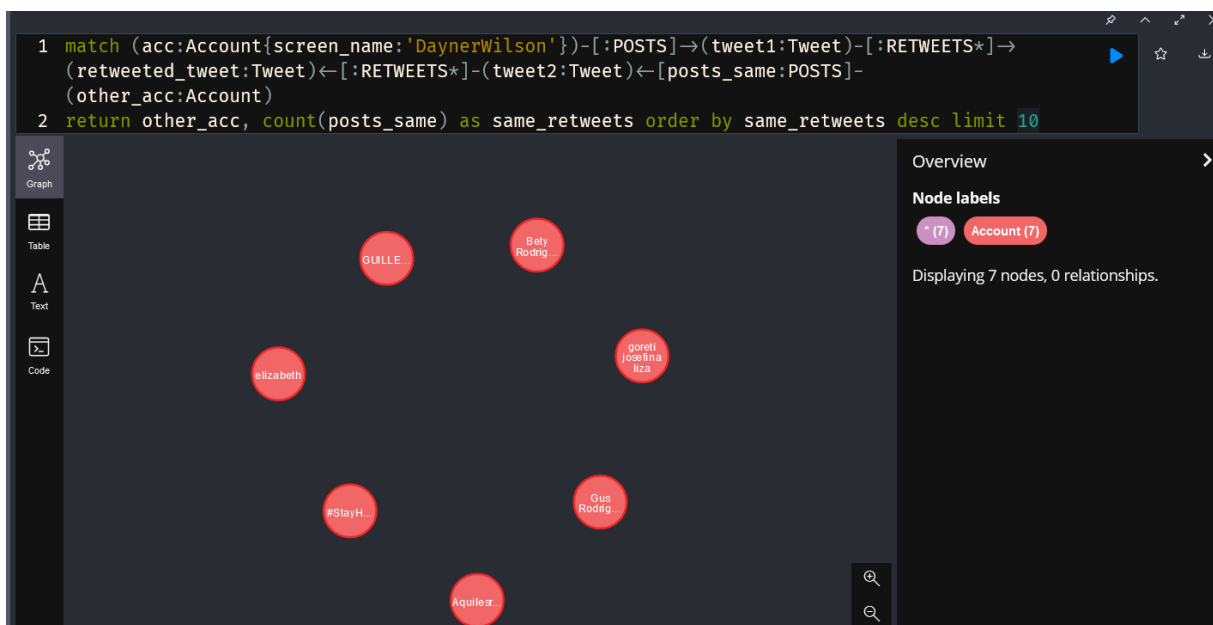
statuses_c 15583
ount

Aj s vizualizáciou vzťahov:

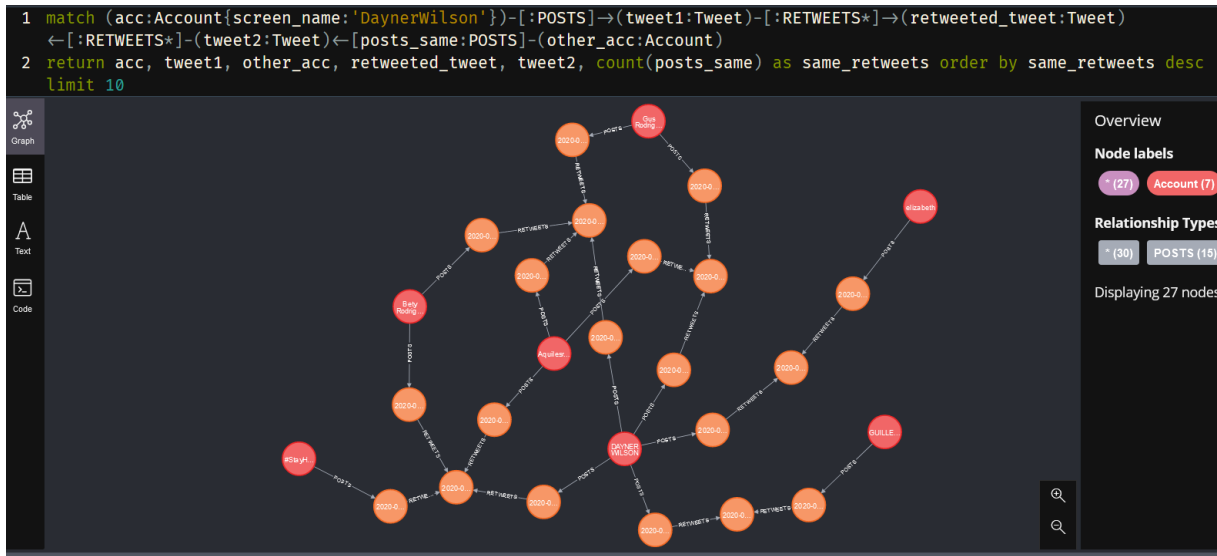


7. Odporučte používateľovi {screen_name:"DaynerWilson"} followovanie ďalších Accountov na základe zhody v retweetovaní rovnakých tweetov: Vyhľadajte 10 accountov, ktoré retweetli najviac tých istých tweetov, ako náš používateľ. Ak tweet ktorý retweetujeme, je už tiež retweetom, rátajte za zhodu aj retweetovanie jeho parent tweetu – retweetovanie teda zohľadňujte rekurzívne.

Riešenie:

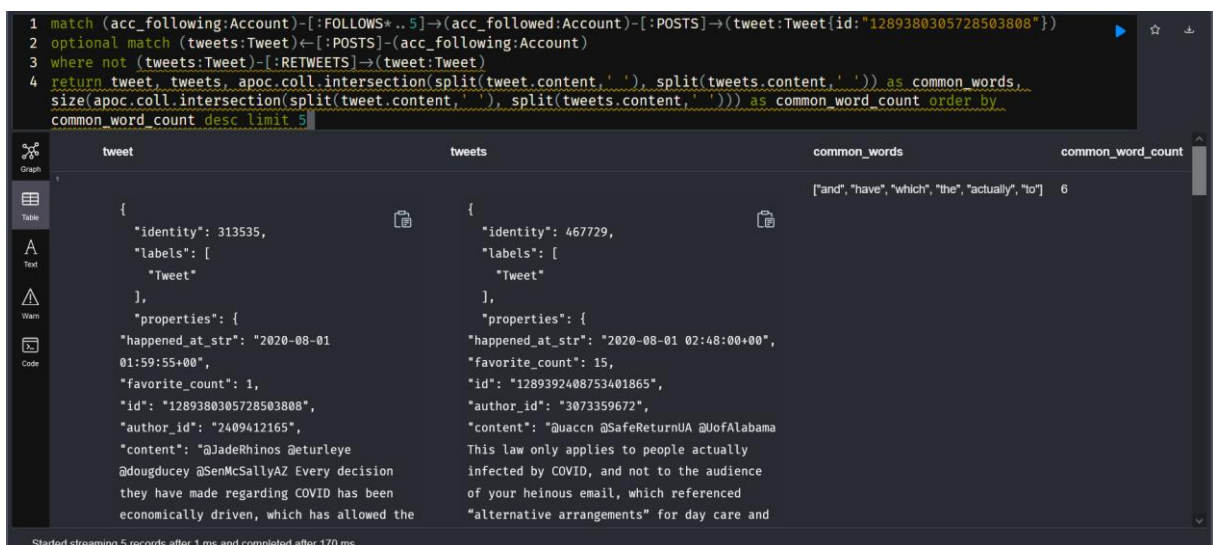


Aj s vizualizáciou vzťahov:

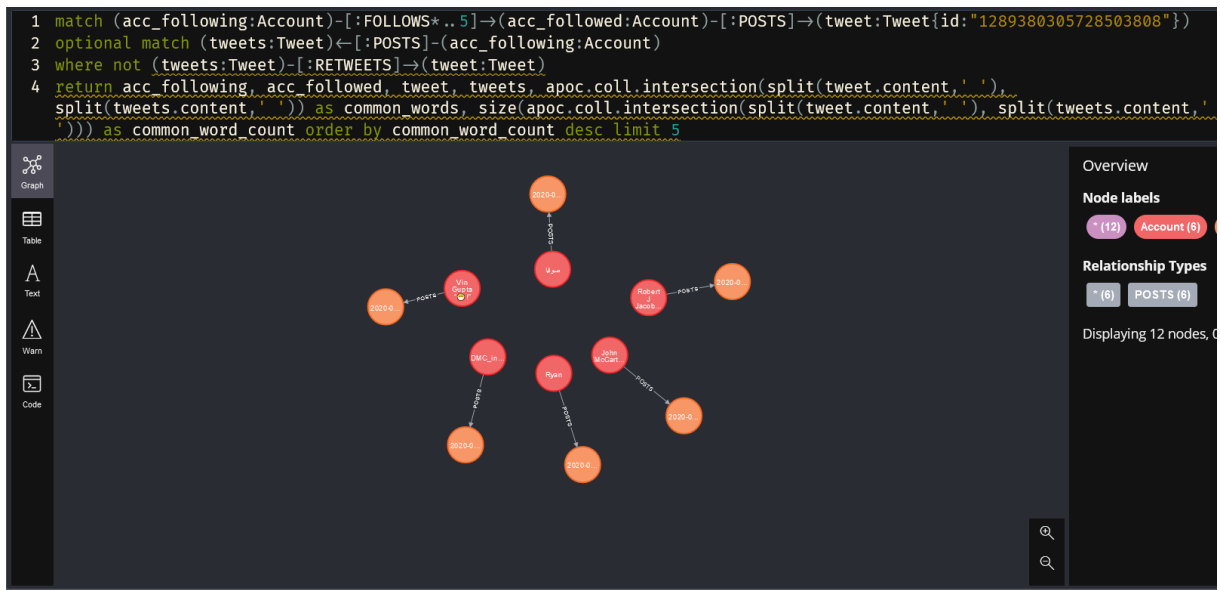


8. Vyhľadajte 5 tweetov ostatných Accountov, ktoré do hĺbky 5 followujú account, ktorý napísal tweet {id: "1289380305728503808"}, ktoré síce nie sú retweetom vybraného tweetu, ale napriek tomu majú čo najviac rovnakých slov v poli content zhodných s vybraným tweetom (stačí rozdeliť content na slová cez `split(tweet.content, " ")`). Account, ktorý followuje Account, ktorý followuje nami vybraný Account rozumieme hĺbkou 2. Odporúčam pozrieť si procedúry v knižnici APOC pracujúce s collections, ale nie je to podmienkou na zvládnutie úlohy.

Riešenie:



Aj s vizualizáciou vzťahov:



BONUS: Nájdite najkratšie cesty medzi Katy Perry (katyperry) a Donaldom Trumpom cez vzťah RETWEETS (a tým pádom aj POST). Všetky cesty, kde Katy Perry retweetla post Accountu, ktorý retweetol post Accountu, ktorý..., ktorý retweetol post Donalda Trumpa.

(Account) -> [POSTS] -> (Tweet) – [RETWEETS] -> (Tweet) <- [POSTS] – (Account) – [POSTS] -> (Tweet) – [RETWEETS] -> (Tweet) <- [POSTS] - (Account) -> ...

Riešenie:

