

Fakulta informatiky a informačných technológií STU v Bratislave

**Zadanie 4 – Neo4j**

**29.9.2022**

**Matej Delinčák**

**Github classroom:** <https://github.com/FIIT-DBS/zadanie-pdt-mateju25>

1. Vytvorte volaním iba jednej query nového Autora s Vaším menom (vyplňte minimálne name a username) a tým istým volaním vytvorte tweet ktorý bude obsahovať aspoň jeden vami vybraný Hashtag (vzťah :HAS) a ktorý bude retweetovať najretweetovanejší tweet Vladimíra Zelenského{screen\_name:"ZelenskyyUa"}.

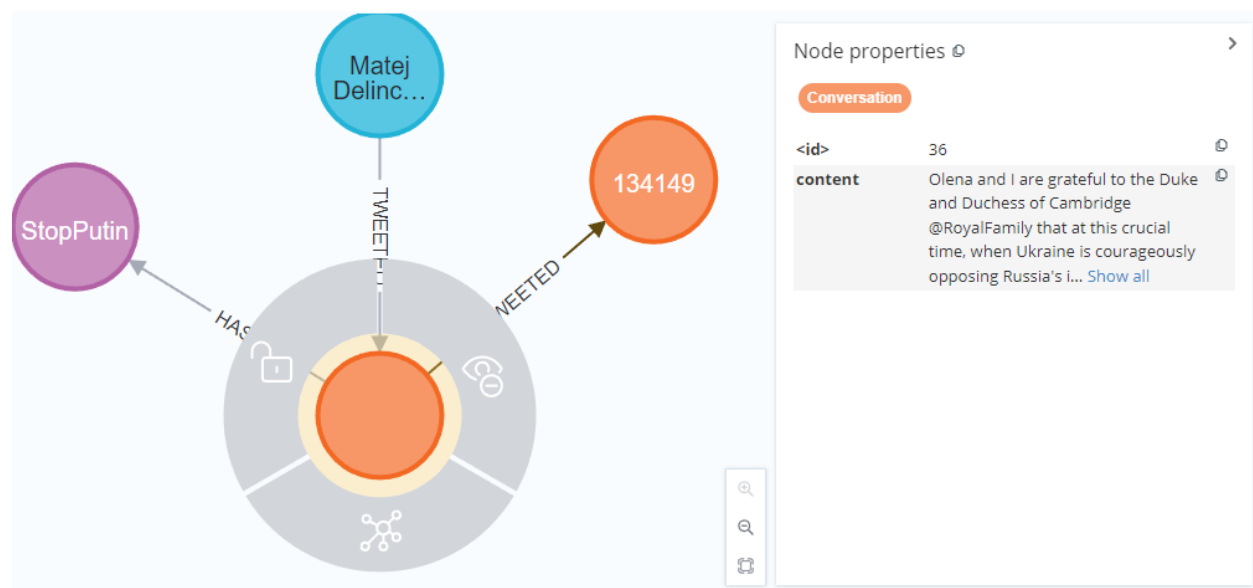
Ešte pred prvou otázkou som si importol dáta cez neo4j-admin príkaz, ktorý som našiel v bin. Verzia DBMS musela byť 4.4.5 aby to prebehlo úspešne.

Najskôr som si vytiahol tweety, ktoré patria zelenskému a cez agregáciu funkciu som si zrátal, koľko replied\_to vzťahov existuje na daný tweet. Z nich som vytiahol ten s najväčším počtom. Ešte som si našiel hocijaký hashtag. Následne som si vytvoril autora a tweet, ktorý ma HAS reláciu s hashtagom, moj autor ho TWEETol a ešte k tomu RETWEETuje zelenského tweet.

```
1 MATCH (a:Author)-[r:TWEETED]→(t:Conversation)←[rt:REPLIED_TO]-()
2 WHERE a.username='ZelenskyyUa'
3 WITH t, count(rt) as cnt ORDER BY cnt DESC LIMIT 1
4 MATCH ()-[r:HAS]→(h:Hashtag) WITH t, h LIMIT 1
5 CREATE (acc:Author {username:'matej.delincak', name: 'Matej Delincak'})
6 CREATE (new_tweet:Conversation {content: t.content})
7 CREATE (new_tweet)-[:HAS]→(h)
8 CREATE (new_tweet)-[:RETWEETED]→(t)
9 CREATE (acc)-[:TWEETED]→(new_tweet)
```

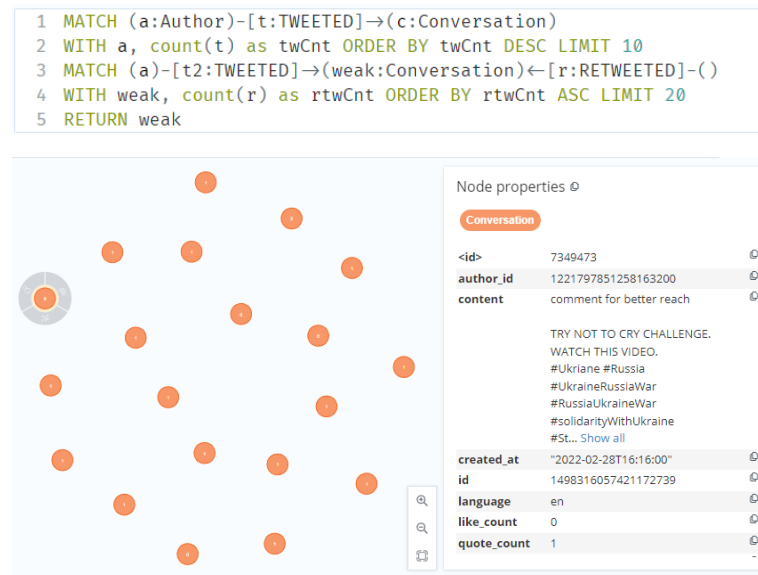
Added 2 labels, created 2 nodes, set 3 properties, created 3 relationships, completed after 1796 ms.

Na overenie som si selectol mňa a rozbalil som si graf pomocou gui. Dva tweety od zelenského mali rovnaký počet, druhý začínal „To the world...“.



**2. Vyhľadajte zlyhania influencerov. Vyhľadajte 20 najmenej retweetovaných tweetov od Accountov, ktoré sú na prvých 10 miestach v celkovom počte retweetov. Aj keď taký príklad v datasete nie je, počítajte aj s prípadom, ak by niektorý tweet bol retweetnutý 0 krát.**

Prvý krok bolo nájsť všetkých autorov, ktorý majú najviac tweetov. Podobne ako v prvej úlohe som si cez agregačnú funkciu selectol 10 najaktívnejších autorov. Následne som selectol také tweety, ktoré títo autori vytvorili a niekto iní ich retweetol aspoň raz. Z nich som vybral 20 najhorších.



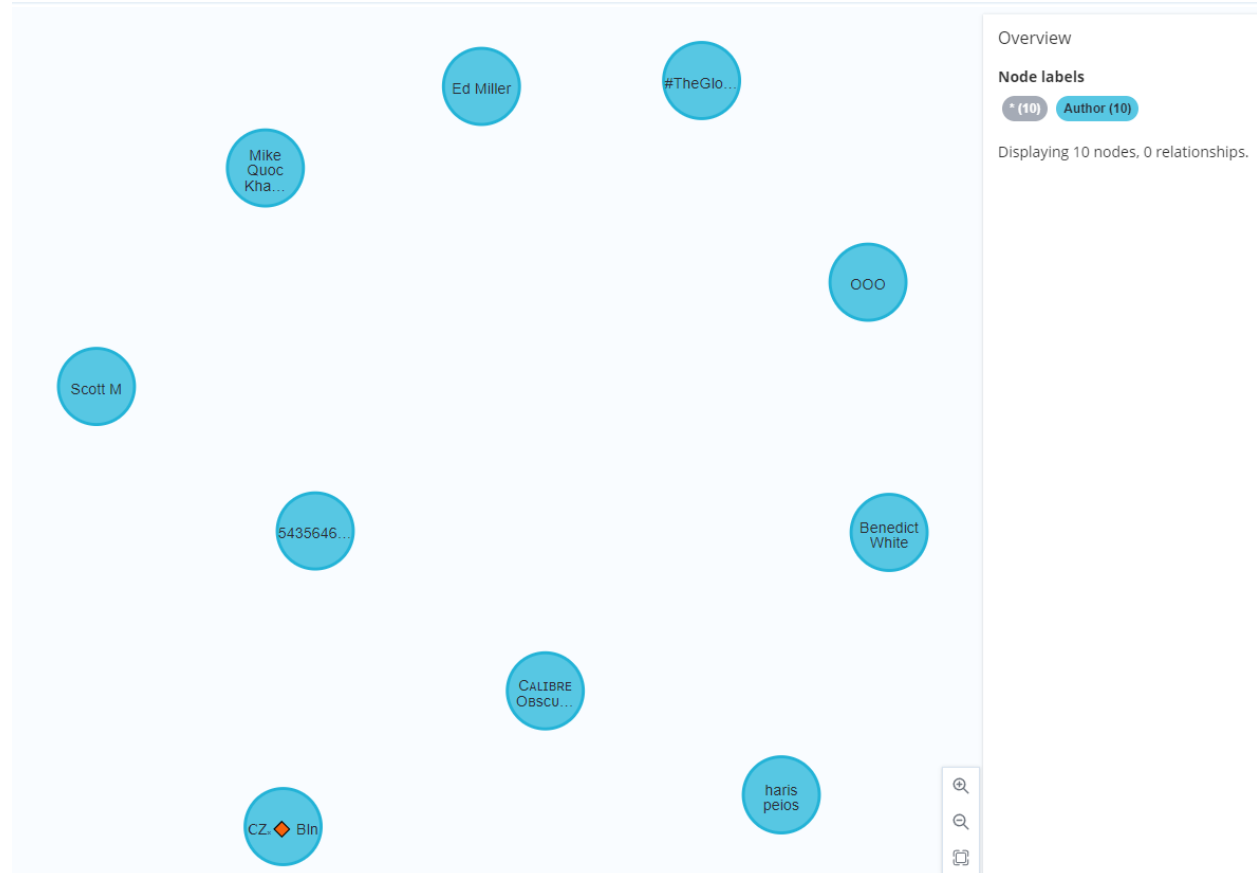
Prvé tweety – id a ich počet retweetov.

"weak.id"	"rtwCnt"
1498129025927360512	1
1498184901572849664	1
1498542678971396097	1
1498361187712782336	1
1498393754733977602	1
1498359267707207681	1
1498359024781508610	1
1498364068532752385	1
1498357524457996289	1
1498354497265233921	1

**3. Odporučte používateľovi (username: Marios59885699) followovanie ďalších autorov na základe zhody v retweetovaní rovnakých tweetov: Vyhľadajte 10 autorov, ktorí retweetli najviac tweetov rovnakých, ako používateľ Marios59885699. Počítajú sa aj retweety tweetov, ktoré retweetujú rovnaký tweet.**

Ako prvú vec som si našiel daného autora. Pomocou riadku tri som vytiahol všetky tweety také, aby niekto retweetol daný tweet až do hĺbky 2 (retweet tweetu) a to isté aj z autorovej strany. A opäť cez agregačnú funkciu som si zrátal, koľko daní autori vytvorili takýchto tweetov, ktoré sú prepojené a vrátil som 10 najlepších.

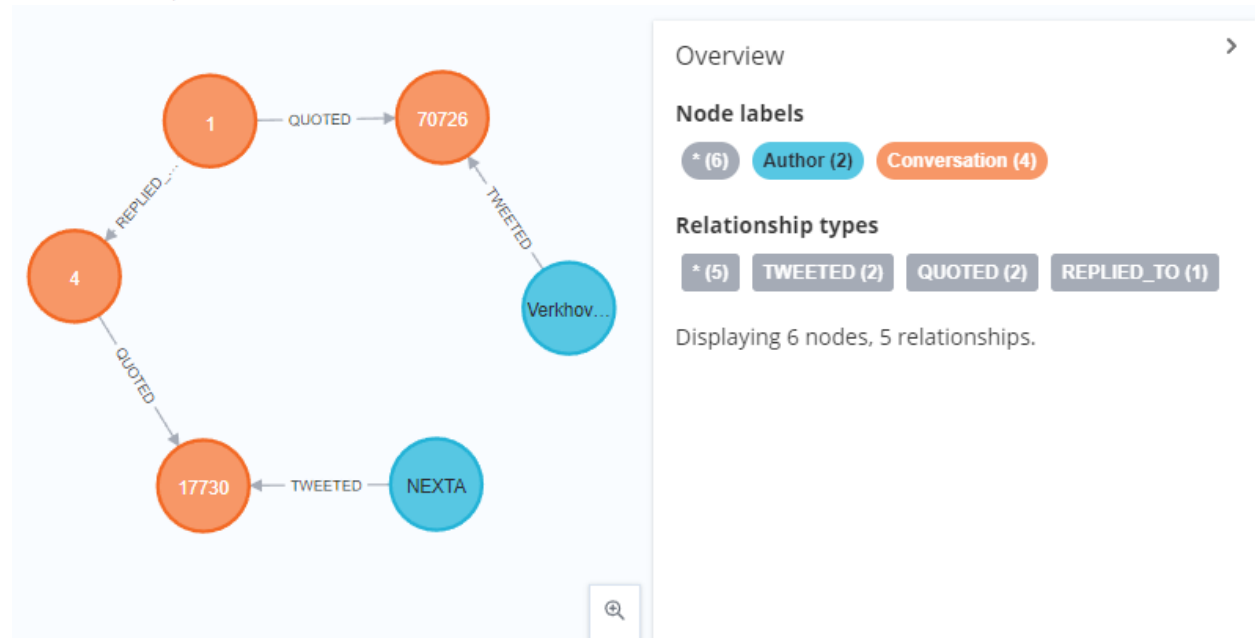
```
1 MATCH (auth:Author {username: 'Marios59885699'})
2 WITH auth
3 MATCH (auth)-[:TWEETED]->(:Conversation)-[:RETWEETED*1..2]->(c:Conversation)-[:RETWEETED*1..2]->(:Conversation)->[t:TWEETED]-(recomm:Author)
4 WITH recomm, count(t) as twCnt ORDER BY twCnt DESC
5 LIMIT 10
6 RETURN recomm
```



4. Nájdite najkratšie cesty medzi Ukrajinským parlamentom (username: "ua\_parliament") a NextaTV (username: "nexta\_tv") cez vzťahy TWEETED, RETWEETED, REPLIED\_TO a QUOTED. Hľadajte do hĺbky maximálne 10. (hint: allShortestPaths)

Najprv som si selectol daných autorov. A cez príkaz v hinte som našiel všetky najkratšie cesty medzi nimi dvoma do hĺbky 10. Nanešťastie sa našla len jedna cesta o hĺbke 5 a žiadna iná.

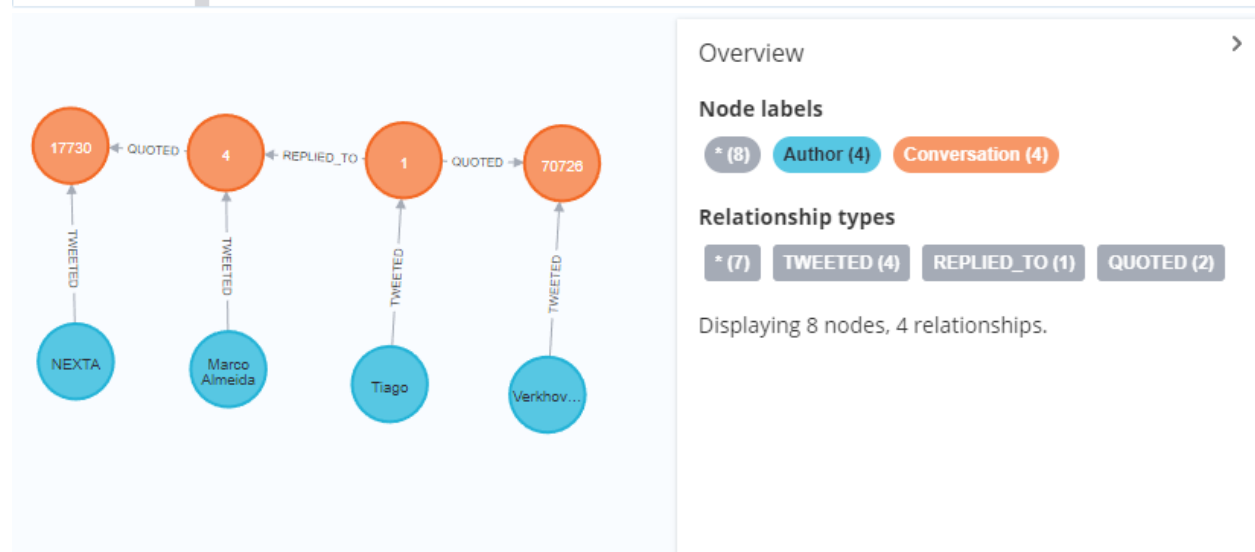
```
1 MATCH (n:Author {username: 'ua_parliament'})
2 MATCH (m:Author {username: 'nexta_tv'})
3 MATCH path = allShortestPaths( (n)-[:TWEETED|RETWEETED|REPLIED_TO|QUOTED*..10]-(m) )
4 RETURN path
```



5. Upravte query z predchádzajúcej úlohy, aby vrátila iba nájdenú najkratšiu cestu a jednotlivé uzly typu Conversation v nej aj spolu z autorom, ktorý ich tweetol. (hint: UNWIND)

V tejto query som už použil príkaz na nájdenie len jednej najkratšej cesty. Následne som si cez UNWIND rozbalil path na nodes a selectol som takých autorov, ktorý tweetli dané tweety.

```
1 MATCH (n:Author {username: 'ua_parliament'})
2 MATCH (m:Author {username: 'nexta_tv'})
3 MATCH path = shortestPath( (n)-[:TWEETED|RETWEETED|REPLIED_TO|QUOTED*..10]-(m) )
4 WITH nodes(path) AS x
5 UNWIND x AS c
6 MATCH (a:Author)-[t:TWEETED]→(c)
7 RETURN *
```



**6. Vypíšte 10 najpoužívanejších hashtagov v datasete aj s celkovým počtom použitií a autorom, ktorý daný hashtag najviac krát použil. (hint: collect)**

Tu som nepoužil hint ale subqeury. Selectol si najprv všetky potrebné hashtagy, ktoré majú najviac HAS. A potom pre každý jeden som našiel autora, ktorý mal najviac tweetov obsahujúce daný hashtag. Nižšie je tabuľka hashtagov a ich autorov, čo ich najviac používali.

```
1 MATCH ()-[r:HAS]→(h:Hashtag)
2 WITH h, COUNT(r) as hCnt ORDER by hCnt DESC LIMIT 10
3 CALL {
4   WITH h
5   MATCH (a:Author)-[t:TWEETED]→(c:Conversation)-[f:HAS]→(h)
6   WITH a, COUNT(t) as tCnt ORDER by tCnt DESC LIMIT 1
7   RETURN a
8 }
9 RETURN h, hCnt, a
```

"h"	"hCnt"	"a.username"
{"tag":"Ukraine"}	942018	"shdgsjksad"
{"tag":"Russia"}	235078	"indiaainstats1"
{"tag":"UkraineRussiaWar"}	197036	"indiaainstats1"
{"tag":"StandWithUkraine"}	159507	"indiaainstats1"
{"tag":"Putin"}	139129	"igeusllc"
{"tag":"StopRussia"}	70032	"shdgsjksad"
{"tag":"StopPutin"}	61875	"shdgsjksad"
{"tag":"Kyiv"}	59901	"ahmedbabaad21"
{"tag":"Kharkiv"}	49519	"PresseMondial"
{"tag":"NATO"}	44055	"TV9Bharatvarsh"