



به نام خدا
دانشگاه تهران
دانشکده مهندسی برق و کامپیوتر



درس آزمایشگاه پایگاه داده دستورکار سوم

نام و نام خانوادگی	میثاق محقق
شماره دانشجویی	810199484
تاریخ ارسال گزارش	1402.08.13

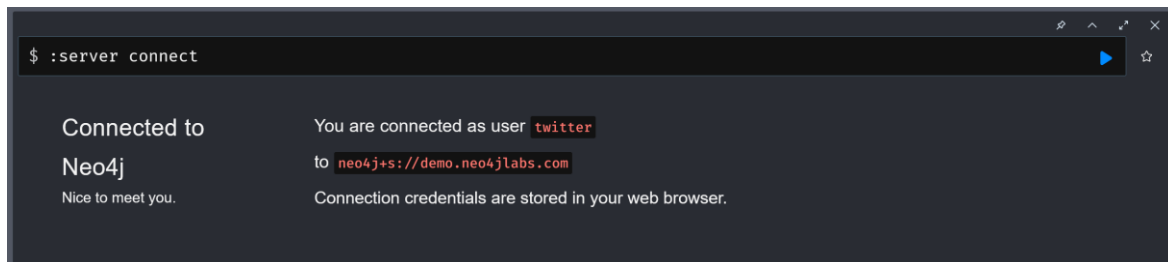
فهرست

2	پایگاه داده نمونه Twitter
2	Connect 1-0
2	Play 1-1
3	Graph your Application 1-2
3	Your Mentions 1-3
4	As Table 2-3
4	Most Influential Followers 1-4
5	Most Tagged 1-5
5	Followback Rate 1-6
6	Follower Recommendations 1-7
7	Links from Interesting Retweets 1-8
7	Common Hashtags 1-9

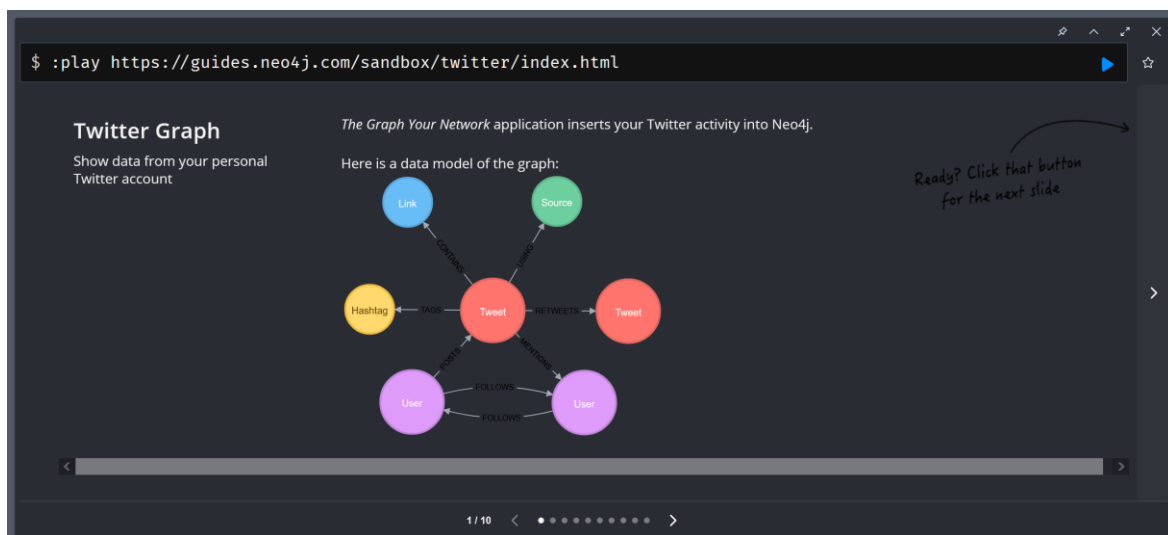
پایگاه داده نمونه Twitter

Connect 1-0

به لینک داده شده متصل می‌شویم و رمز twitter را وارد می‌کنیم:

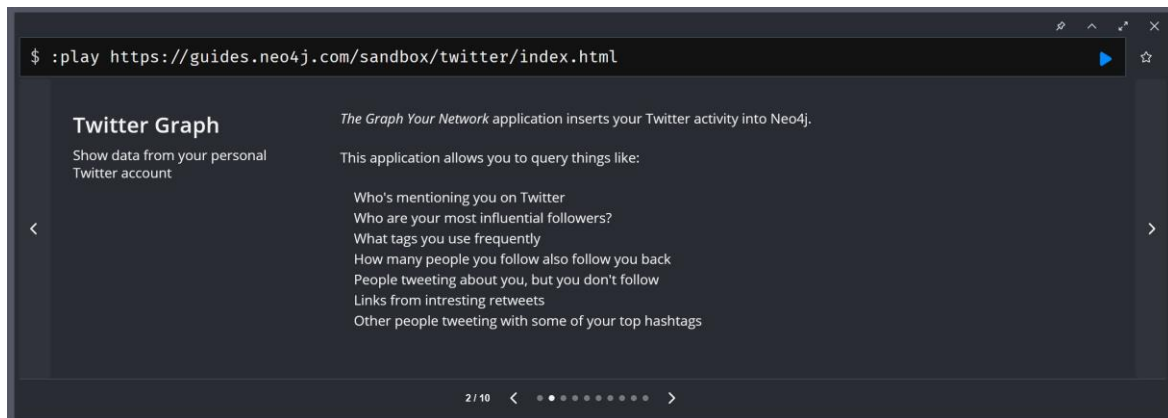


Play 1-1



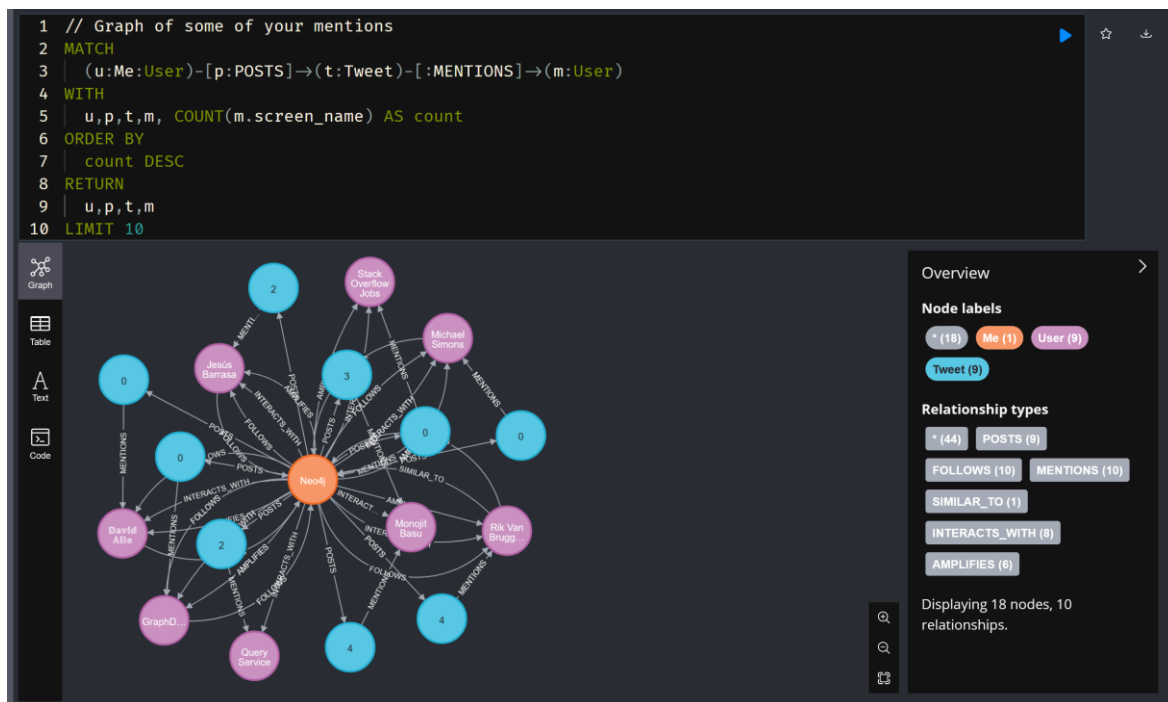
در اینجا می‌توانیم مدل گرافی دیتاها را مشاهده کنیم. نودهای این دیتابیس از چند label تشکیل شده اند که شامل توییت، لینک، هشتگ، کاربر و منبع ارسال است. طبق گراف، یک توییت می‌تواند ری‌توییت باشد، حاوی لینک بوده و هشتگ‌هایی داشته باشد. همچنان می‌توان در یک توییت یک کاربر را mention کرد و کاربرها می‌توانند همدیگر را follow کنند. پست‌های کاربران شامل توییت‌های آنها است.

Graph your Application 1-2



در این صفحه مثال‌هایی از کوئری‌هایی که می‌توان انجام داد را نوشته است.

Your Mentions 1-3



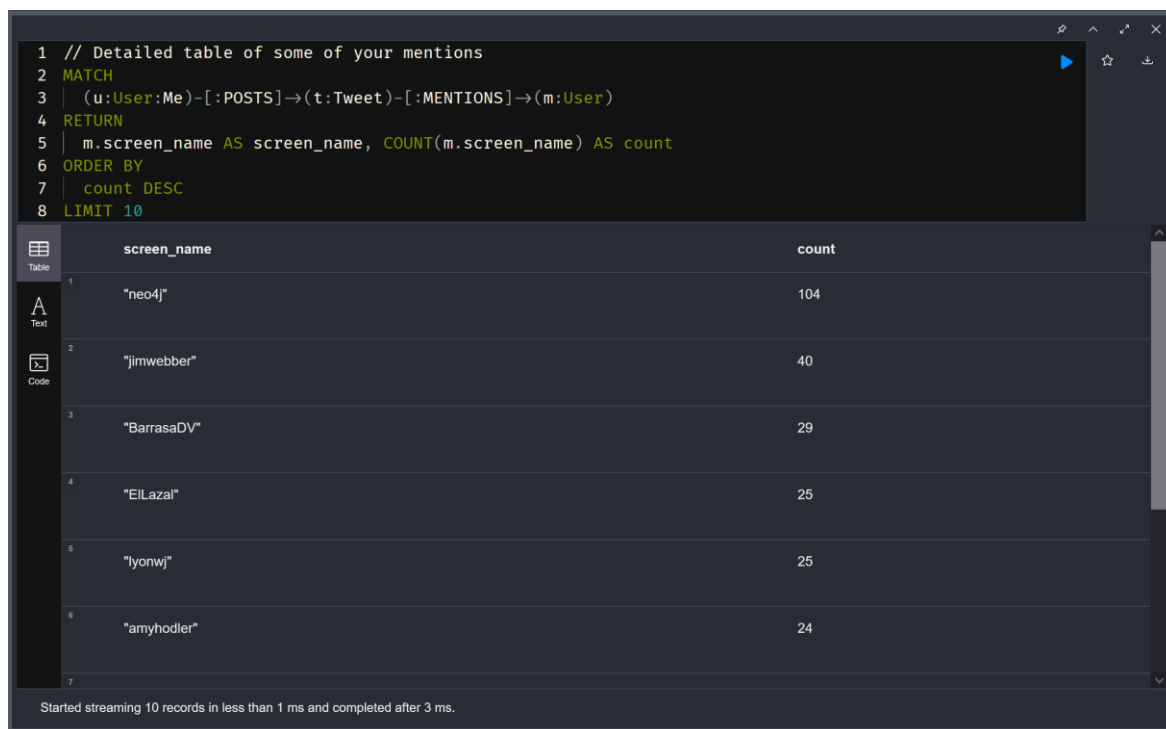
این Cypher ابتدا u که نودی است که هم لیبل Me دارد و هم لیبل User (در این دیتابیس فقط یکی هست) را گرفته و همه توییت‌هایی که Me زده و یک کاربر دیگری را داخل آن mention می‌کند را دریافت می‌کند. سپس با group by کرده و توییت‌ها را بر حسب تعداد منشن‌های داخلش سورت می‌کند.

در خروجی، از هر label حداکثر 10 تا آمده که همانطور که می‌بینیم، مثلاً از Me به یک توییت رابطه posts هست که آن پست یک کاربر را mention می‌کند.

(اعداد روی توییت‌ها تعداد favorite-ها اند)

As Table 2-3

در نمایش جدولی، می‌بینیم که مثلاً کاربر jimwebber را 40 بار در میان توییت‌هایمان منشن کرده‌ایم.



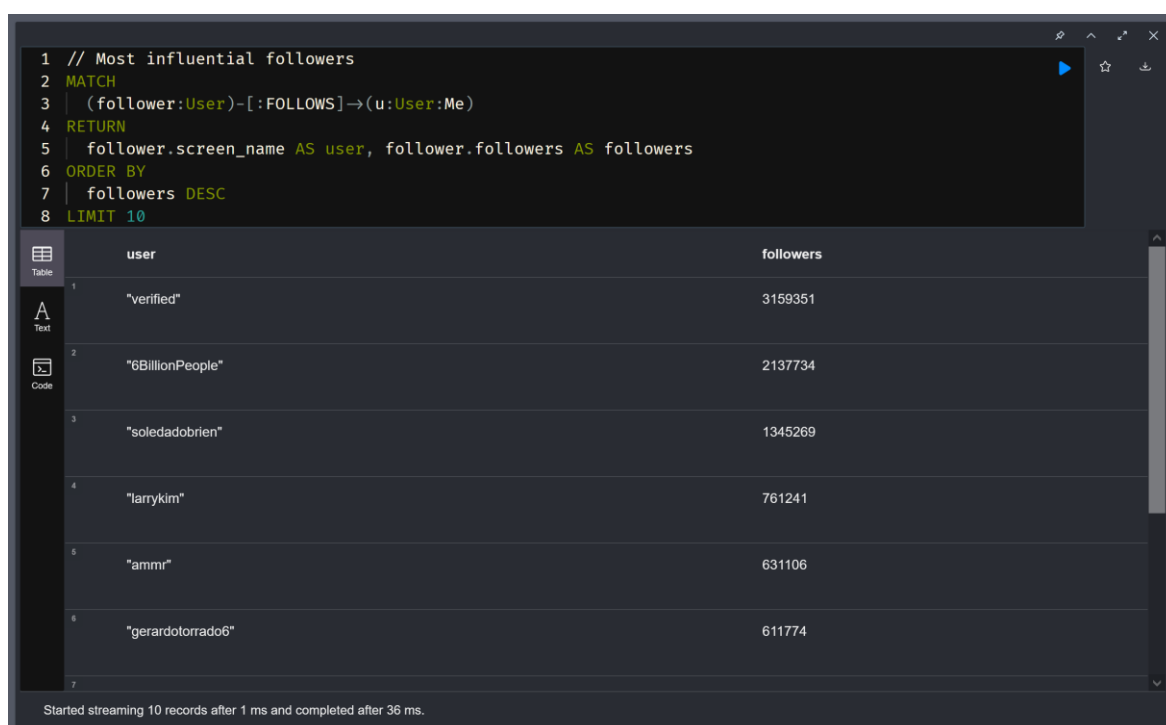
```
1 // Detailed table of some of your mentions
2 MATCH
3   (u:User:Me)-[:POSTS]->(t:Tweet)-[:MENTIONS]->(m:User)
4 RETURN
5   m.screen_name AS screen_name, COUNT(m.screen_name) AS count
6 ORDER BY
7   count DESC
8 LIMIT 10
```

	screen_name	count
1	"neo4j"	104
2	"jimwebber"	40
3	"BarrasaDV"	29
4	"EILazal"	25
5	"lyonwj"	25
6	"amyhodler"	24
7		

Started streaming 10 records in less than 1 ms and completed after 3 ms.

Most Influential Followers 1-4

در این بخش دنبال کسانی می‌گردیم که ما را follow کرده‌اند و تعداد follower-های آنها زیاد است. یعنی بین افرادی که ما را follow کرده‌اند، 10 نفری را انتخاب می‌کنیم که بیشترین تعداد follower را دارند.



```
1 // Most influential followers
2 MATCH
3   (follower:User)-[:FOLLOWS]->(u:User:Me)
4 RETURN
5   follower.screen_name AS user, follower.followers AS followers
6 ORDER BY
7   followers DESC
8 LIMIT 10
```

	user	followers
1	"verified"	3159351
2	"6BillionPeople"	2137734
3	"soledadobrien"	1345269
4	"larrykim"	761241
5	"ammr"	631106
6	"gerardotorrado6"	611774
7		

Started streaming 10 records after 1 ms and completed after 36 ms.

Most Tagged 1-5

در اینجا MATCH به صورت برعکس نوشته شده است و در نهایت تگ‌هایی که بیشتر از همه در توییت‌ها استفاده کرده‌ایم را (به همراه تعداد استفاده) برمی‌گرداند.

```
1 // The hashtags you have used most often
2 MATCH
3   (h:Hashtag)←[:TAGS]-(t:Tweet)←[:POSTS]-(u:User:Me)
4 WITH
5   h, COUNT(h) AS Hashtags
6 ORDER BY
7   Hashtags DESC
8 LIMIT 10
9 RETURN
10  h.name, Hashtags
```

	h.name	Hashtags
1	"neo4j"	241
2	"graphdatabases"	75
3	"twin4j"	68
4	"graphcast"	34
5	"graphtechnology"	29

Followback Rate 1-6

کد این بخش با سینتکس ارور برخورد می‌کند:

```
1 // Followback rate
2 MATCH
3   (me:User:Me)-[:FOLLOWS]→(f)
4 WITH
5   me, f, size((f)-[:FOLLOWS]→(me)) as doesFollowBack
6 RETURN
7   SUM(doesFollowBack) / toFloat(COUNT(f)) AS followBackRate
```

ERROR Neo.ClientError.Statement.SyntaxError

A pattern expression should only be used in order to test the existence of a pattern. It can no longer be used inside the function size(), an alternative is to replace size() with COUNT {}. (line 5, column 15 (offset: 77))

```
" me, f, size((f)-[:FOLLOWS]→(me)) as doesFollowBack"
```

با توجه به نکته ارور، با جایگزینی size با COUNT به دستور صحیح می‌رسیم:

```

1 // Followback rate
2 MATCH
3   (me:User:Me)-[:FOLLOWS]→(f)
4 WITH
5   me, f, COUNT{(f)-[:FOLLOWS]→(me)} as doesFollowBack
6 RETURN
7   SUM(doesFollowBack) / toFloat(COUNT(f)) AS followBackRate

```

Table	followBackRate
1	0.606775977874358

Started streaming 1 records in less than 1 ms and completed after 17 ms.

همانطور که می‌بینیم، ابتدا همه کسانی که ما follow می‌کنیم را می‌گیریم. سپس از بین آنها، آنهایی که ما را follow می‌کنند را پیدا کرده و نسبت آن به کل تعداد following‌های خود را حساب می‌کنیم.

7-1. Follower Recommendations

با این کوئری، لیست کاربرانی که ما را در تویییتی mention کرده‌اند را دریافت می‌کنیم که ما او را follow نمی‌کنیم.

استفاده از distinct برای این است که اگر کاربری چند بار mention شده باشد، یک رای بیشتر حساب نشود.

```

1 // Follower Recommendations - tweeting about you, but you don't follow
2 MATCH
3   (ou:User)-[:POSTS]→(t:Tweet)-[mt:MENTIONS]→(me:User:Me)
4 WITH
5   DISTINCT ou, me
6 WHERE
7   (ou)-[:FOLLOWS]→(me)
8   AND NOT
9   (me)-[:FOLLOWS]→(ou)
10 RETURN
11   ou.screen_name

```

Table	ou.screen_name
1	"drlynnchiu"
2	"jscarp"
3	"jughh"
4	"danfeindanfein"
5	"madjid_net"

Links from Interesting Retweets 1-8

```
1 // Links from interesting retweets
2 MATCH
3   (:User:Me)-[:POSTS]→
4   (t:Tweet)-[:RETWEETS]→(rt)-[:CONTAINS]→(link:Link)
5 RETURN
6   t.id_str AS tweet, link.url AS url, rt.favorites AS favorites
7 ORDER BY
8   favorites DESC
9 LIMIT 10
```

	tweet	url	favorites
5	"1362720758909329408"	"https://twitter.com/i/web/status/1362086924077436931"	36
6	"1356549869918380033"	"https://twitter.com/i/web/status/1356548651867385856"	31
7	"1354812704310009860"	"https://twitter.com/i/web/status/1354439155930030085"	22
8	"1350467524685684738"	"https://twitter.com/i/web/status/1350176362196316160"	21

با این کوئری، پست‌هایی از خود را دریافت می‌کنیم که یک retweet هستند و در آنها لینک وجود دارد. سپس بر حسب سورت تعداد favorite-های آن، 10 تای اول (بیشترین fav) را می‌گیریم.

Common Hashtags 1-9

```
1 // Users tweeting with your top hashtags
2 MATCH
3   (me:User:Me)-[:POSTS]→(tweet:Tweet)-[:TAGS]→(ht)
4 MATCH
5   (ht)←[:TAGS]-(tweet2:Tweet)←[:POSTS]-(sugg:User)
6 WHERE
7   sugg <> me
8   AND NOT
9   (tweet2)-[:RETWEETS]→(tweet)
10 WITH
11   sugg, collect(distinct(ht)) as tags
12 RETURN
```

	friend	common
1	"danielcfng"	23
2	"JAdP"	10
3	"ADyczkowsky"	6
4	"PuraPaper"	5


```

1 // Users tweeting with your top hashtags
2 MATCH
3 | (me:User:Me)-[:POSTS]→(tweet:Tweet)-[:TAGS]→(ht)
4 MATCH
5 | (ht)←[:TAGS]-(tweet2:Tweet)←[:POSTS]-(sugg:User)
6 WHERE
7 | sugg <> me
8 | AND NOT
9 | (tweet2)-[:RETWEETS]→(tweet)
10 WITH
11 | sugg, collect(distinct(ht)) as tags
12 RETURN
13 | sugg.screen_name as friend, size(tags) as common
14 ORDER BY
15 | common DESC
16 LIMIT 20

```

این کوئری، ابتدا همه توییت‌های ما که هشتگ دارند را می‌کند. سپس همه توییت‌های بقیه کاربرها که همان هشتگ‌ها را دارند را می‌کند.

سپس با اعمال یک فیلتر WHERE، مطمئن می‌شویم که کاربرهای دیگر شامل خود ما نمی‌شود، و پست آنها یک retweet از توییت ما نیست.

حال هشتگ‌های متمایز استفاده شده توسط هر کاربر را گرفته و در نهایت 20 سطر، شامل نام کاربری که از هشتگ ما استفاده کرده، و تعداد هشتگ‌های مشترکی که استفاده کرده را به صورت سورت نزولی ریترن می‌کنیم.