stackoverfloooooooow

Questions | Tags | Users | Badges | Unanswered | Ask Question

find paths with specific length combined with nodes with specific degree neo4j

asked   4 months ago
viewed   74 times
active   4 months ago

0

I am trying to write a query to find all nodes with outdegree X and only return the paths that contains those nodes when path length is equal to Y

If I want to get only nodes with outdegree X I use the following Cypher query

```
MATCH (s:URL)-[r:VISITED*]->(t:URL)
WITH s, count(t) as degreeout
WHERE 73 in s.job_id and degreeout <4
return s, degreeout
```

If I want to get only paths with length = X I use the following query

```
MATCH p=(s:URL)-[r:VISITED*]->(t:URL)
WHERE length(p)=7
return p
```

I tried the combine the previous two queries in the following query

```
MATCH (s:URL)-[r:VISITED*]->(t:URL)
WITH s, COLLECT(DISTINCT id(s)) as matched, count(t) as degreeout
WHERE 73 in s.job_id and degreeout <4
MATCH p=(s2:URL)-[r:VISITED*]-(t2:URL)
WHERE id(s2) in matched and length(p) >=1
RETURN p
```

Whenever I execute the query, the machine keeps processing and then I get an error no enough memory.

It seems like there is an infinite loop !!

neo4j   cypher   graph-databases

share improve this question

asked Apr 13 at 13:01
l.el-sayed
65 ● 1 ● 10

add a comment

1 Answer

active   oldest   votes

1

If you're just interested in traversing the relationship an exact number of times you can include that in the path expression:

```
MATCH p=(s:URL)-[r:VISITED*7]->(t:URL)
 return p
```

In general you should avoid doing traversals of unlimited length, i.e. `:VISITED*` . If you want to keep the depth variable because it's unknown it's good practice to set a max. value, i.e. `:VISITED*..7` .

If I understood correctly, your original query can be adjusted, just be setting the variable length to 7 in the path:

```
MATCH (s:URL)-[r:VISITED*7]->(t:URL)
WITH s, count(t) as degreeout
WHERE 73 in s.job_id and degreeout <4
return s, degreeout
```

Blog

Why Stack Overflow is a
Good Workplace for Women

Related

1 Optimize Neo4j Cypher
path finding with limited
paths in an undirected
graph

1 cypher query to find paths
between set of nodes takes
a very long time

2 Find all paths with smallest
weight if there are many
relations between two
nodes using cypher in
Neo4j

0 Extract all nodes from a
path query with cypher
(from Neo4j)?

0 Neo4j first node to meet
relationship in a movie
model

0 Finding Paths with Neo4j
Cypher Below a Given
Length, but Excluding
Those with an Nodes with a
Specific Property set to a
Specific Value

2 Neo4j: Sum relationship
properties where node
properties equal Value A
and Value B (intersection)

0 Find the path between "n"
nodes in neo4J using
cypher or java

1 Cypher - neo4j find
recursive end nodes

You should see some performance improvement because now paths of length > 7 will be excluded from the results and they won't be traversed. Again, always avoid unlimited depth traversals unless there is a really good reason for it and you have enough computer resources + time for the query to complete.

Regarding performance best practices, this query will still not perform very well, since it's forcing a graph scan to find the start node. I understand the nodes labeled URL contain a property `job_id` of type Array because of the `in` operator? Neo4j needs to read all URL nodes and their properties, then scan through those arrays just to find the start node.

I would recommend changing your data model to use schema indexes based on an exact property value. Example:

```
(j:Job {job_id: 73})-[:SOMETHING?]->(u:URL {...})
```

We would also add a schema index:

```
CREATE INDEX ON :Job(job_id)
```

Then you can query like this:

```
MATCH (j:Job {job_id: 73})-[:SOMETHING?]->(s:URL)-[r:VISITED*7]->(t:URL)
WITH s, count(t) as degreeout
WHERE degreeout <4
return s, degreeout
```

share improve this answer

answered Apr 13 at 14:38

albertoperdomo
1,132 ● 3 ● 9

I label the starting node as STARTING_URL ... so the starting node will have two labels "URL" and "STARTING_URL" – l.el-sayed Apr 14 at 6:19

Regarding the job_id I can't have it as an int property value. I should have it as an array property. I don't think that this way (j:Job {job_id: 73}) (property as dictionary) would enable me to handle the job_id attribute as an array ... !?? – l.el-sayed Apr 14 at 6:29

The problem with that approach is that locating a node by an array property is not efficient, since it won't use any indexes and Neo4j will have to scan all nodes with matching labels. If you have n job_ids for each STARTING_URL, then you can probably create one node for each job_id and have a relationship to the STARTING_URL as I tried to explain in the last section of my answer. This approach would take more advantage of Neo4j's graph performance capabilities. – albertoperdomo Apr 14 at 18:50

I did create the JOB node as u suggested and changed the design. I just have a question now. How I can return all paths with length >=3 and the same time exclude all nodes with outdegree > 10?? How I can do this in a cypher query...?? – l.el-sayed Apr 14 at 23:16

You can add a minimum depth: MATCH (j:Job {job_id: 73})-[:SOMETHING?]->(s:URL)-[r:VISITED*3..]->(t:URL) WITH s, count(t) as degreeout WHERE degreeout <11 return s, degreeout – albertoperdomo Apr 15 at 9:01

show **1** more comment

Your Answer

| **B** | *I* | | 🔗 | 66 | {} | 🖼 | | ⌗ | ⌗ | ⌗ | ⌗ | | ↩ | ↪ |

Post Your Answer

*By posting your answer, you agree to the *privacy policy* and *terms of service*.*

Not the answer you're looking for? Browse other questions tagged neo4j   cypher   graph-databases or
ask your own question.

tour   help   blog   chat   data   legal   privacy policy   work here   advertising info   mobile   contact us   feedback

TECHNOLOGY

Stack Overflow

Server Fault

Super User

Web Applications

Ask Ubuntu

Webmasters

Game Development

TeX - LaTeX

Programmers

Unix & Linux

Ask Different (Apple)

WordPress
Development

Geographic Information
Systems

Electrical Engineering

Android Enthusiasts

Information Security

Database
Administrators

Drupal Answers

SharePoint

User Experience

Mathematica

Salesforce

ExpressionEngine®
Answers

more (13)

LIFE / ARTS

Photography

Science Fiction &
Fantasy

Graphic Design

Movies & TV

Seasoned Advice
(cooking)

Home Improvement

Personal Finance &
Money

Academia

more (9)

**CULTURE /
RECREATION**

English Language &
Usage

Skeptics

Mi Yodeya (Judaism)

Travel

Christianity

Arqade (gaming)

Bicycles

Role-playing Games

more (21)

SCIENCE

Mathematics

Cross Validated (stats)

Theoretical Computer
Science

Physics

MathOverflow

Chemistry

Biology

more (5)

OTHER

Stack Apps

Meta Stack Exchange

Area 51

Stack Overflow Careers