**BIG DATA REPORT**

First of all, csv file which is gene\_associations.csv is loaded and node type is given as TrialGene and edges are defined as  AssociationType when it loaded.This query is used:

Note that :CSV file is put in “neo4jDatabases\Your-graph-place \installation-3.5.5\import” .

LOAD CSV WITH HEADERS FROM "file:///C:/gene\_associations.csv" AS line

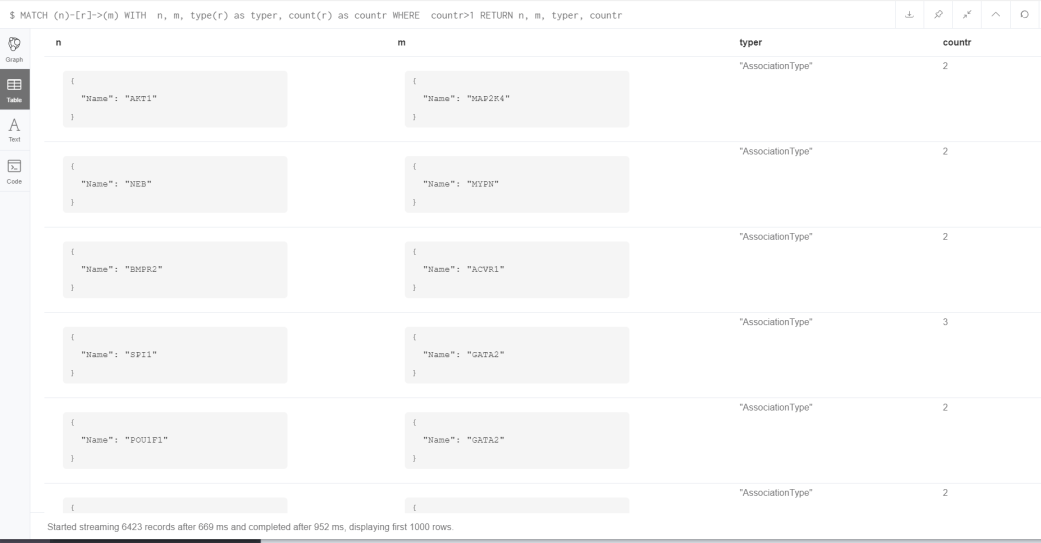
MERGE (n:TrialGene {Name:line.OFFICIAL\_SYMBOL\_A})

MERGE (m:TrialGene {Name:line.OFFICIAL\_SYMBOL\_B})

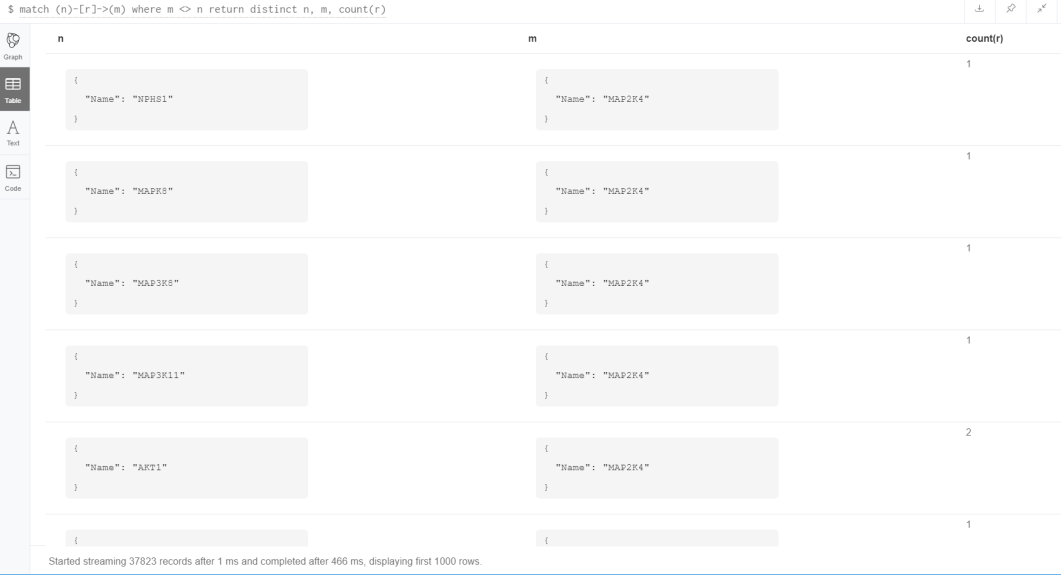
MERGE (n) -[:AssociationType { AssociatedWith:line.EXPERIMENTAL\_SYSTEM}]-> (m)

This steps are followed:

1. After that ,for counting node we use this query: MATCH (n) RETURN count(\*) -9656 node
2. For calculating number of loops in the graph this query is used:



3) ”match (n)-[r]->(m) where m <> n return distinct n, m, count(r) “query’s result.

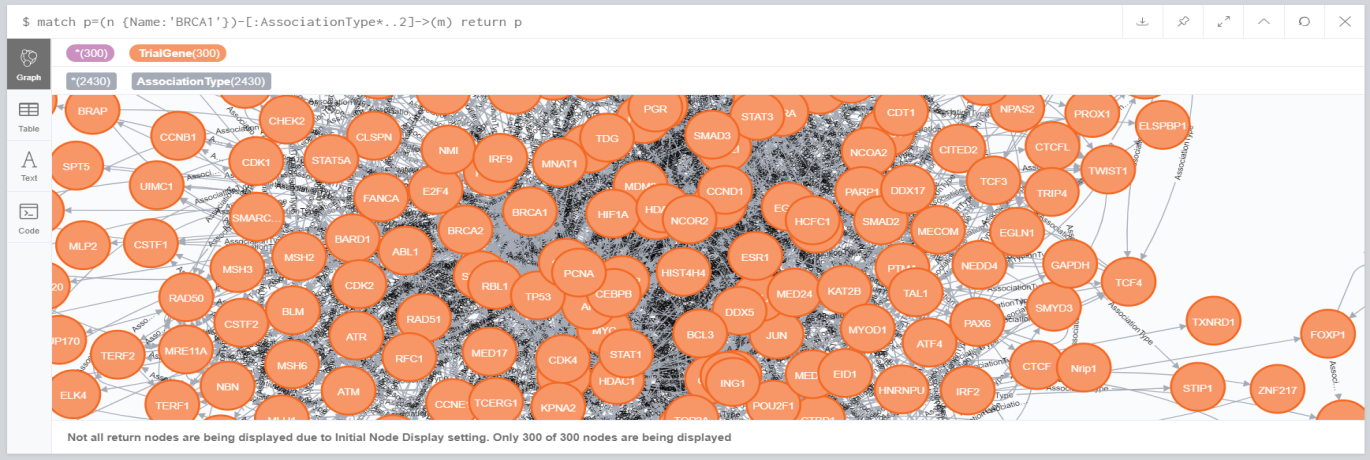


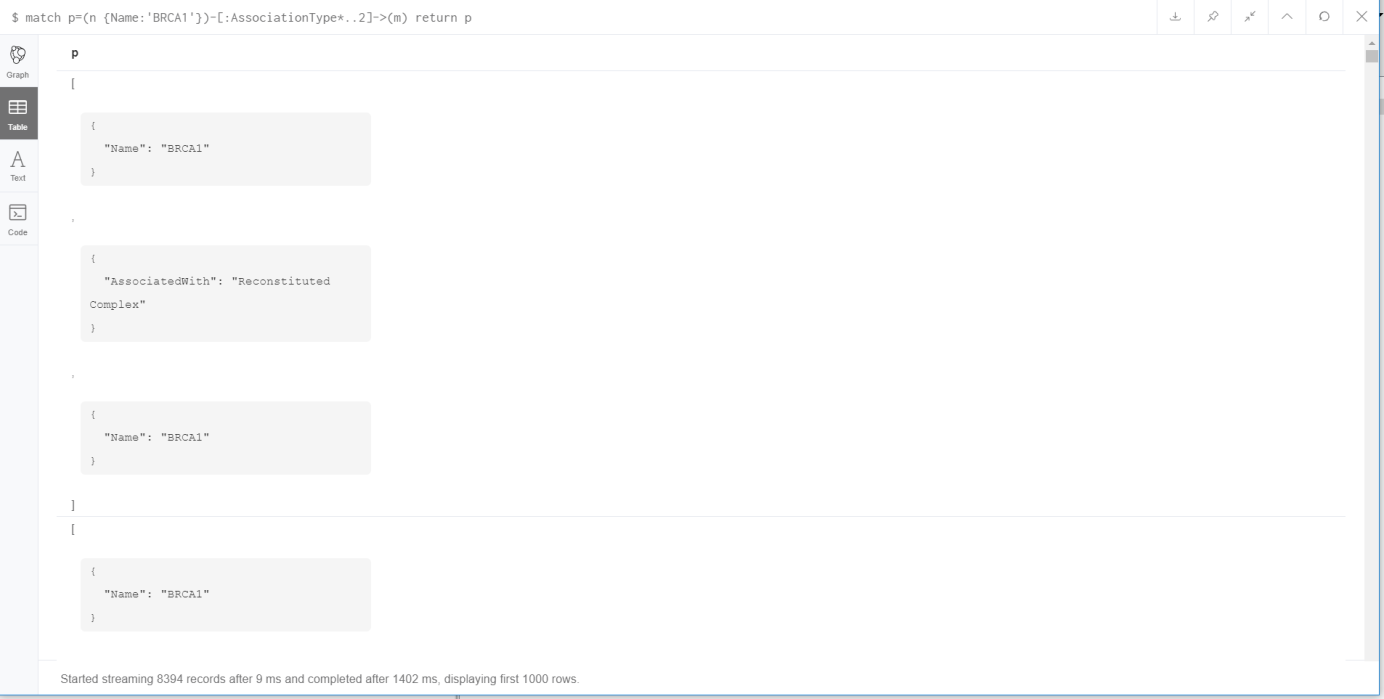
4) For finding number of nodes which are not included loop edges this query is used in the third step.

5 ) “match (n)-[r]->(m) where m <> n return distinct n, m, count(r) as myCount order by myCount desc limit 1” , report of the results:

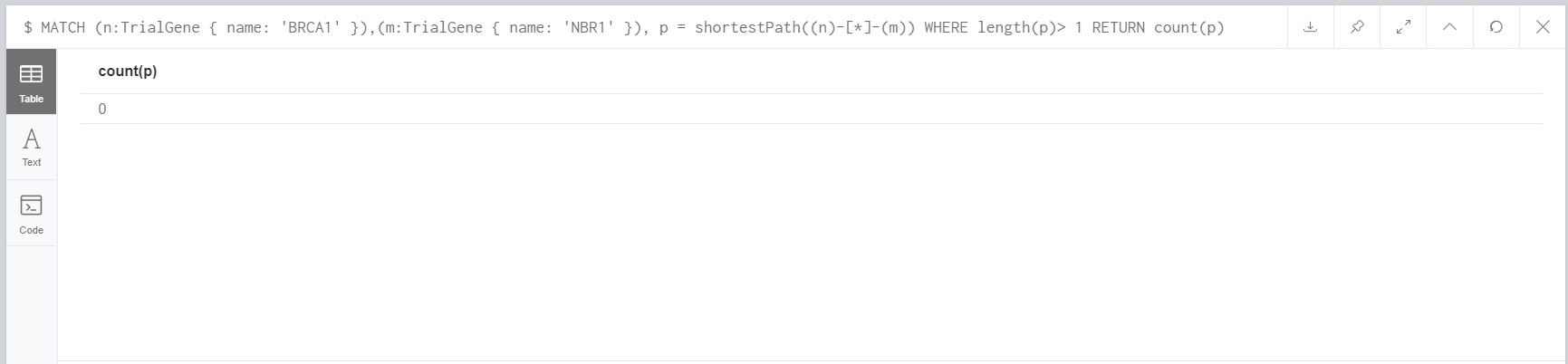


6)” match p=(n {Name:'BRCA1'})-[:AssociationType\*..2]->(m)”

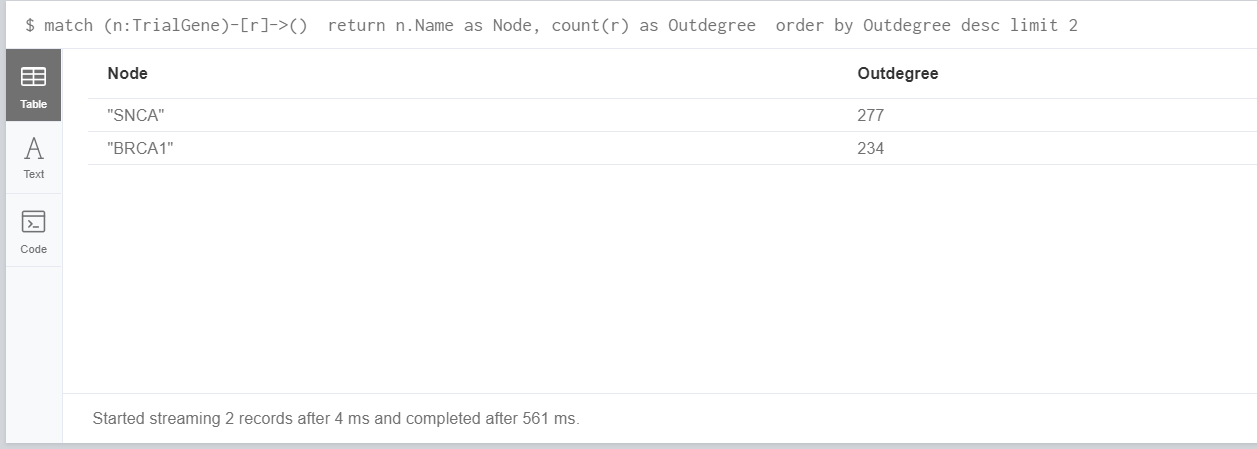




Returns all nodes related to **‘BRCA1’** by 0 to 2 hops.

7)Shortest path of between ‘BRCA1’ node and ‘NBR1’ node.

8) Top 2 nodes with the highest outdegree



9) Histogram of node’s degrees

Nodes which have degree of 3:

