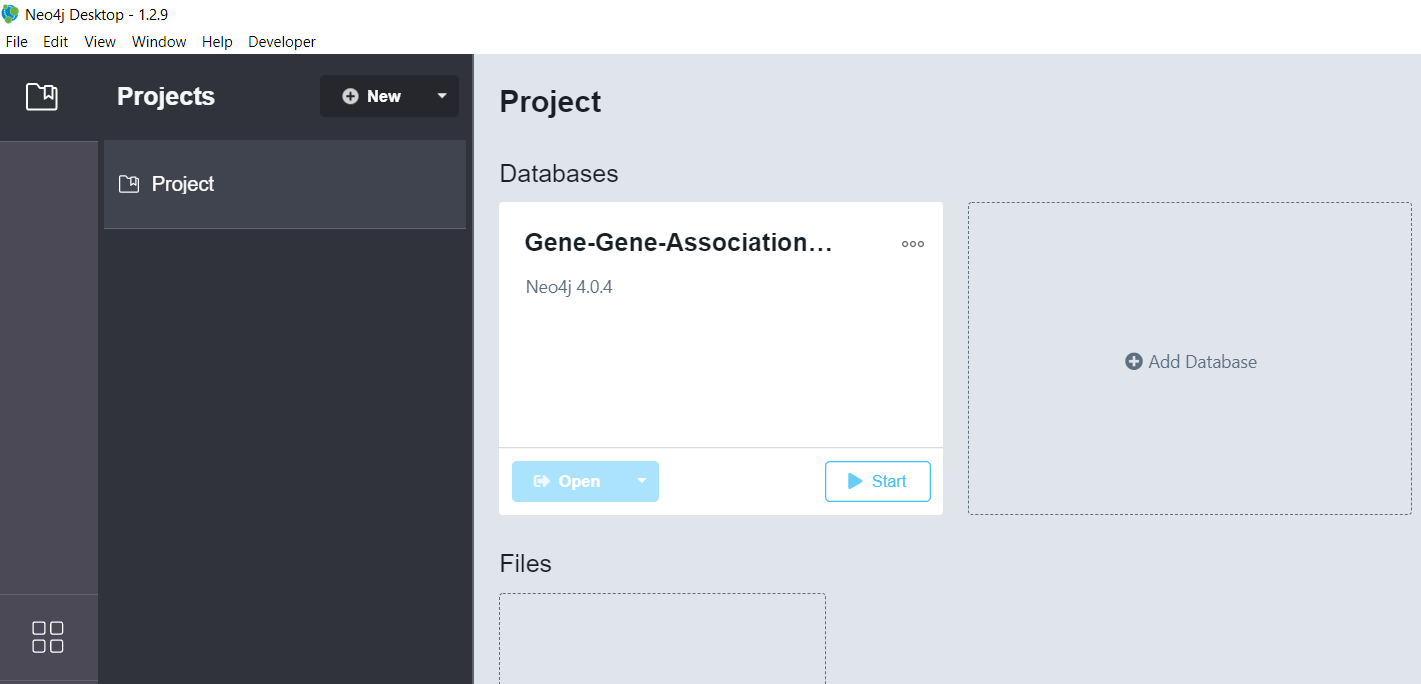
**درس : Big Data استاد : خانم دکتر صفری**

**دانشجو : رقیه یزدانی تمرین : Neo4j**

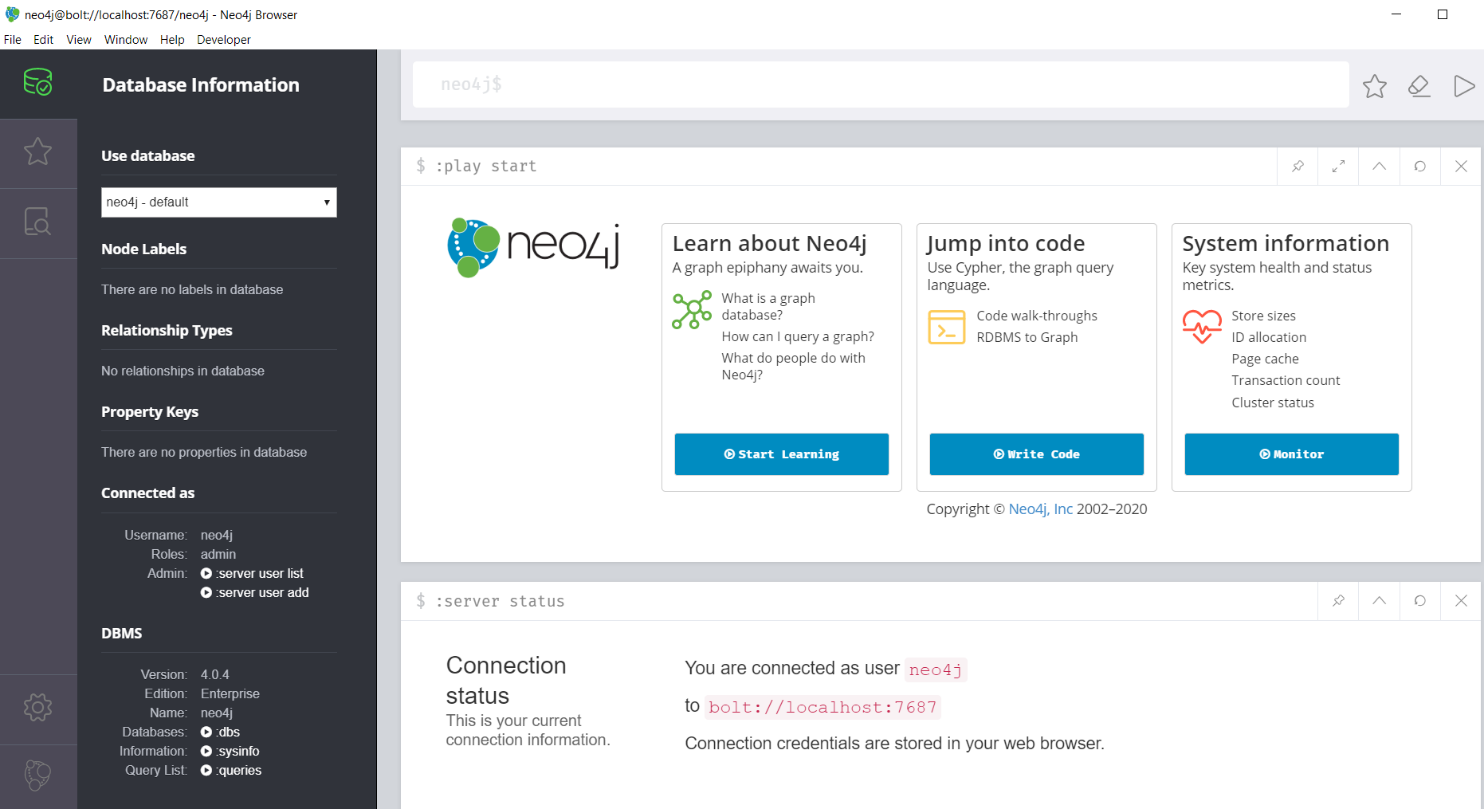
# Neo4j

# نصب Neo4j و import دیتاست

ابتدا از آدرس <https://neo4j.com/download/> نسخه Neo4j Desktop که free می‌باشد، دانلود و نصب گردید.(حجم فایل دانلودی حدود 530 MB است). پس از نصب ابتدا یک Project ساخته شده و در آن با کلیک بر روی گزینه Add Database، دیتابیس موردنظرمان را از نوع «Create a Local Graph»با نام Gene-Gene-Associations-50k می‌سازیم.



حال با انتخاب گزینه Start، دیتابیس را start کرده و سپس با زدن کلید open، دیتابیس روی پورت 7687 در browser بالا می‌آید که مشابه شکل زیر است. همانگونه که مشاهده می‌شود، هیچ Node و Relationship ای وجود ندارد. در قسمت بالای صفحه و بخش neo4j$ امکان تایپ دستورات وجود دارد که با زدن کلید ctrl+enter اجرا خواهند شد.



ابتدا فایل gene\_gene\_associations\_50k.csv به مسیر import در پوشه‌ای که دیتابیس ساخته شده که بصورت زیر بوده، کپی شد:

F:\Programs\.Neo4jDesktop\neo4jDatabases\database-81944467-6bc1-469a-a0dd-337376ae0f61\installation-4.0.4\import

سپس دستور زیر در محیط neo4j browser نوشته و اجرا شد:

LOAD CSV WITH HEADERS FROM "file:///gene\_gene\_associations\_50k.csv" AS row

merge(n:GeneNode{Name:row.OFFICIAL\_SYMBOL\_A})

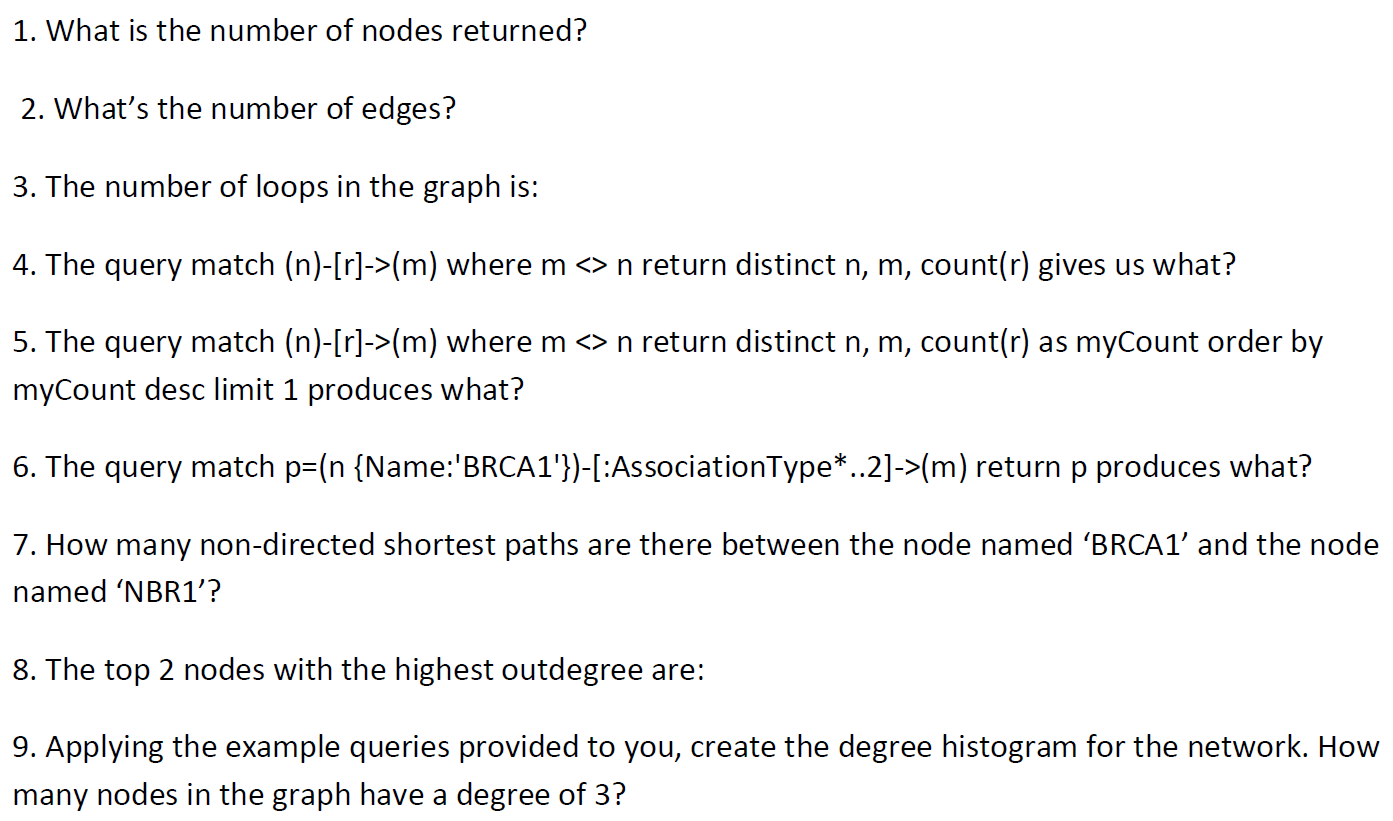
merge(m:GeneNode{Name:row.OFFICIAL\_SYMBOL\_B})

merge(n)-[:TO{experimental\_system:row.EXPERIMENTAL\_SYSTEM}]->(m)

پس از اجرای دستورات بالا که چند دقیقه ای زمان می‌برد، تعداد 9656 نود و تعداد 46621 رابطه ایجاد می‌شود.

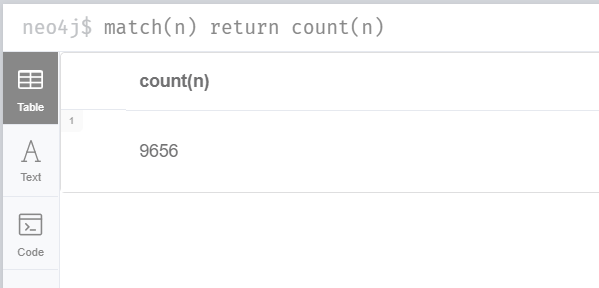
# تمرینها

باید بر اساس داده های import شده به سوالات زیر پاسخ داده شود :



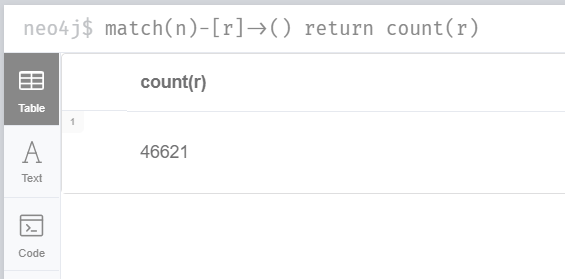
# تمرین اول

برای تعیین تعداد نودها؛ query زیر اجرا می‌شود :



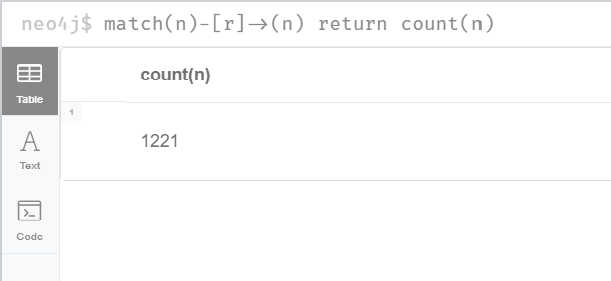
# تمرین دوم

برای تعیین تعداد یالها یا رابطه ها query زیر اجرا می‌شود :



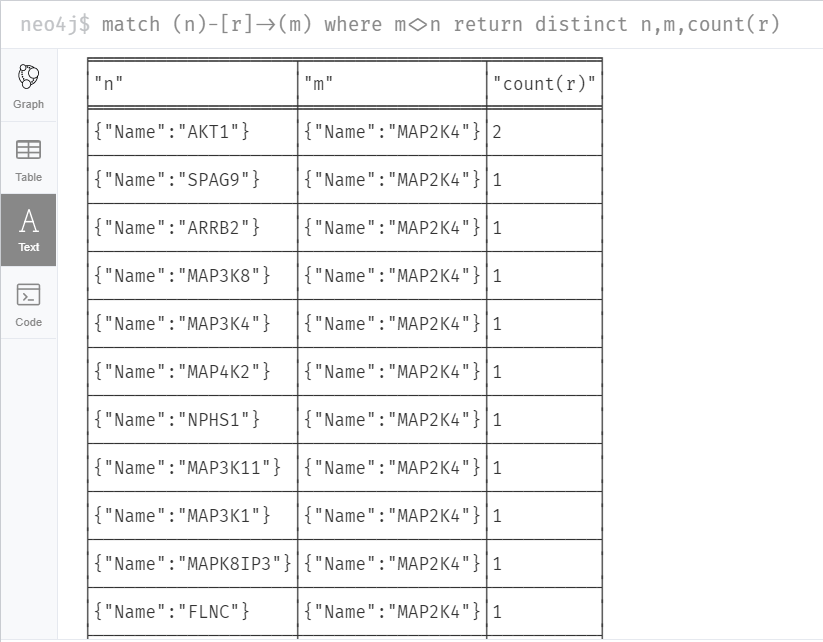
# تمرین سوم

# برای تعیین تعداد loop ها، اسکریپت زیر نوشته شد:



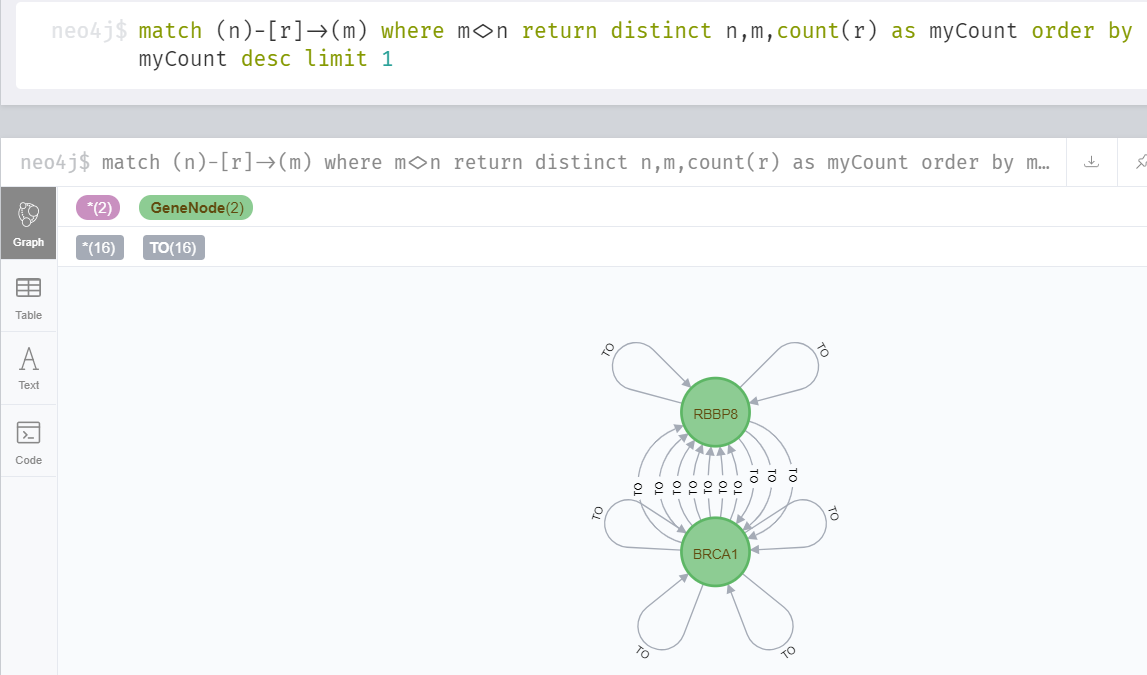
# تمرین چهارم

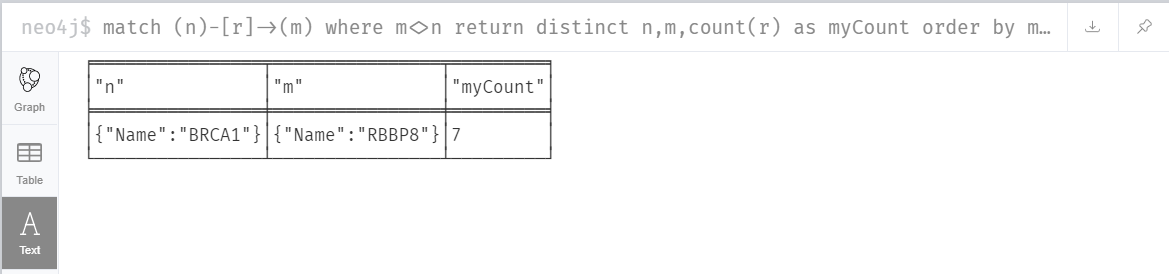
در این query، روابطی که به شکل loop نیستند، شناسایی شده و به ازای هر ترکیب غیر تکراری از دو نود مبدأ و مقصد، تعداد روابط بین آنها آورده شده است.



# تمرین پنجم

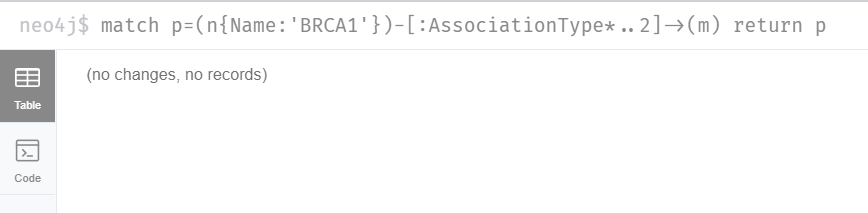
در این query، بخش اول آن، مشابه تمرین قبلی است و شرایطی به انتهای آن اضافه شده است که بدین معناست که خروجی تمرین بالا بصورت نزولی بر اساس ستون myCount که به عنوان یک alias برای count(r) تعریف شده مرتب شود و سپس اولین سطر آن به عنوان خروجی برگردانده شود که نشان می دهد بیشترین تعداد رابطه از نود BRCA1 به نود RBBP8 به تعداد 7 رابطه می باشد که گراف متناظر خروجی هم، این را نشان می دهد. البته در گراف سایر روابط این دو نود هم نشان داده شده است.





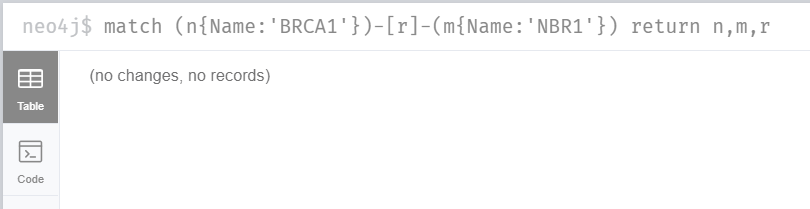
# تمرین ششم

این query خروجی ندارد و بدین معناست که از نود BRCA1 به هر نود دیگری رابطه ای از نوع AssociationType تا دو همسایه بعد وجود ندارد.



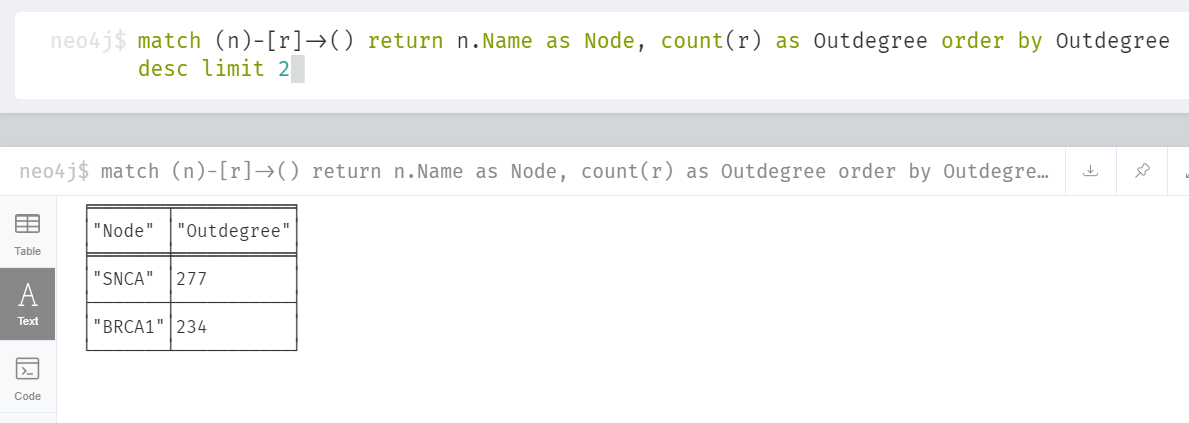
# تمرین هفتم

این query بصورت زیر نوشته شد و خروجی ندارد. می توان بعد از return نوشت count(r). ولی در هر حال خروجی ندارد و یعنی اینکه بین این دو نود، رابطه ای برقرار نیست چه از BRCA1 به NBR1 و یا از NBR1 به BRCA1.



# تمرین هشتم

برای این تمرین، query بصورت زیر نوشته شد تا دو نود با بالاترین outdegree (تعداد رابطه های خروجی از نود) بدست آید:



# تمرین نهم

برای این تمرین برای هر نود، degree (تعداد رابطه های ورودی و خروجی) آن محاسبه می‌شود و سپس نودهای دارای degree یکسان گروه‌بندی شده و بصورت صعودی بر اساس میزان degree نمایش داده می‌شوند. در شکل زیر، query و نتیجه آن آورده شده است. همانطور که پیداست، تعداد 821 نود دارای degree برابر 3 می‌باشند.

