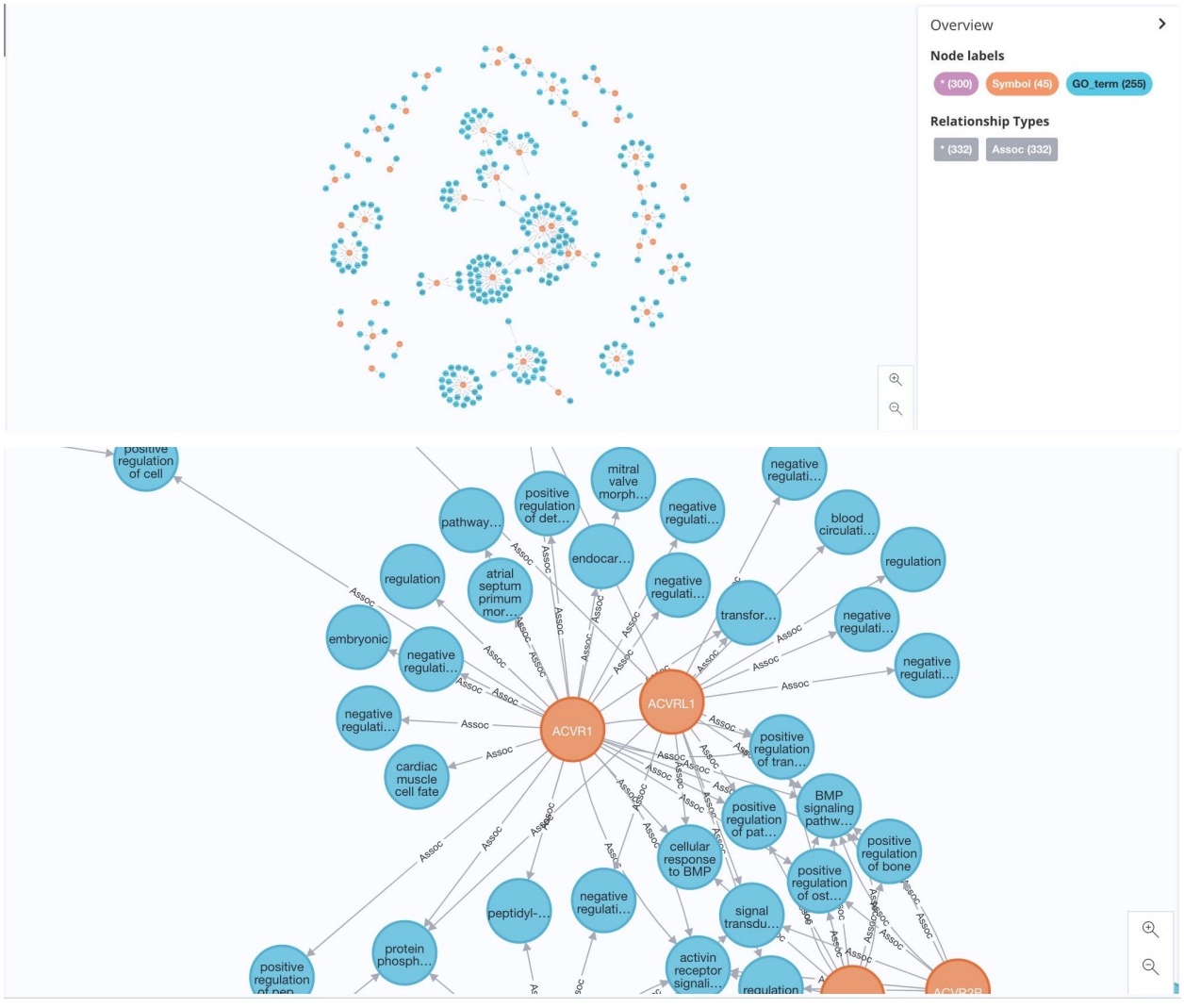
DS4300: HW4 Report

Sara Adra, Anika Das, Mirah Gordon, Genny Jawor

STEP 2: Build a graph model

```
call
apoc.load.jdbc("jdbc:mysql://localhost:3306/HW4?serverTimezone=EST5EDT&user=root&password=p@ssword", "genes") YIELD row
merge (gene:Symbol {Symbol:row.Symbol})
merge (goTerm:GO_term {GO_term:row.GO_term, id:row.GO_ID})
merge (gene)-[:Assoc {qualifier:row.qualifier, evidence:row.evidence}]->(goTerm)
return gene, goTerm
```



Graph Design

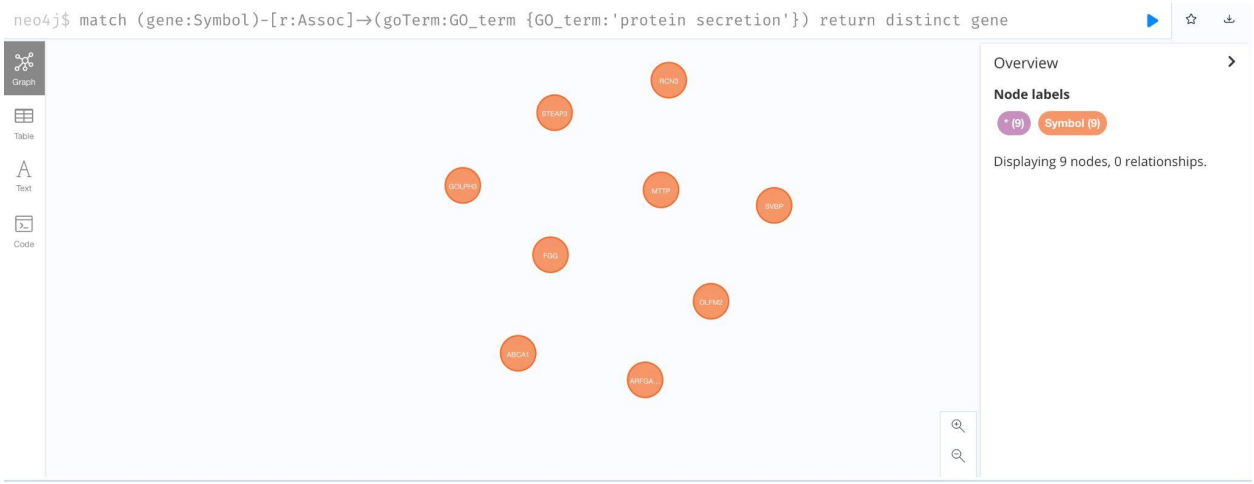
The way in which we chose to design this graph is by having the genes and GO terms as nodes where the gene node has the gene symbol as its property and the GO term node has the GO term and the GO ID as properties. We also created a relationship between the genes and go terms (where the genes are associated with the go terms) with the relationship's properties being the qualifier (ex. involved_in, NOT involved_in, etc) and evidence (ex. IDA, IEP, etc).
We did choose to filter on gene symbols that contained "LOC####", per the suggestion of step 1, in order to make our query answers more meaningful.

STEP 3: Query the database using Cypher

1. Pick a function (GO term) that seems interesting to you. What genes are linked to that GO term?

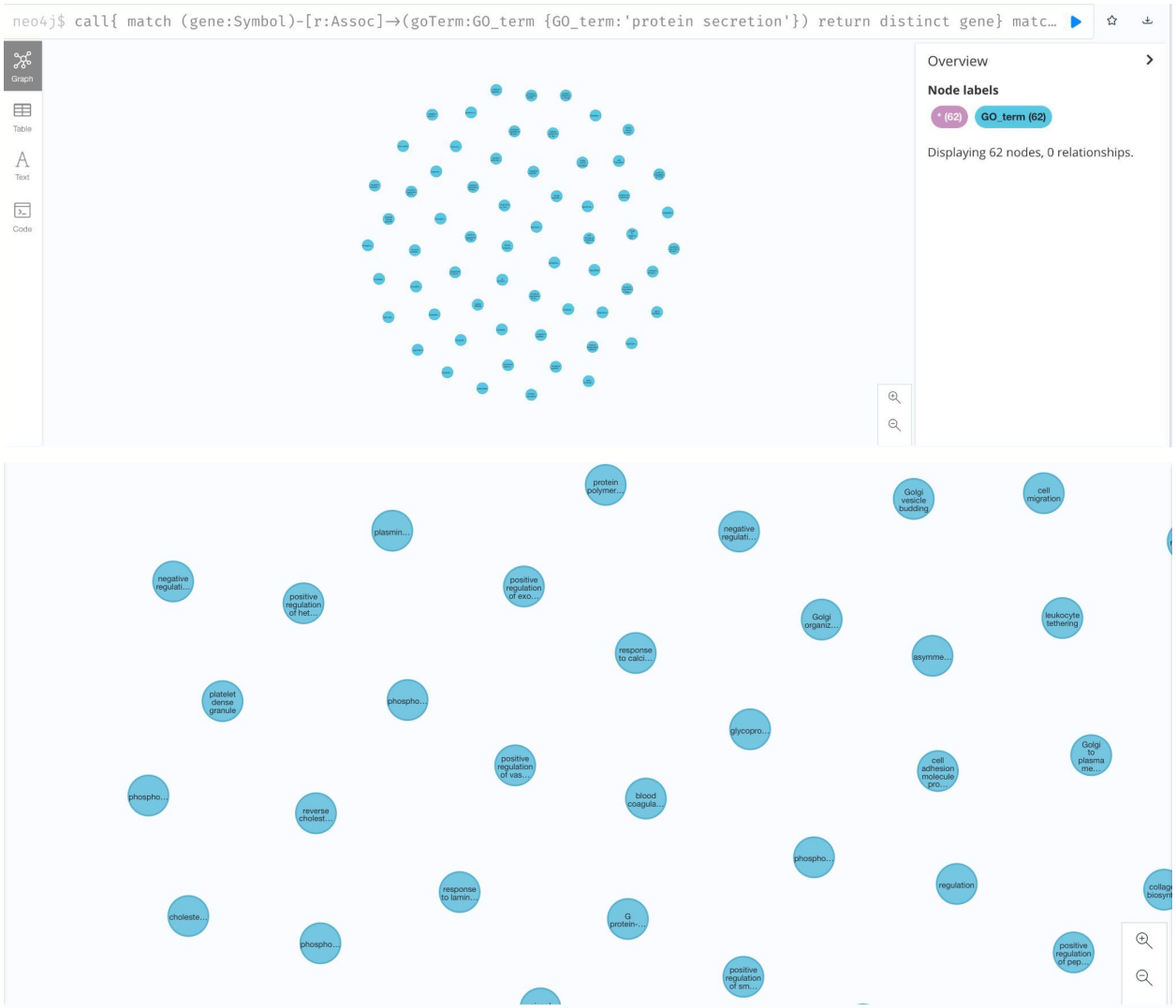
The GO term that we selected is protein secretion. Below is the query that determines what genes are linked to this GO term of protein secretion.

```
match (gene:Symbol)-[r:Assoc]->(goTerm:GO_term {GO_term:'protein secretion'})
return distinct gene
```



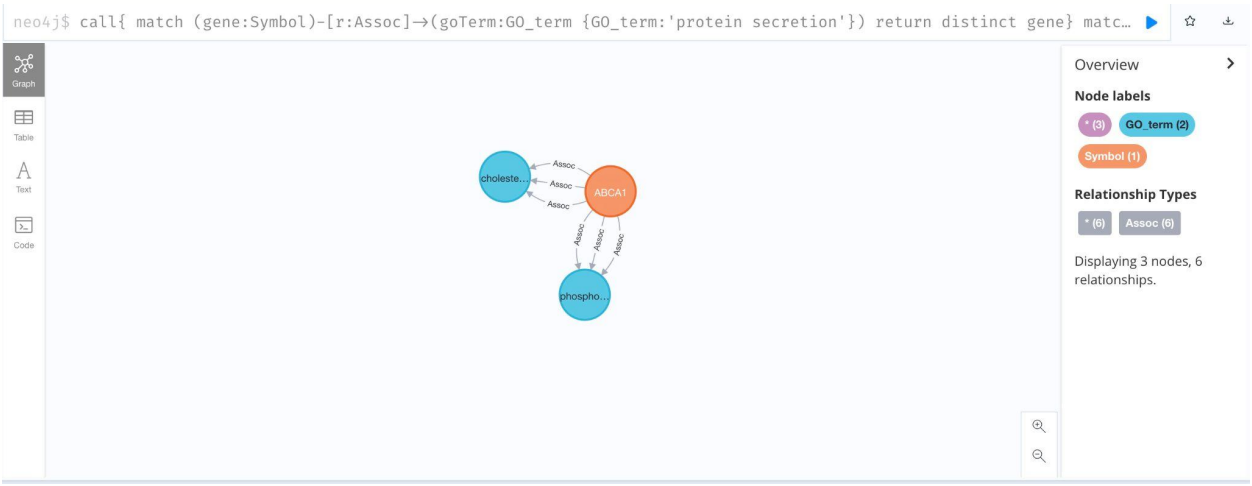
2. What other GO Terms are also associated with the genes you just identified?

```
call{
match (gene:Symbol)-[r:Assoc]->(goTerm:GO_term {GO_term:'protein secretion'})
return distinct gene}
match(gene:Symbol)-[r:Assoc]->(goTerm: GO_term)
return distinct goTerm
```



3. For the top two GO terms found in 2, what genes do they all have in common?

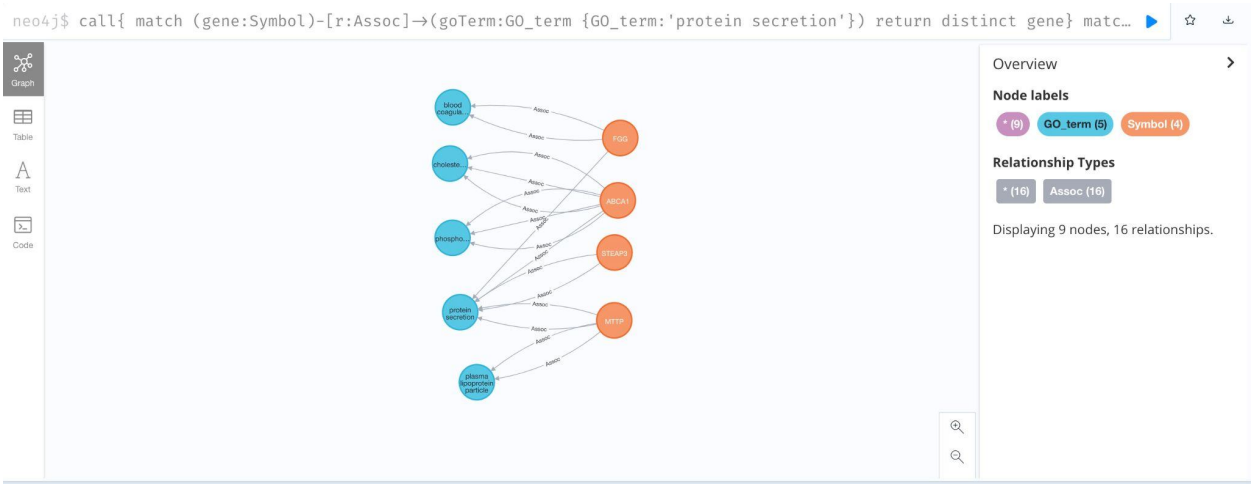
```
call{
match (gene:Symbol)-[r:Assoc]->(goTerm:GO_term {GO_term:'protein secretion'})
return distinct gene}
match(gene:Symbol)-[r:Assoc]->(goTerm: GO_term)
return goTerm, gene, COUNT(goTerm) as occurence_count
ORDER BY occurence_count DESC LIMIT 2
```



They have gene ABCA1 in common with each other.

4. Construct a bipartite graph between your term-linked genes and the top five GO terms found to be commonly associated with these genes.

```
call{
match (gene:Symbol)-[r:Assoc]->(goTerm:GO_term {GO_term:'protein secretion'})
return distinct gene}
match(gene:Symbol)-[r:Assoc]->(goTerm: GO_term)
return goTerm, gene, COUNT(goTerm) as occurence_count
ORDER BY occurence_count DESC LIMIT 5
```



5. Do an internet search. Is there any reason to think that these biological processes are linked somehow?

We found that there was scientific evidence on the connection of these processes and their similar genes. While this holds true for multiple genes and processes, one instance of this is with the ABCA1 gene. The following information from Medline Plus discusses how the gene "belongs to a group of genes called the ATP-binding cassette family, which provides instructions for making proteins that transport molecules across cell membranes." Further, the protein "moves cholesterol and certain fats called phospholipids across the cell membrane to the outside of the cell," and both phospholipid efflux and cholesterol efflux are seen as two of the top five GO_terms.
Source: https://medlineplus.gov/genetics/gene/abca1/

STEP 4: Exploration and Discovery

a) Ask a specific question

What genes are linked to positive or negative regulation of apoptotic processes?

b) Define a cypher query that produces a result

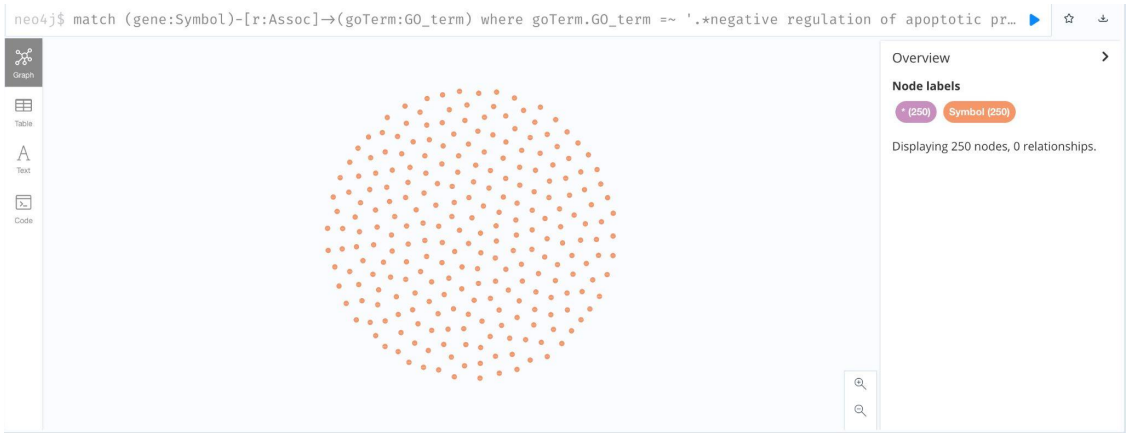
```
match (gene:Symbol)-[r:Assoc]->(goTerm:GO_term)
where goTerm.GO_term =~ '.*negative regulation of apoptotic process.*'
return distinct gene

match (gene:Symbol)-[r:Assoc]->(goTerm:GO_term)
where goTerm.GO_term =~ '.*positive regulation of apoptotic process.*'
return distinct gene
```

c) Present your answer

We found that there are 188 distinct genes linked to positive regulation of apoptotic processes and 250 genes linked to the negative regulation. This is shown in the below pictures and makes intuitive sense, given how important something like apoptosis is inside the body.

Negative Regulation:



Positive Regulation: