

# 22S-CSB150-EEB159 Lab8

LISA WANG

TOTAL POINTS

**105 / 105**

QUESTION 1

**1 Ex.1 20 / 20**

**✓ - 0 pts** *Correct*

**- 5 pts** Missing one network

**✓ - 0 pts** *Correct*

**- 5 pts** Click here to replace this description.

QUESTION 2

**2 Ex.2 20 / 20**

**✓ - 0 pts** *Correct*

QUESTION 3

**3 Ex.3 20 / 20**

**✓ - 0 pts** *Correct*

**- 4 pts** Correctly argue that the random network

doesn't look like the real one

**- 2 pts** Include pictures of the graphs

QUESTION 4

**4 Ex.4 20 / 20**

**✓ - 0 pts** *Correct*

**- 20 pts** not attempted

**- 2 pts** Include node and edge files

QUESTION 5

**5 Ex.5 20 / 20**

**✓ - 0 pts** *Correct*

**- 20 pts** Not attempted

QUESTION 6

**6 Ex.6 (optional) 5 / 5**

Lisa Wang  
UID: 105502901  
Lab 8 Submission

Exercise 1  
1. CCLE data

**Nodes:** 297

**Edges:** 2345

Directed Graph

2. Yeast

**Nodes:** 2361

**Edges:** 7182

Directed Graph

3. Diseaseasome

**Nodes:** 1419

**Edges:** 3926

Directed Graph

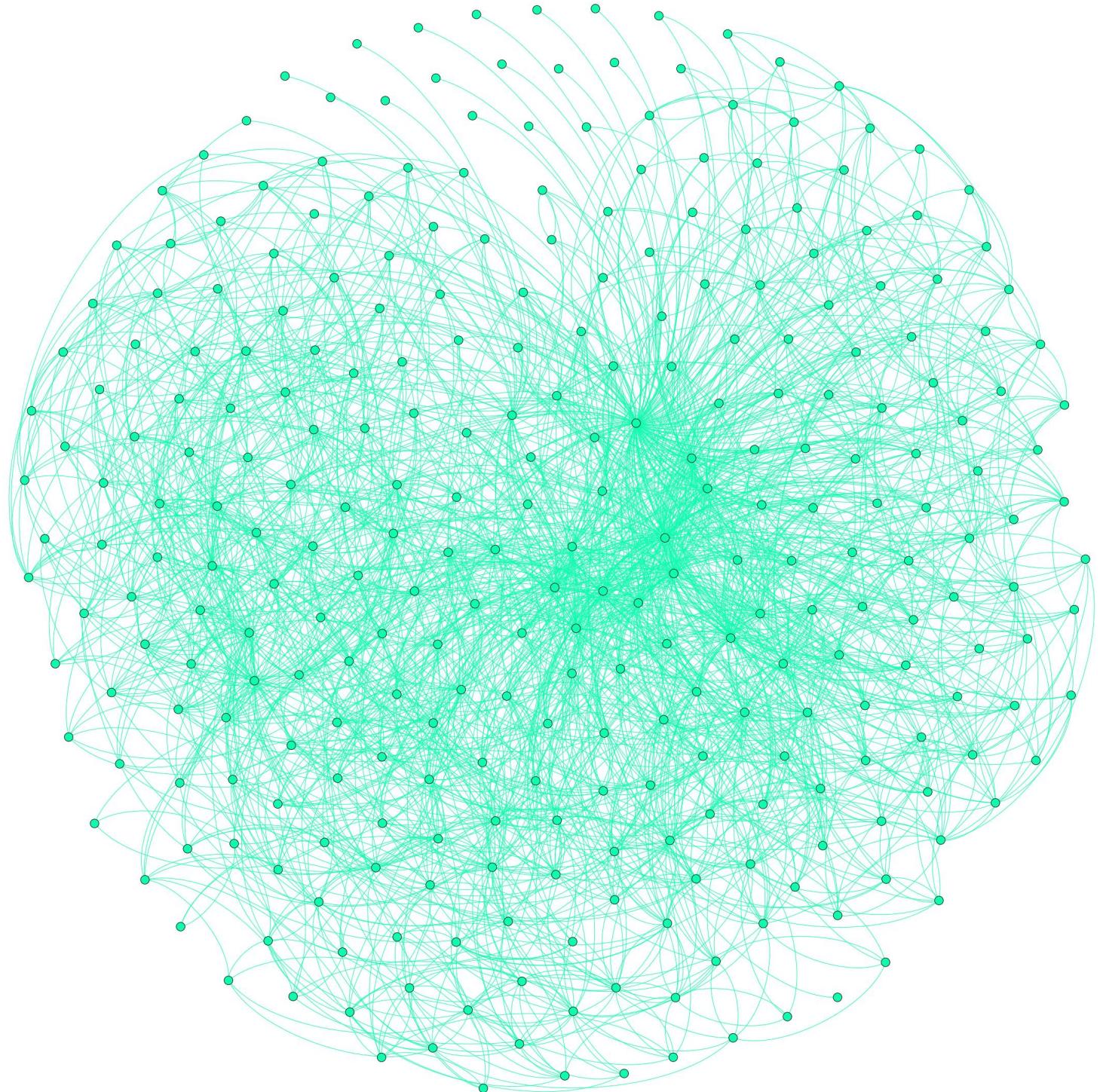
4. C. Elegans

**Nodes:** 306

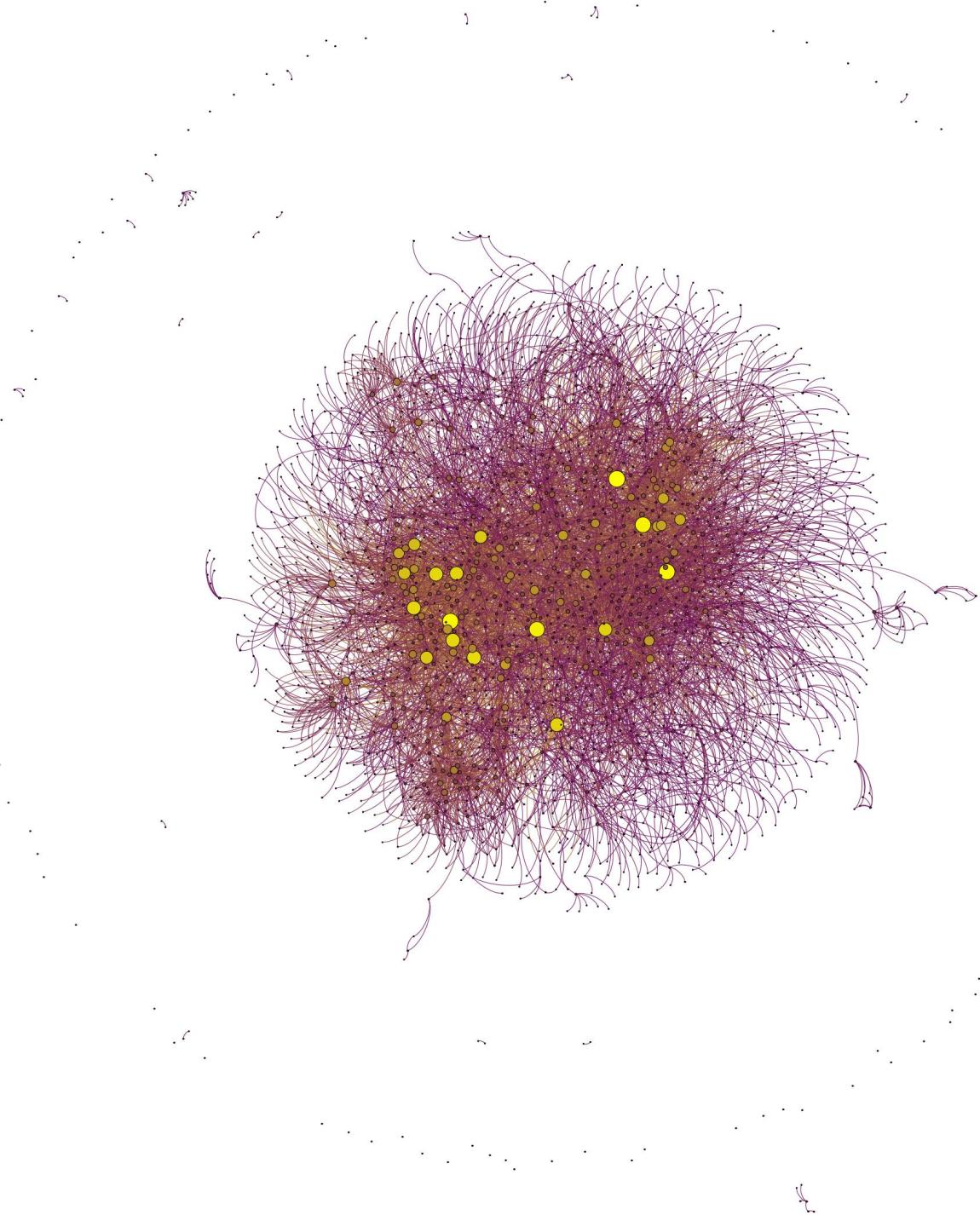
**Edges:** 2345

Directed Graph

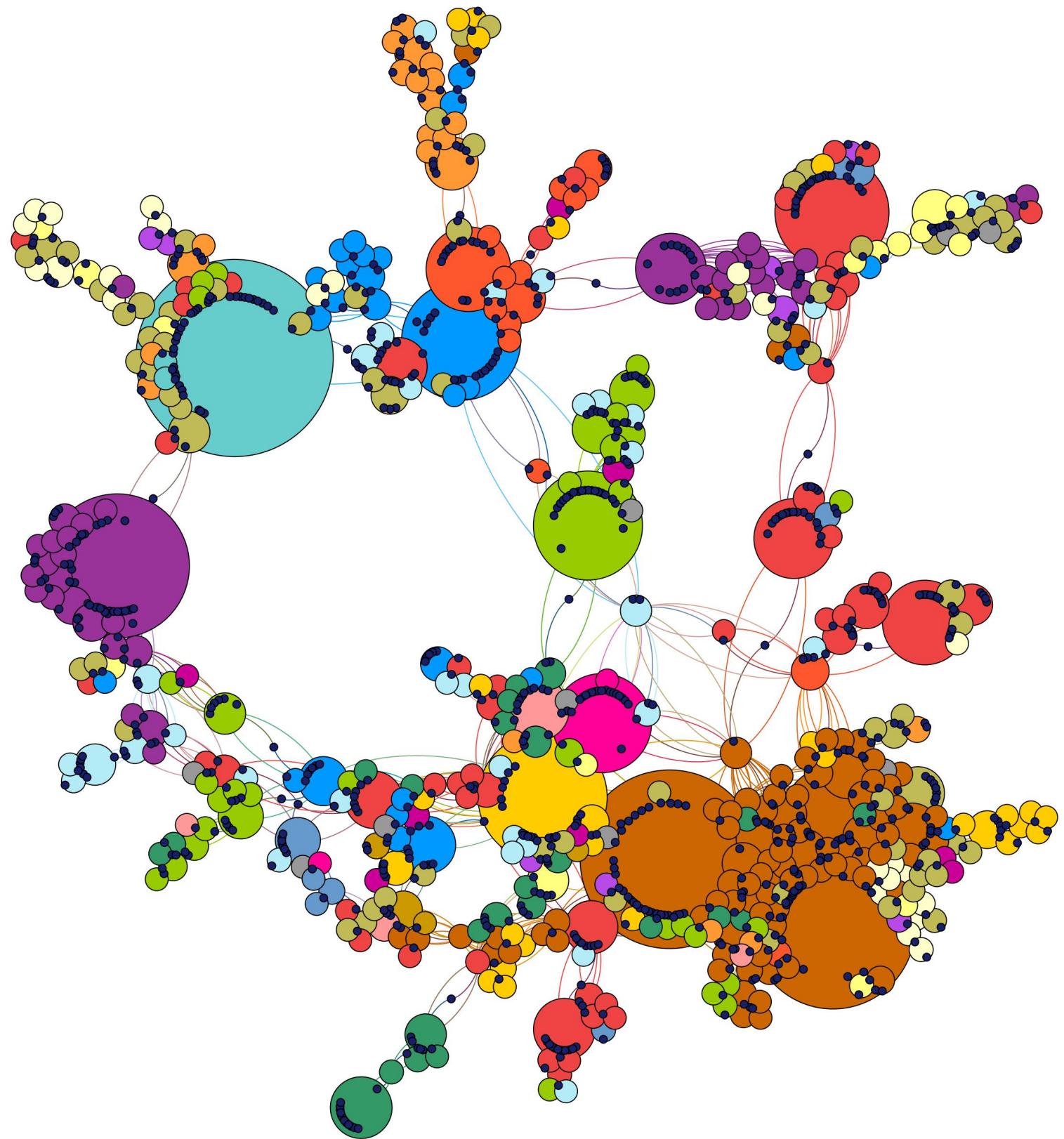
Exercise 2  
Comparison table



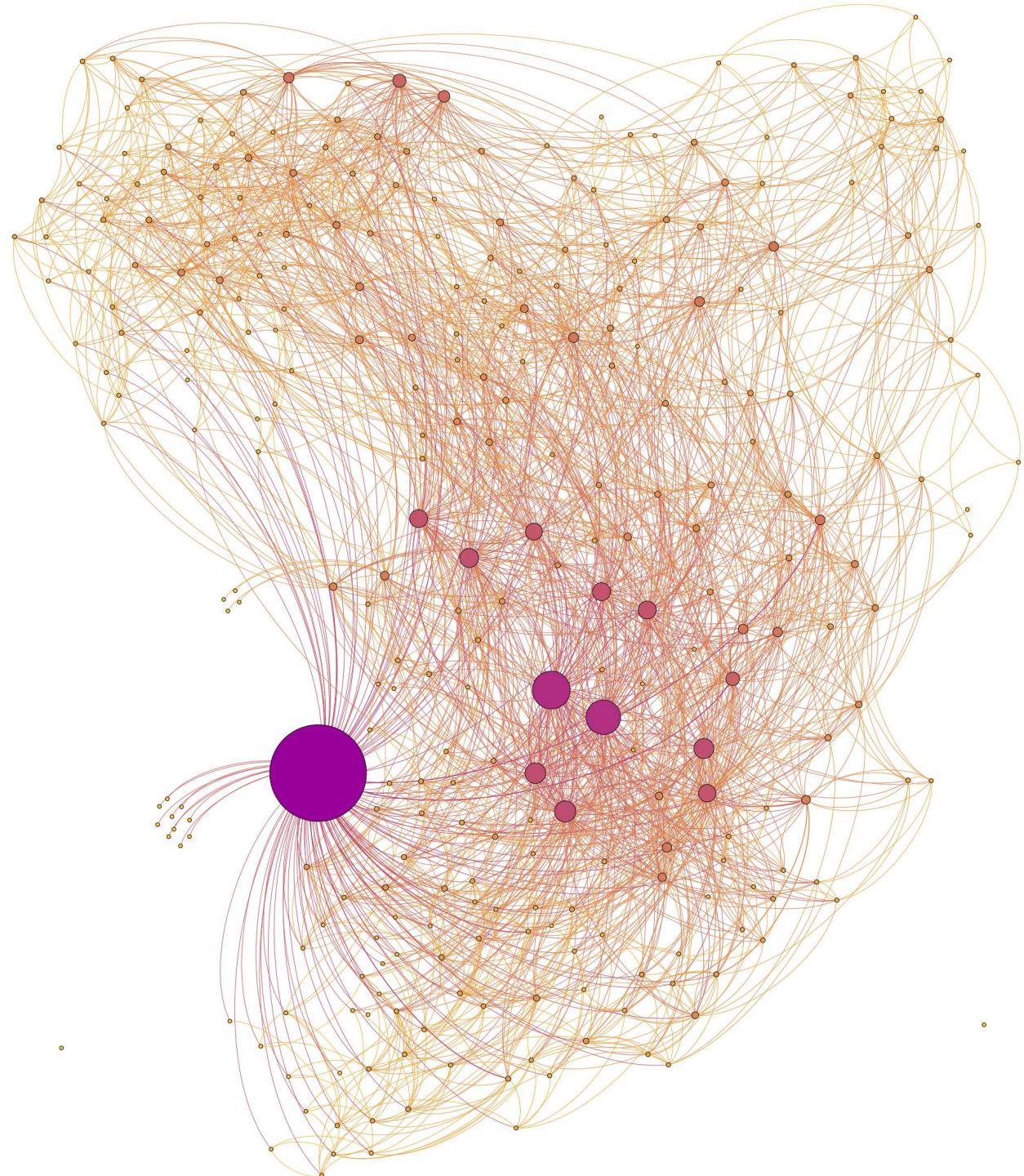
# Yeast



# Diseasome



*C. Elegans*



1 Ex.1 20 / 20

✓ - 0 pts Correct

- 5 pts Missing one network

Lisa Wang  
UID: 105502901  
Lab 8 Submission

Exercise 1  
1. CCLE data

**Nodes:** 297

**Edges:** 2345

Directed Graph

2. Yeast

**Nodes:** 2361

**Edges:** 7182

Directed Graph

3. Diseaseasome

**Nodes:** 1419

**Edges:** 3926

Directed Graph

4. C. Elegans

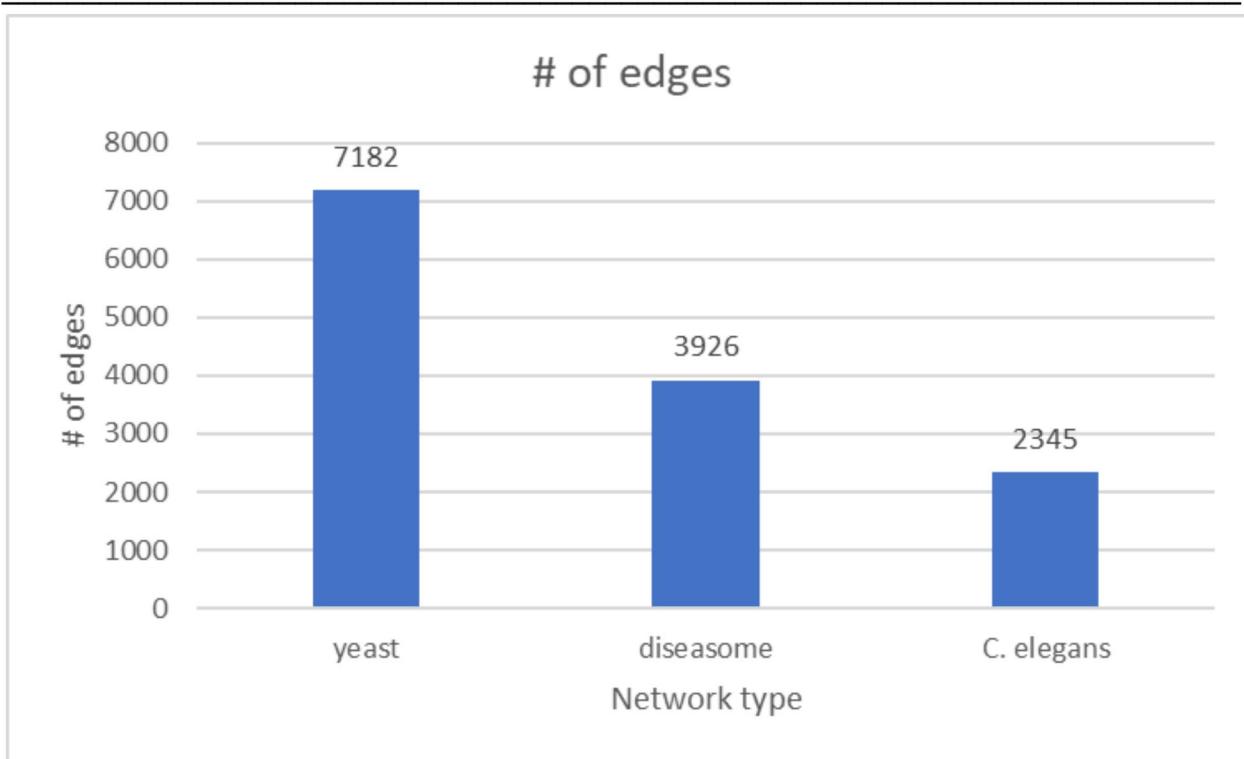
**Nodes:** 306

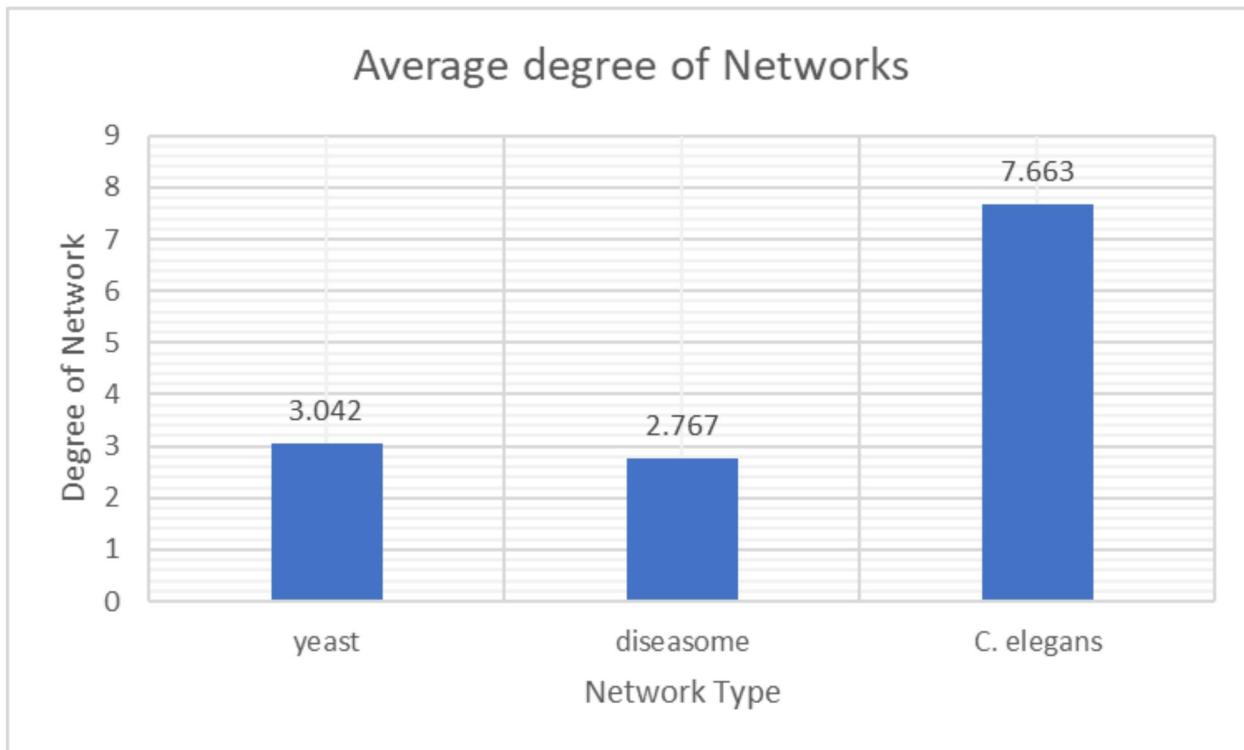
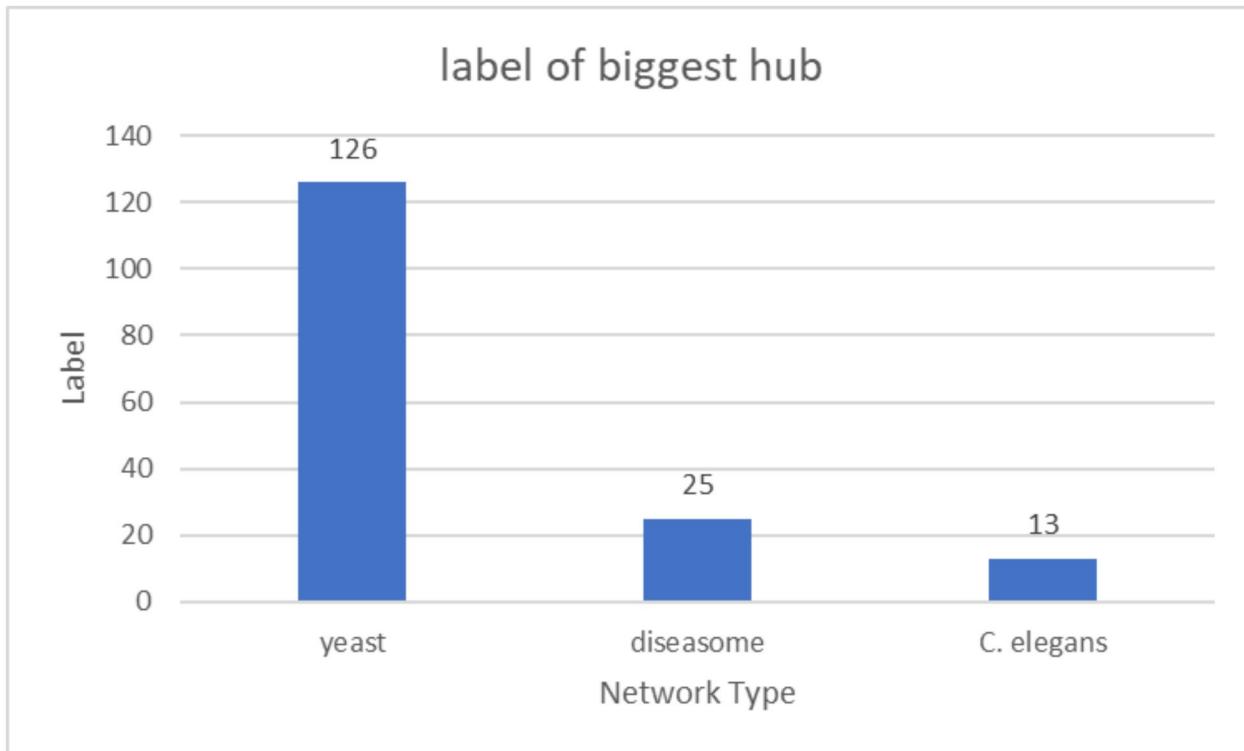
**Edges:** 2345

Directed Graph

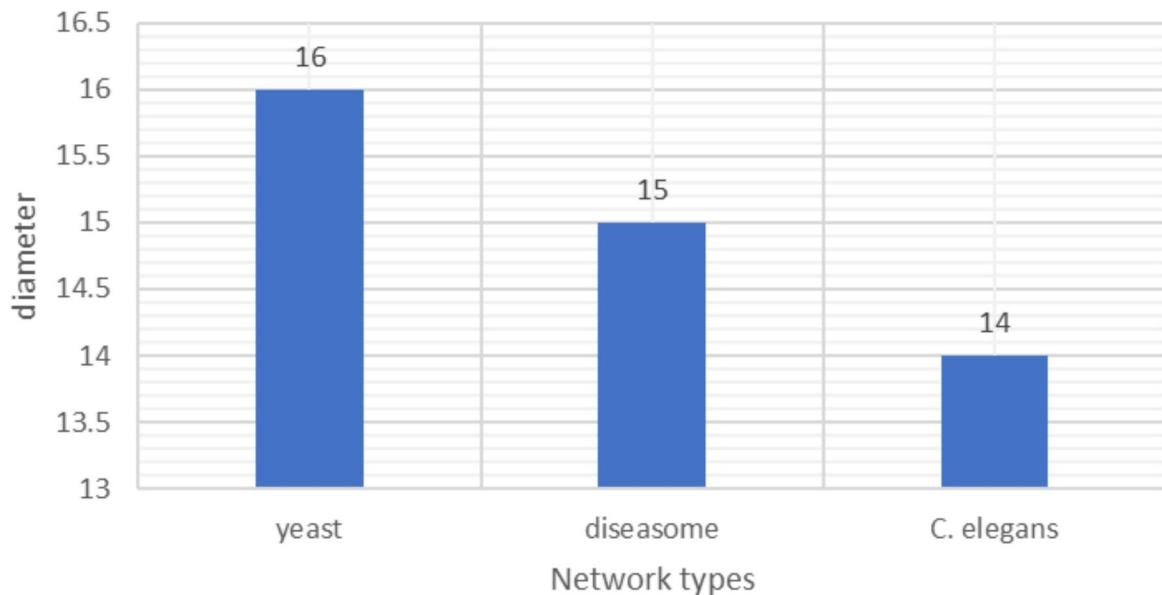
Exercise 2  
Comparison table

1		yeast	diseasome	C. elegans	
2	# of edges	7182	3926	2345	
3	label of biggest hub ( biggest modularity)	126	25	13	
4	ave degree	3.042	2.767	7.663	
5	shape of degree distribution	exponential decay	exponential decay	right-skewed unimodal (not an exponential decay)	
6	Network diameter	16	15	14	
7	Avg. Path length	4.648	6.648708193	3.9918839	
8	graph density	0.001	0.002	0.025	
9	number of modular community	126	25	13	
10	Modularity	0.591	0.876	0.371	
11	Clustering coefficient	0.096	0.414	0.164	
12	Mean connectivity (# edges/ # node)	3.041931385	2.766737139	7.663398693	
13	Wiring probability = Number of actual edges / Number of possible edges	0.002	0.004	0.05	
14					
15					

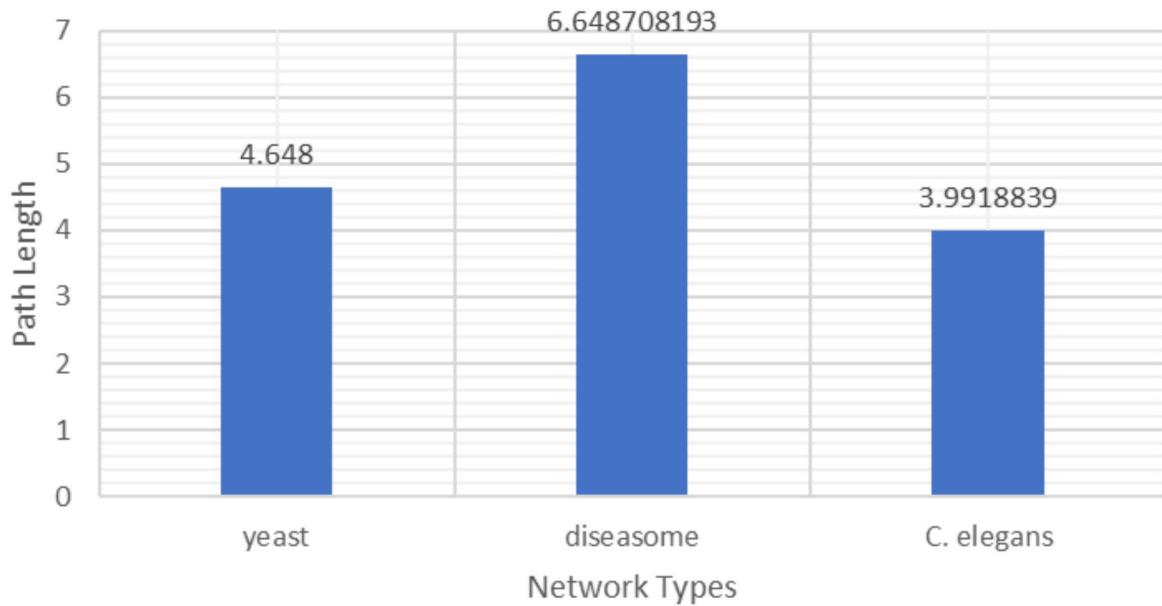


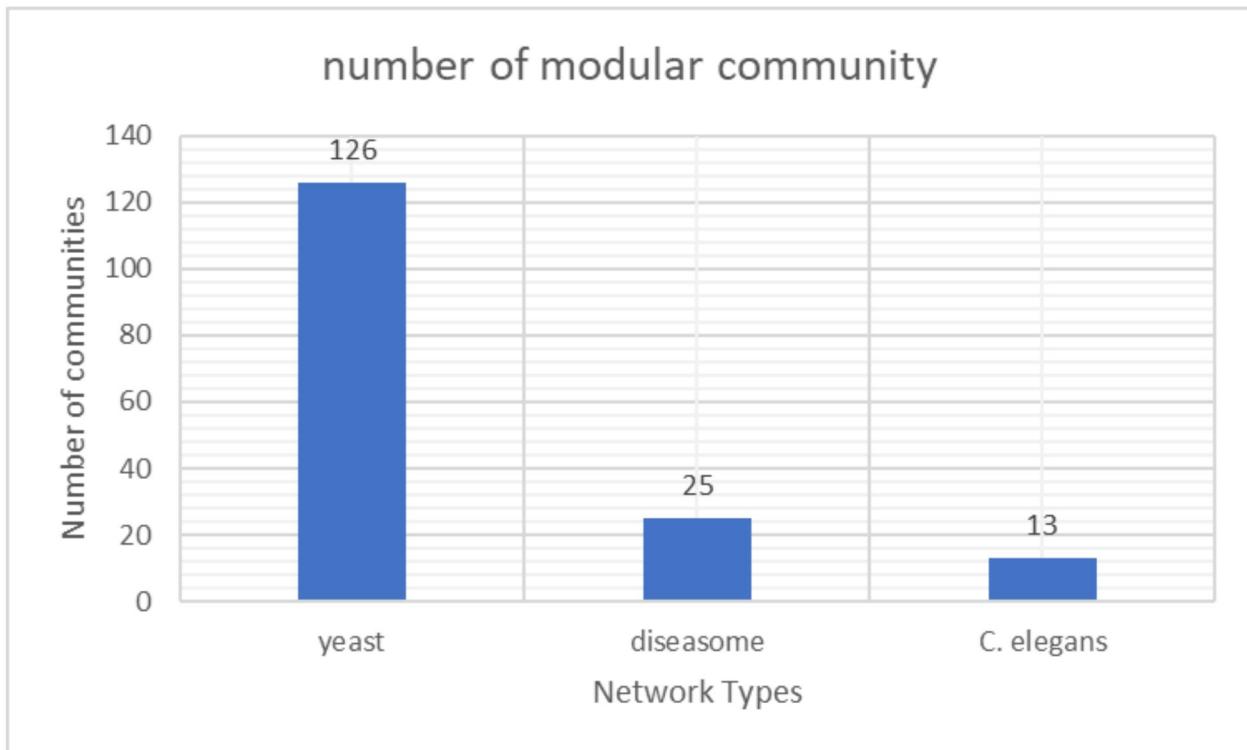
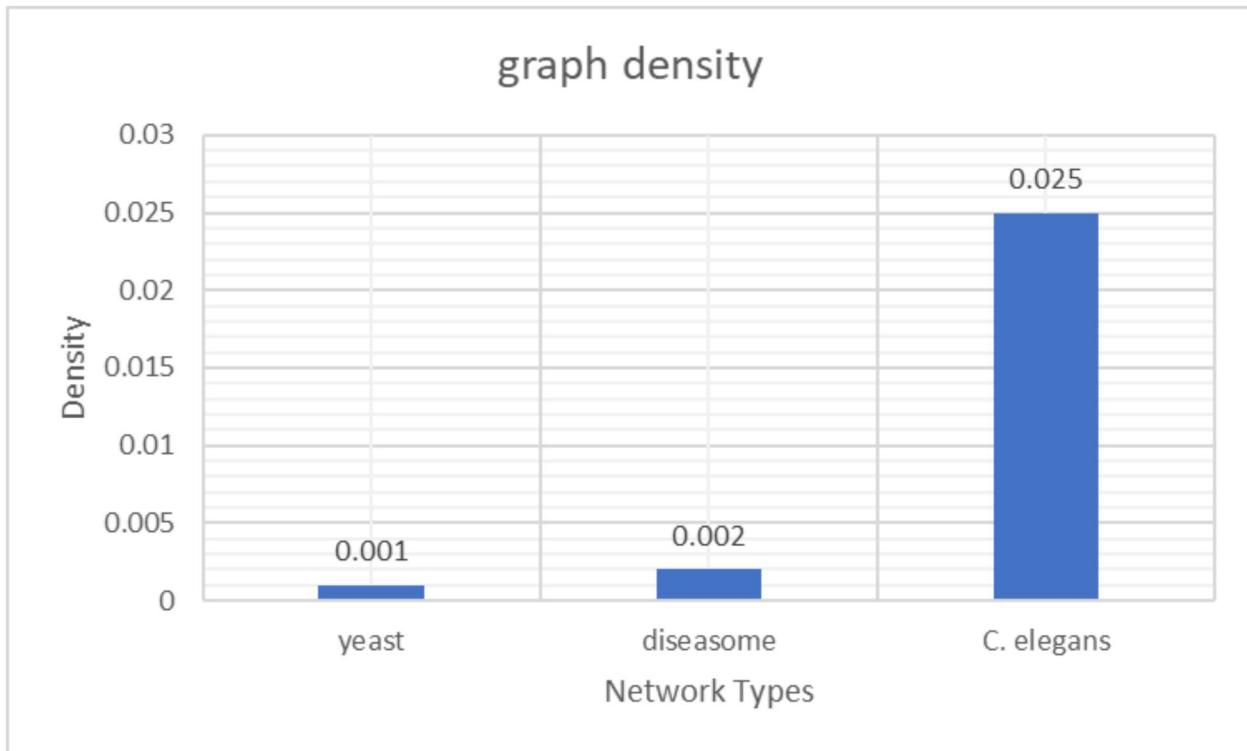


### Network diameter

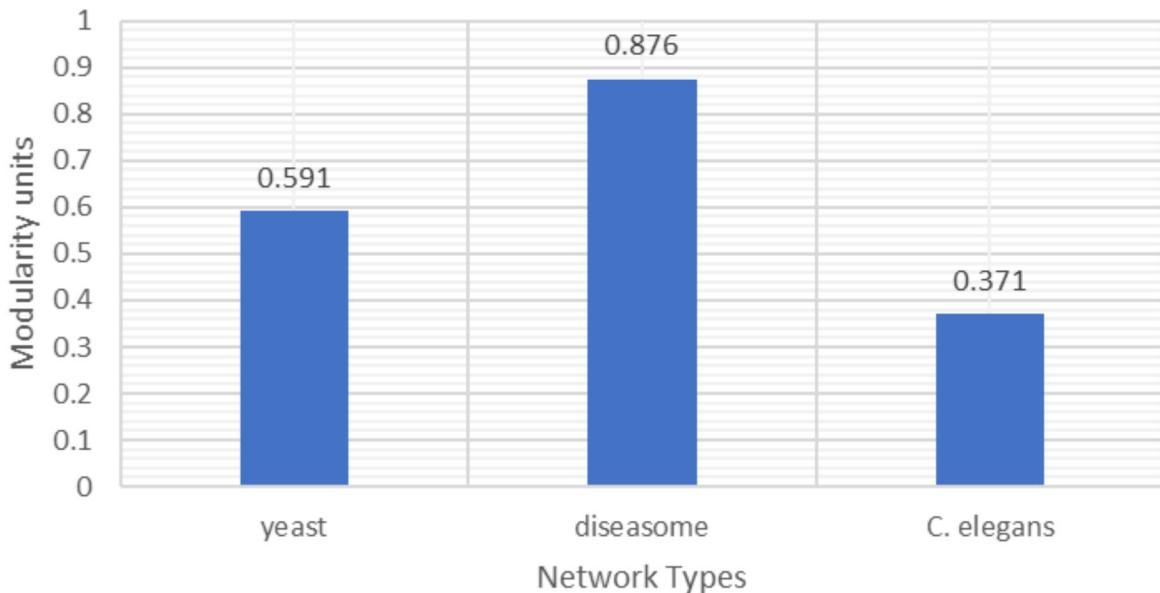


### Avg. Path length

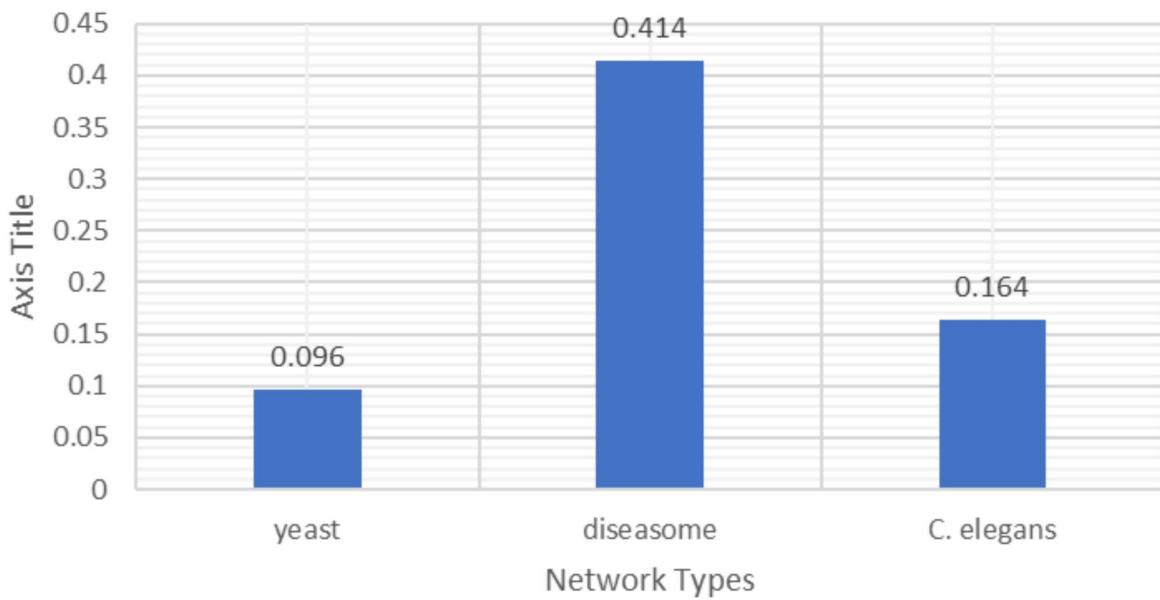




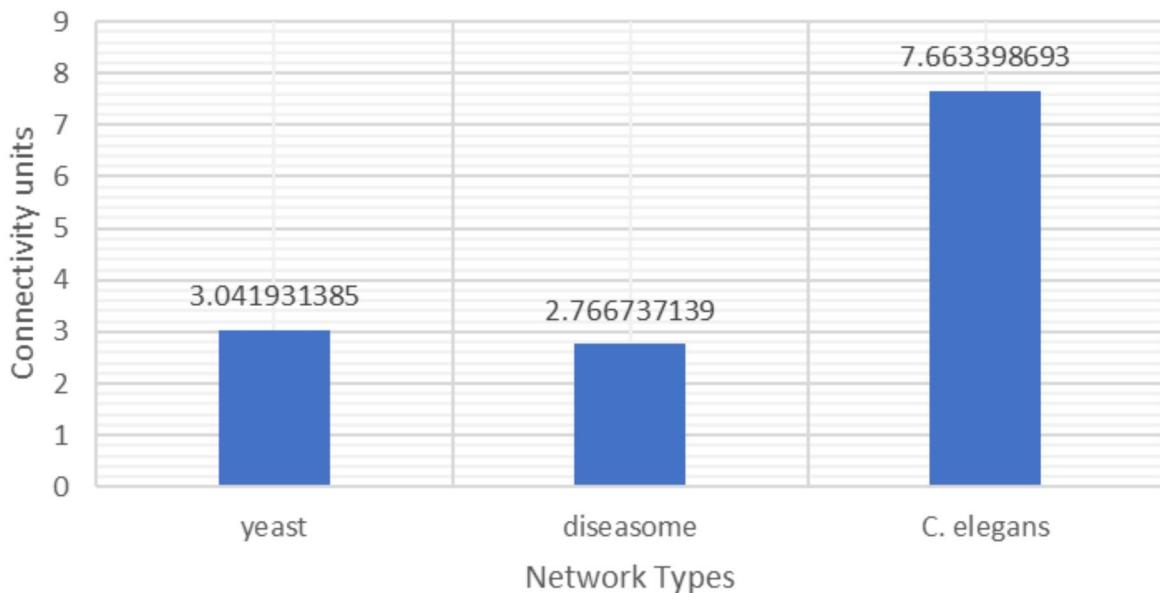
### Modularity



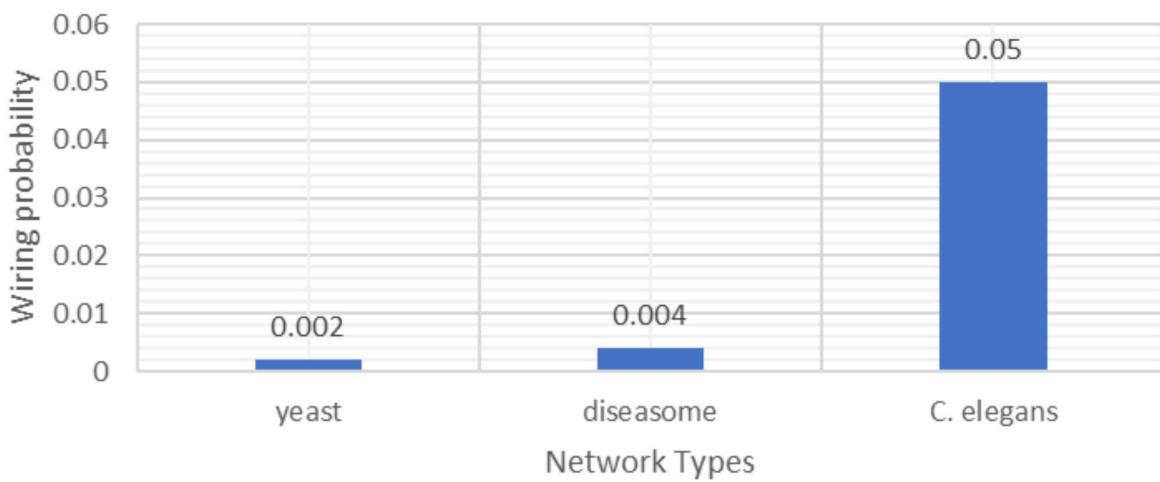
### Clustering coefficient



Mean connectivity (# edges/ # node)



Wiring probability = Number of actual edges/Number of possible edges =  $(2 * \text{Graph Density})$



We notice that the three network types have variability in all these parameters.

---

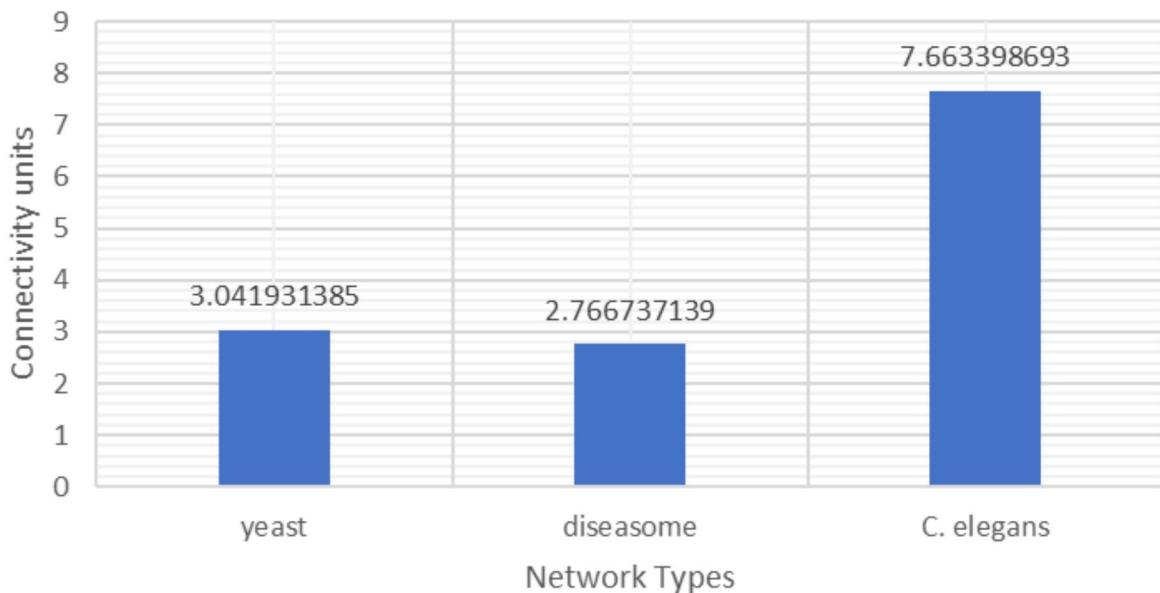
### Exercise 3

Random Yeast using Yifan

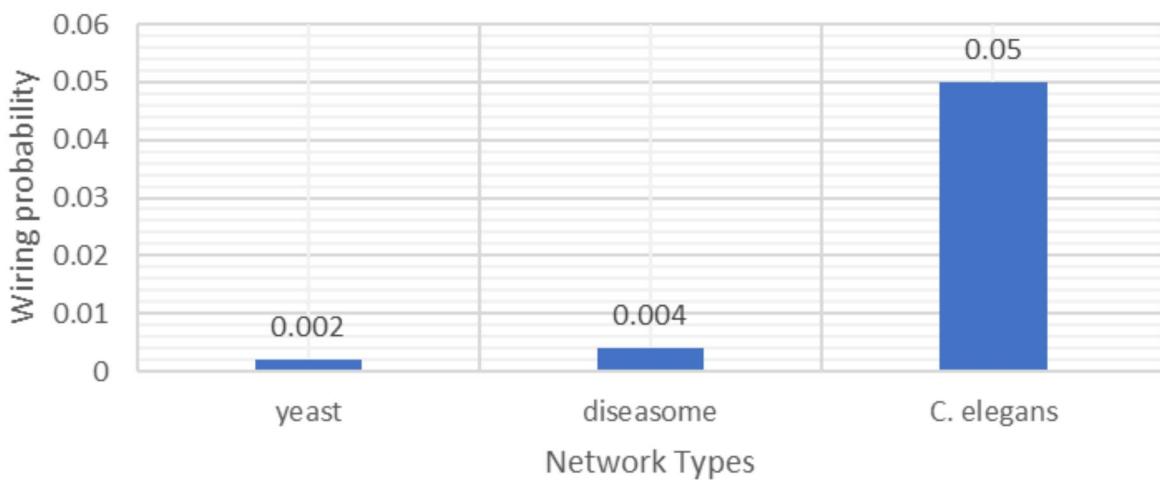
2 Ex.2 20 / 20

✓ - 0 pts Correct

Mean connectivity (# edges/ # node)



Wiring probability = Number of actual edges/Number of possible edges =  $(2 * \text{Graph Density})$

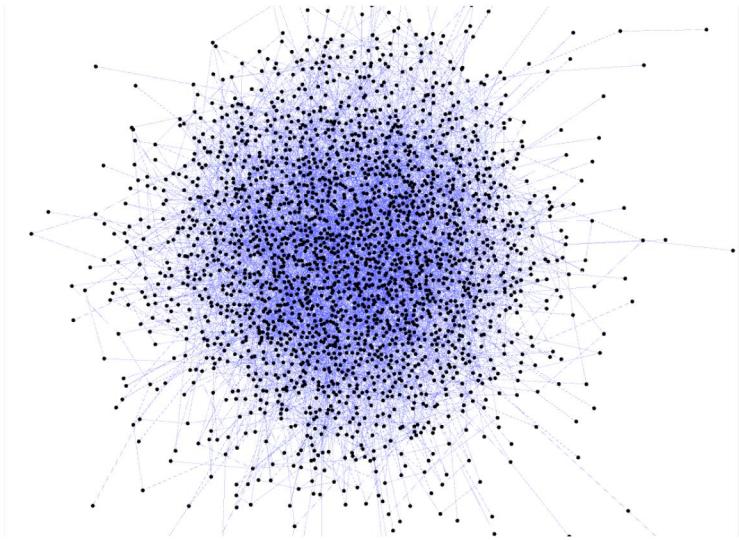


We notice that the three network types have variability in all these parameters.

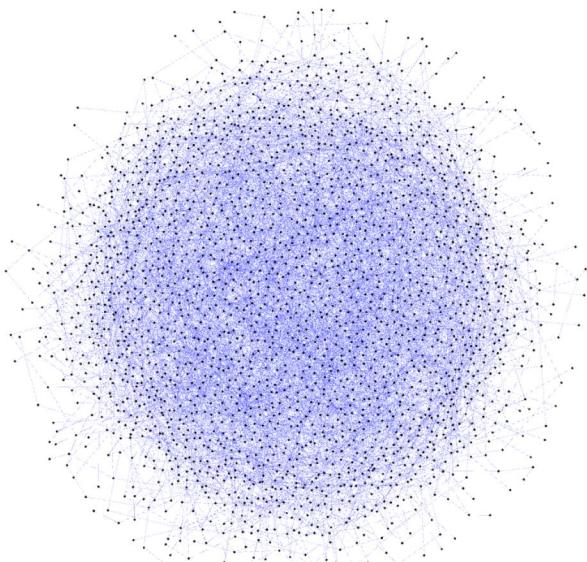
---

### Exercise 3

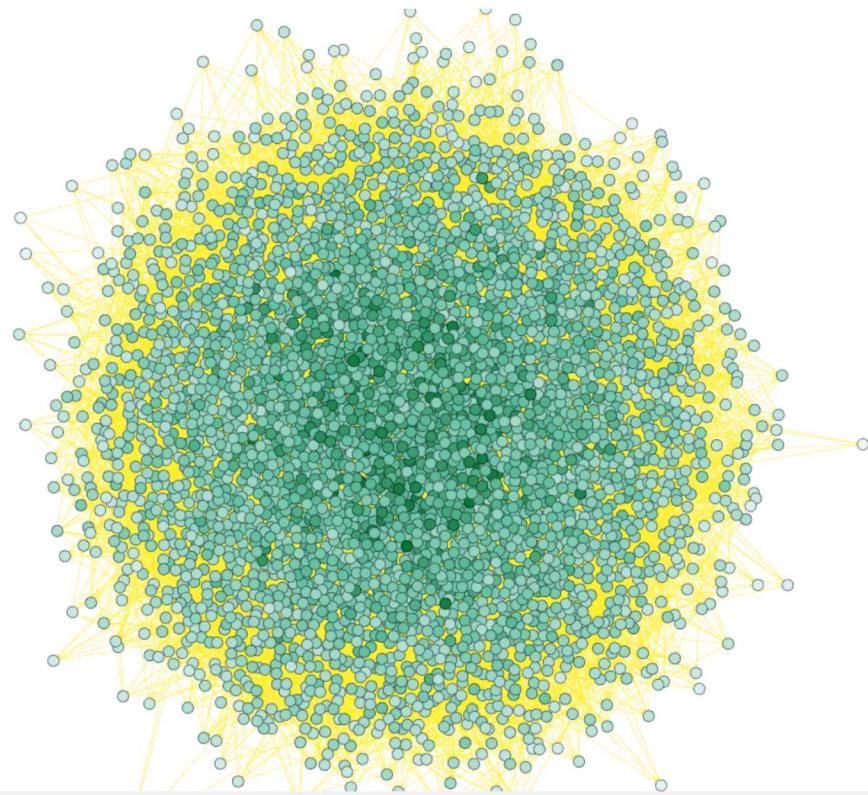
Random Yeast using Yifan



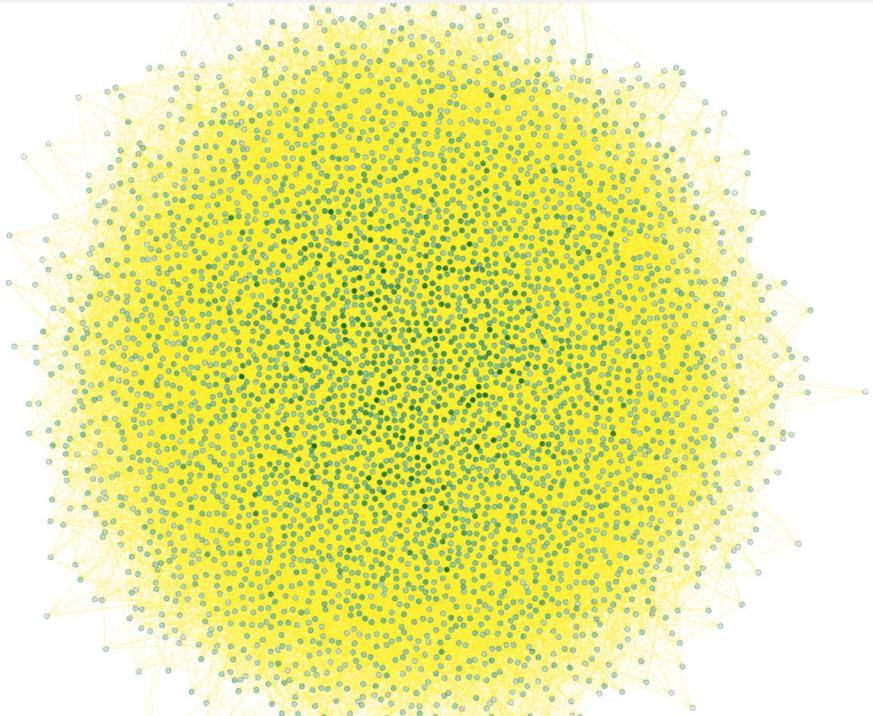
Random Yeast using FR



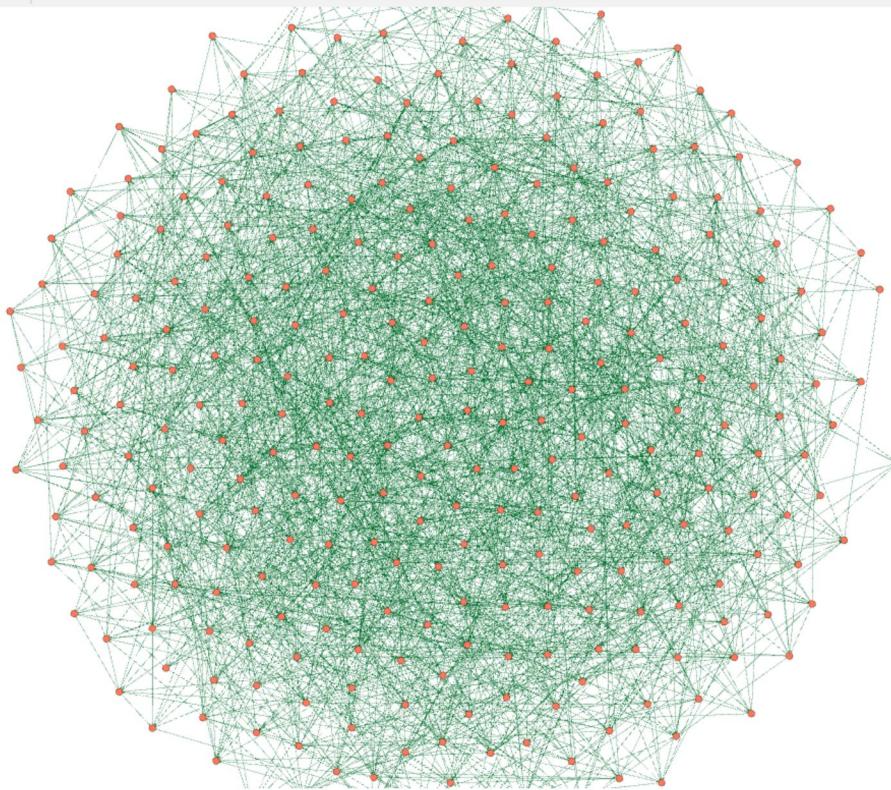
Random diseaseosome using Yifan Hu



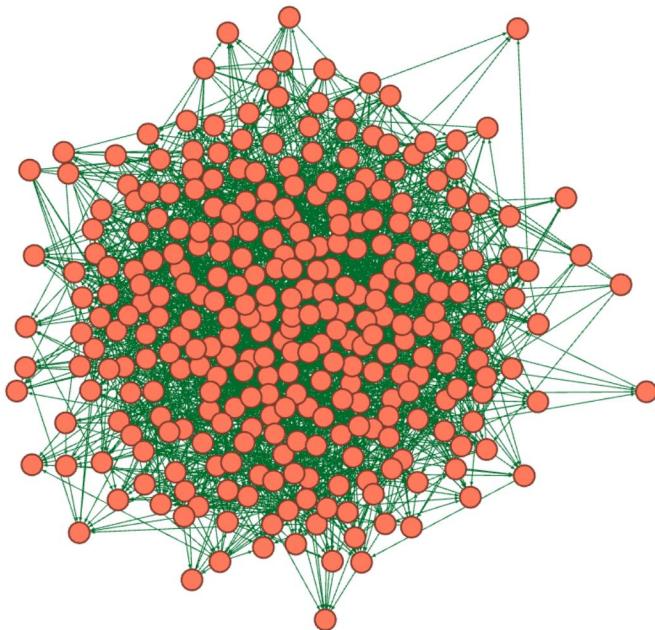
Random diseaseosome using FR



Random *C. elegans* using FR

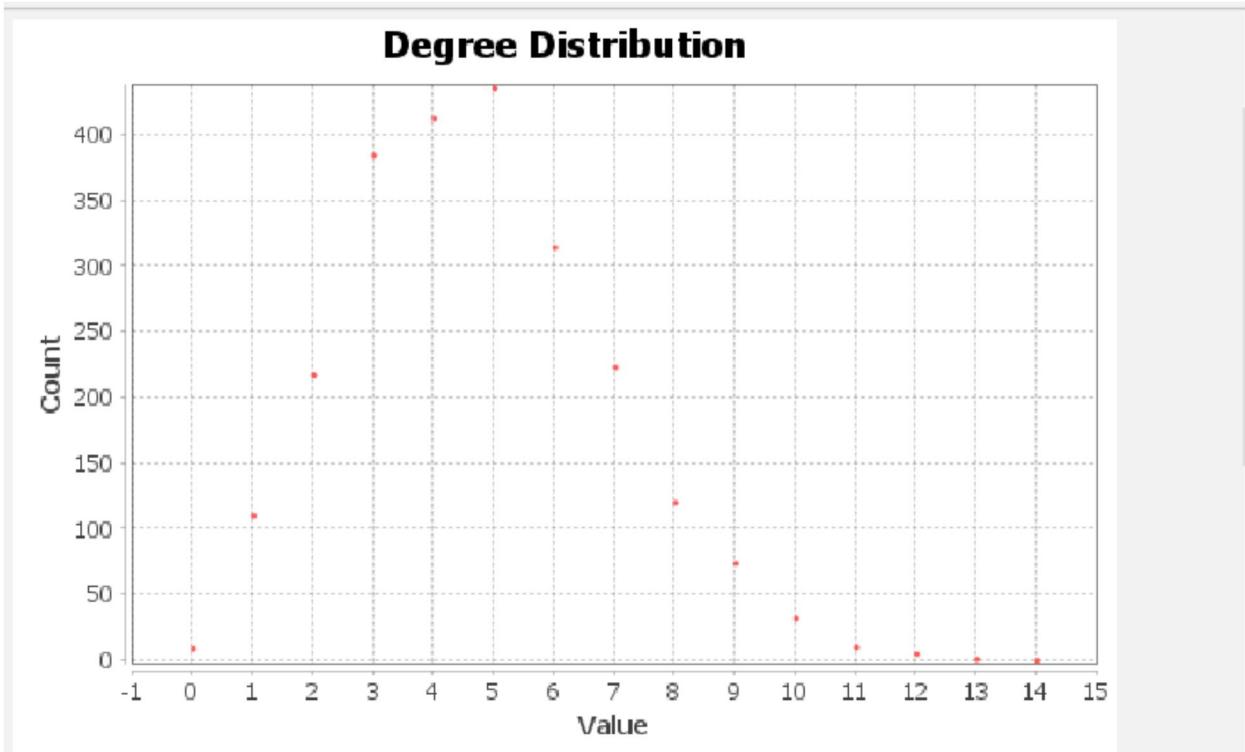


Random *C. elegans* using Yifan Hu



For the yeast figure, when I tried the above two methods I thought they didn't exactly resemble as much as expansion of the random figure to a scale of 1.2 did. Although keeping the edges

and nodes, the graph density the same, there are some parameters that are inherent to the graph of non-stochastic model. This changes the degree distribution so that the degree distribution is normal as opposed to the exponential decay distribution. It can not look exactly like that one, and the normal distribution graph of the degree of edges for a randomized Yeast system is shown below.

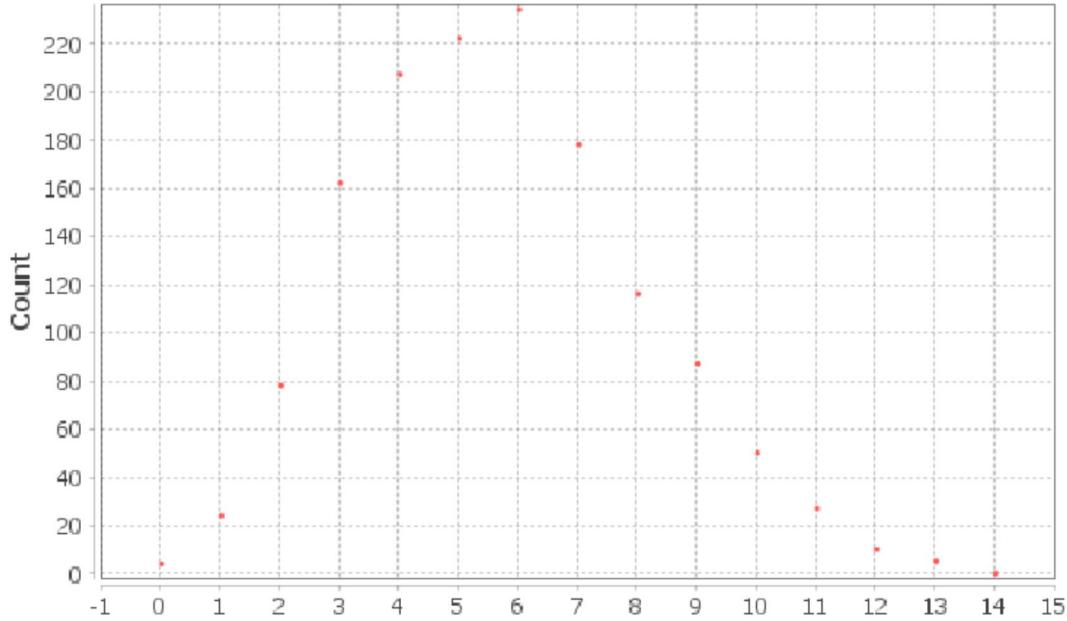


## Degree Report

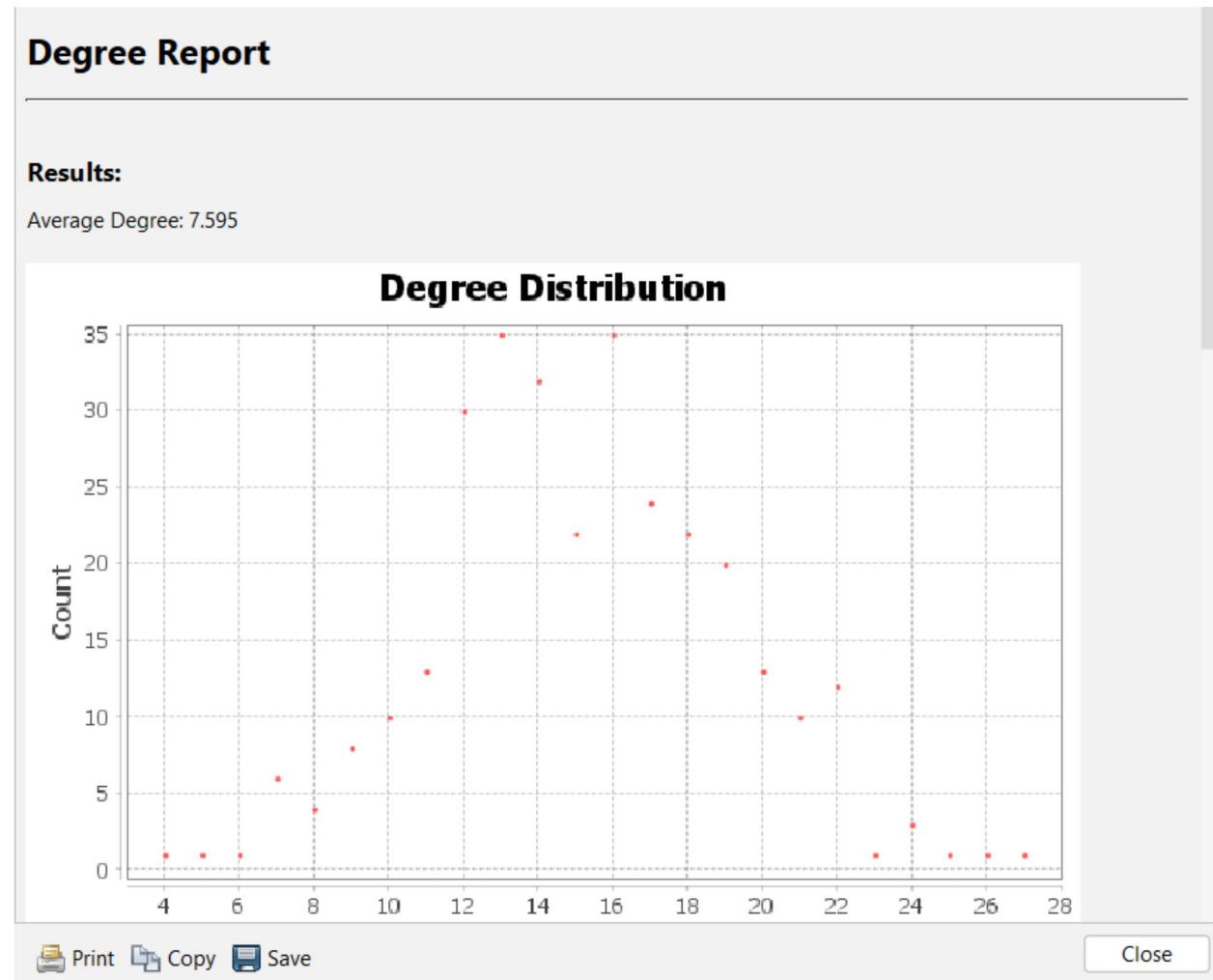
### Results:

Average Degree: 2.837

**Degree Distribution**



Below is the degree report for *C. elegans*, which is again normal.



So this is why even when expanded and manipulated, the random graph does not resemble the true network system.

---

Exercise 4

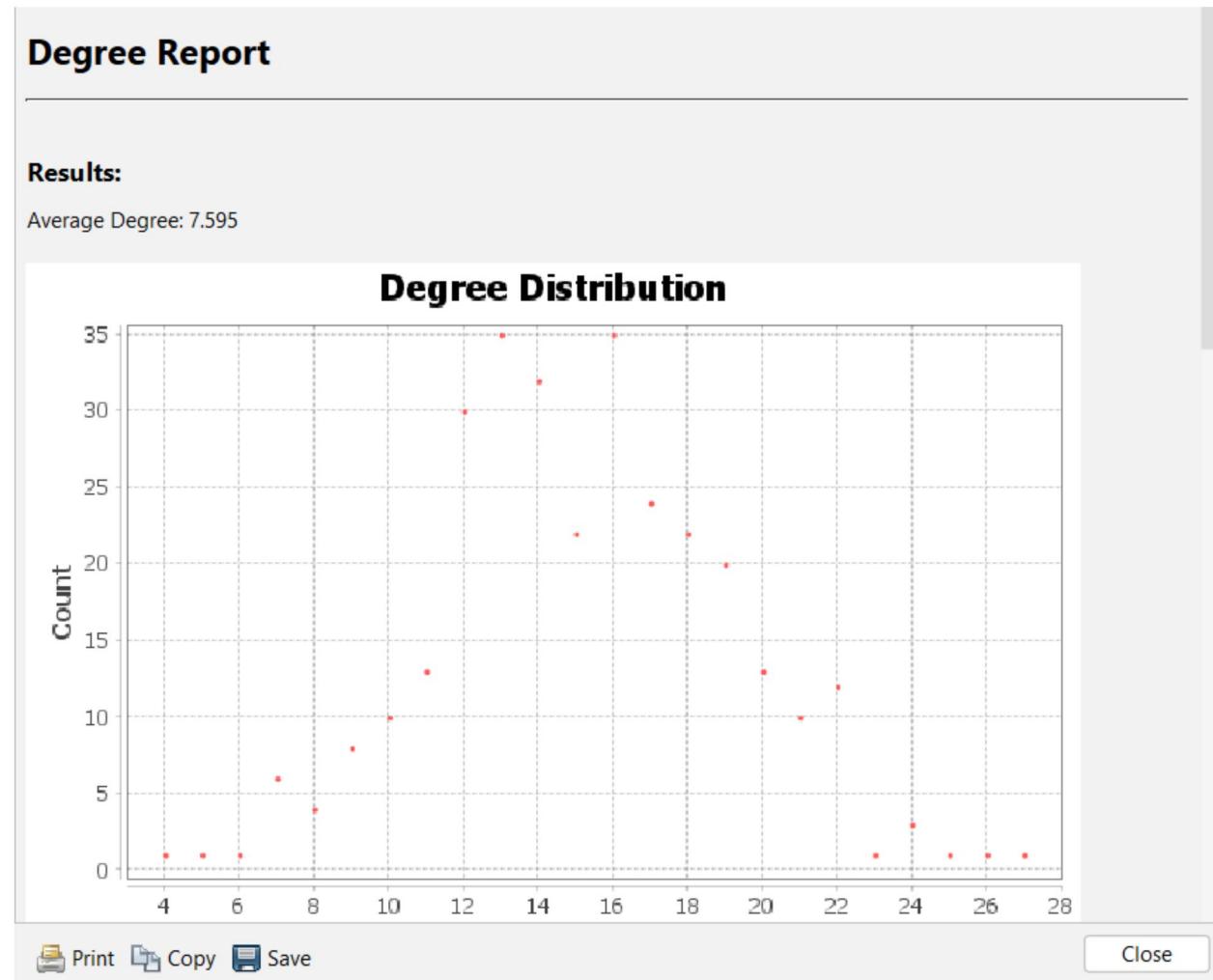
3 Ex.3 20 / 20

✓ - 0 pts Correct

- 4 pts Correctly argue that the random network doesn't look like the real one

- 2 pts Include pictures of the graphs

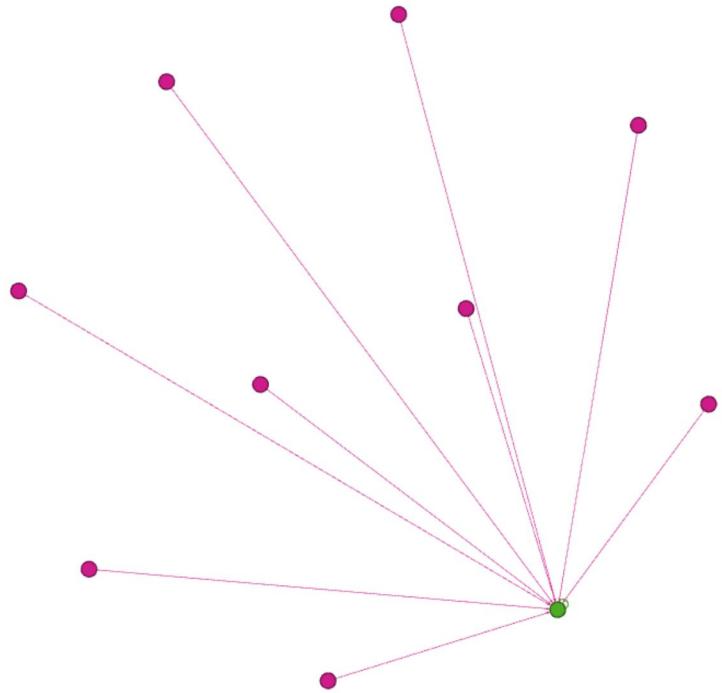
Below is the degree report for *C. elegans*, which is again normal.



So this is why even when expanded and manipulated, the random graph does not resemble the true network system.

---

Exercise 4



Id	Label	timeset
0	1	
1	2	
2	3	
3	4	

Source	Target	Type	Id	Label	timeset	Weight
1	6	Directed	71			1
2	6	Directed	17			1
3	6	Directed	145			1
4	6	Directed	239			1
5	6	Directed	453			1
6	6	Directed	710			1
7	6	Directed	1069			1
8	6	Directed	1142			1
9	6	Directed	1341			1
10	6	Directed	1491			1

## Exercise 5

Compare and contrast the values for each network with the biological details, knowledge, and intuition about each network. And give some comments on your observations.

The budding yeast network interaction contains mainly query genes and crossing mutations. When we evaluate its global dynamics, properties, and stability, we can see that it has the least graph density. This means that the yeast network is a sparse graph which every pair of vertices is connected by only a few edges. We can not say that the budding yeast network has the most associatedness because even though it has the most amount of # of edges, the mean connectivity is still lower than *C. elegans*. As an organism, then, we can conclude that yeast has genetic connectivity between each gene of *C. elegans*. A feature of budding yeast network is its clustering strength, with the largest label of biggest hub being 126, much greater than that of the diseasesome and *C. elegans*. The strength to be categorized under a community-like structural understanding of its genes.

Looking at the Diseasesome network, we can see that there are clearly some extremely large nodes being clustered. A property of a human disease network is that associatedness is derived from gene information and symptoms. Based the disease's classes, some diseases may connect to a large number of disease whereas most diseases have fewer links to others. Like the yeast network, the average degree will be low because of the disease's relatedness to other diseases, with few that is linked to many. Unlike *C. elegans*, which the worm's neural connectivity being high enough that it is commonly used as a subject in neuroscience laboratories. If relatedness is low, then the average path length will be higher, which is what is displayed to us.

4 Ex.4 20 / 20

✓ - 0 pts Correct

- 20 pts not attempted
- 2 pts Include node and edge files

Source	Target	Type	Id	Label	timeset	Weight
1	6	Directed	71			1
2	6	Directed	17			1
3	6	Directed	145			1
4	6	Directed	239			1
5	6	Directed	453			1
6	6	Directed	710			1
7	6	Directed	1069			1
8	6	Directed	1142			1
9	6	Directed	1341			1
10	6	Directed	1491			1

## Exercise 5

Compare and contrast the values for each network with the biological details, knowledge, and intuition about each network. And give some comments on your observations.

The budding yeast network interaction contains mainly query genes and crossing mutations. When we evaluate its global dynamics, properties, and stability, we can see that it has the least graph density. This means that the yeast network is a sparse graph which every pair of vertices is connected by only a few edges. We can not say that the budding yeast network has the most associatedness because even though it has the most amount of # of edges, the mean connectivity is still lower than *C. elegans*. As an organism, then, we can conclude that yeast has genetic connectivity between each gene of *C. elegans*. A feature of budding yeast network is its clustering strength, with the largest label of biggest hub being 126, much greater than that of the diseasesome and *C. elegans*. The strength to be categorized under a community-like structural understanding of its genes.

Looking at the Diseasesome network, we can see that there are clearly some extremely large nodes being clustered. A property of a human disease network is that associatedness is derived from gene information and symptoms. Based on the disease's classes, some diseases may connect to a large number of diseases whereas most diseases have fewer links to others. Like the yeast network, the average degree will be low because of the disease's relatedness to other diseases, with few that are linked to many. Unlike *C. elegans*, which the worm's neural connectivity being high enough that it is commonly used as a subject in neuroscience laboratories. If relatedness is low, then the average path length will be higher, which is what is displayed to us.

Thirdly, the *C. elegans* network is important because so far it is the only organism that has a complete connectome. Its genome encodes for a multitude of proteins that contribute to neuronal activity and communication. When I first looked at the *C. elegans* graph, there were a lot of bolded arrows that point to a single enlarged node. This organism has the largest mean connectivity and least average path length due to its feature. It means that for each node, there are on average about 7.6 edges creating a walk. For each gene, there are on average about 7.6 associated genes that contributes to the connectome.

#### Exercise 6

I looked at the nonweighted social network on Les Mis characters. Hugo's work is to accentuate the religious, political, and social tumultuous nature of the French revolution. However, it traces the redemption of Jean Valjean and so must conform to his social network. The multitude of characters such as Marius, Cosette, and the innkeepers draw out how intricately Hugo crafted this novel. I found that the average degree was 3.299 and the average weighted degree was much higher, 10.649. This means unlike other graphs where the weight was the same = 1, this network has weights. Also, not every character is in the peripheral of Jean Valjean.

5 Ex.5 20 / 20

✓ - 0 pts Correct

- 20 pts Not attempted

Thirdly, the *C. elegans* network is important because so far it is the only organism that has a complete connectome. Its genome encodes for a multitude of proteins that contribute to neuronal activity and communication. When I first looked at the *C. elegans* graph, there were a lot of bolded arrows that point to a single enlarged node. This organism has the largest mean connectivity and least average path length due to its feature. It means that for each node, there are on average about 7.6 edges creating a walk. For each gene, there are on average about 7.6 associated genes that contributes to the connectome.

#### Exercise 6

I looked at the nonweighted social network on Les Mis characters. Hugo's work is to accentuate the religious, political, and social tumultuous nature of the French revolution. However, it traces the redemption of Jean Valjean and so must conform to his social network. The multitude of characters such as Marius, Cosette, and the innkeepers draw out how intricately Hugo crafted this novel. I found that the average degree was 3.299 and the average weighted degree was much higher, 10.649. This means unlike other graphs where the weight was the same = 1, this network has weights. Also, not every character is in the peripheral of Jean Valjean.

6 Ex.6 (optional) 5 / 5

✓ - 0 pts Correct

- 5 pts Click here to replace this description.