HW1 – Generating and identifying random/scale free networks

1. IV. **Type a:**

num_networks = 20, n = 100, p = 0.1
{'degrees_avg': 9.914999999999997, 'degrees_std': 2.95801744241268, 'degrees_min': 3.75, 'degrees_max': 18.0, 'spl': 2.236616161616162, 'diameter': 3.95}

Type b:

num_networks = 20, n = 100, p = 0.6
{'degrees_avg': 59.27000000000002, 'degrees_std': 4.6736997989819375, 'degrees_min': 47.15, 'degrees_max': 71.55, 'spl': 1.4013131313131315, 'diameter': 2.0}

Type c:

num_networks = 10, n = 1000, p = 0.1
{'degrees_avg': 99.86280000000001, 'degrees_std': 9.455184791243514, 'degrees_min': 68.6, 'degrees_max': 131.5, 'spl': 1.9000804804804805, 'diameter': 3.0}

Type d:

num_networks = 10, n = 1000, p = 0.6
{'degrees_avg': 599.2798, 'degrees_std': 15.433643427419153, 'degrees_min': 550.0, 'degrees_max': 648.1, 'spl': 1.4001203203203203, 'diameter': 2.0}

V. **Yes, it makes sense.**

The parameter 'p' represents the probability that any two nodes are connected. Therefore, the higher the 'p', the greater the number of connections between nodes. Our statistics reveal that in Type B, the metrics for average degree ('degrees_avg'), standard deviation of degrees ('degrees_std'), minimum degree ('degrees_min'), and maximum degree ('degrees_max') are all higher than in Type A, attributable to the higher 'p' value in Type B. Additionally, in Types C and D, the number of nodes ('n') is greater, leading to a larger number of edges compared to Types A and B, where the number of nodes is fewer. We also observe variations in these first four statistics, as the 'p' value is lower in Type C then Type D.

Regarding the last two statistics ('spl' and 'diameter'), we observe the opposite trend: the larger the 'p', the smaller these measures tend to be.

This is because the denser the network (more edges), the more routing options there are between vertices. Consequently, the likelihood of shorter paths between any two vertices increases, leading to a decrease in the graph's diameter (shorter routes) and a smaller average shortest path length ('spl') for the same reason."

2. **IV.** First, I pick a network from the list of networks that I have found an optimal theoretical_p:
I run the function 'most_probable_p' on the list, and this is the result:

```
[0.6, 0.1, 0.3, -1, 0.6, 0.3, 0.01, 0.6, 0.01, 0.1]
```

We can see that only the fourth network we don't find optimal theoretical_p.

I pick the first network.
Now I run the function 'rand_net_hypothesis_testing' for different p, this is the results:

```
for p: 0.65 -> (6.101136547689183e-165, 'reject')
for p: 0.55 -> (5.9706581266122484e-133, 'reject')
for p: 0.7 -> (0.0, 'reject')
for p: 0.5 -> (0.0, 'reject')
```

We see that as we change the p we found (up, down) we will probably reject the null hypothesis.
That is, the p-value we found is clearly optimal.

Now we do this for a much bigger network (1000 nodes), this is the result:
The theoretical_p for this network is 0.1

```
for p: 0.11 -> (9.4404455808661e-118, 'reject')
for p: 0.09 -> (1.3489986760305815e-128, 'reject')
for p: 0.2 -> (0.0, 'reject')
for p: 0.01 -> (0.0, 'reject')
```

We can see that we get the same result for a much bigger network.

V. The code:

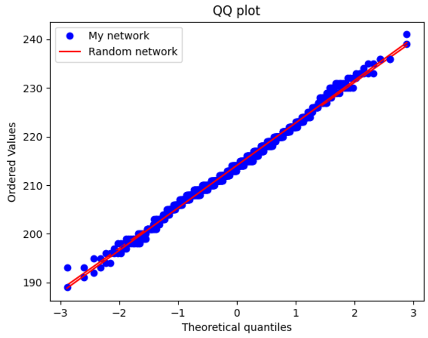
```python
# Load your network list first
my_network = list_networks[0]  # choose the first network

# Generate a random network with the same number of nodes and edges
random_network = nx.gnm_random_graph(my_network.number_of_nodes(), my_network.number_of_edges())

# Calculate the degree distribution for the given network and the random network
given_degrees = list(dict(my_network.degree()).values())
random_degrees = list(dict(random_network.degree()).values())

# Plot the QQ plot
fig, ax = plt.subplots()
stats.probplot(given_degrees, dist="norm", plot=ax)  # Plot for the given network
stats.probplot(random_degrees, dist="norm", plot=ax)  # Plot for the random network

ax.legend(["My network", "Random network"])
ax.set_title("QQ plot")
plt.show()
```



We can see that the network I selected from the pickle file overlaps in the graph with the random network I created.
Therefore, the network we choose does represent a random network.

3. III. The optimal γ of the networks in the lists:

```
network 1 has gamma 2.0354421215471694
network 2 has gamma 2.243689056926761
network 3 has gamma 2.4990852733023576
network 4 has gamma 2.10555389793955
network 5 has gamma 2.148436245556767
network 6 has gamma 2.5757272899518213
network 7 has gamma 2.0979556091503966
network 8 has gamma 2.368179965882712
network 9 has gamma 2.2181074245682737
network 10 has gamma 2.7094221254910655
```

IV. The statistics of the first scale-free network:

{'degrees_avg': 8.225087924970692, 'degrees_std': 24.200820210258257, 'degrees_min': 1, 'degrees_max': 359, 'spl': 2.6766920396942027, 'diameter': 6}

I look for a random network from the list that we load from the first pickle and find that network number 7 is the best (they have almost the same number of nodes and edges.)

The statistics of the random network:

{'degrees_avg': 8.144578313253012, 'degrees_std': 2.7613185697899896, 'degrees_min': 2, 'degrees_max': 18, 'spl': 3.44067900068307, 'diameter': 6}

Yes, it makes sense.
We choose the two networks that are relatively similar in terms of the number of nodes and edges, so we can see that the average degrees in both networks are the same and this is due to the fact that it depends on the number of vertices and arcs.
Also, the types of networks are different, where a scale-free network is characterized by high variability between the nodes, unlike a random network where there is low variability, so it can be seen that the standard deviation is much larger in the scale-free network.
we can also notice that the maximum degree is much higher in a scale-free network and this is because it is characterized by very large hubs.
And the last one thet the diameter in both networks is equal and this is due to the small-world phenomenon.

4. **III.** **This is the classification for each of the networks:**

```
network 1 is a random network
network 2 is a random network
network 3 is a scale-free network
network 4 is a scale-free network
network 5 is a random network
network 6 is a random network
network 7 is a scale-free network
network 8 is a random network
network 9 is a random network
network 10 is a scale-free network
network 11 is a random network
network 12 is a scale-free network
network 13 is a random network
network 14 is a random network
network 15 is a random network
network 16 is a scale-free network
network 17 is a random network
network 18 is a random network
network 19 is a random network
network 20 is a random network
```