Social Network - Home Work 2

Hamid Nemati – 810100492

Assume a Watts-Strogatz with N nodes and <k> = 2c.

1. **For $\boldsymbol{\beta = 0}$ prove clustering coefficient is:**

$$\boldsymbol{C([\beta = 0]) = \frac{3(c-1)}{2(2c-1)}}$$

   a. **Method 1:**

- Watts and Strogatz proposed this clustering coefficient:

$$C_{ws} = \frac{1}{n}\sum_{i=1}^{n} C_i = \frac{1}{n}\sum_{i=1}^{n} \frac{number\ of\ connected\ neighbor\ pairs}{\frac{1}{2}k_i(k_i - 1)}$$

- This method is heavily biased in favor of low degree nodes. The better way is expressed by Newman at el. (2001):

$$C = \frac{3\ \times (number\ of\ triangles\ on\ a\ graph)}{(number\ of\ connected\ triples)}$$

- Triple is any node that is connected to a pair of nodes which may be connected to each other or not. Ina triangle they must be connected.

- Now if a lattice has a linear dimension of L, then the number of triangles on is $\frac{1}{2}Lc(c-1)$ and the number of connected triples are $Lc(2c-1)$. Hence the clustering coefficient is: (c is maximum range of a node)

$$C = \frac{3\ \times \left(\frac{1}{2}Lc(c-1)\right)}{(Lc(2c-1))} = \frac{3(c-1)}{2(2c-1)}$$

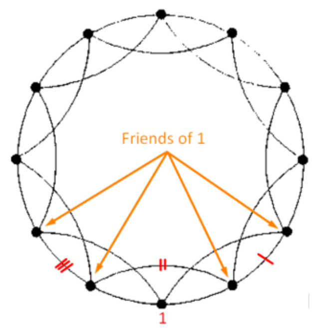
b. Method 2: (Mathematical Induction)

- We calculate it for a small $n = 12$ and $k = 2c = 4$ first:
    i. We calculate C_i for $i = 1$ and because it's a regular lattice the pattern is the same and the average would be the same. The formula for C_i is: (L_i is the number of actual links between one's friends)

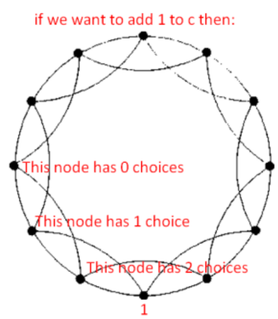
$$C_i = \frac{L_i}{\frac{k_i(k_i-1)}{2}}$$

    ii. k_i is 4 so the Denominator is $4 \times 3 \div 2 = 6$. The Numerator is 3. So:

$$C_i = \frac{1}{2} = \frac{3 \times (c_i - 1)}{2 \times (2c_i - 1)} = \frac{3 \times 1}{2 \times 3}$$



- We assume c=j is right. Is c=j+1 right too? But first we need a way to describe L_i with c. Well we can see we have c node on left side and each one of them has a range and we should find out how many nodes that range can take in; the formula for that can be written this way:



$$0 + 1 + 2 + \cdots + c = \frac{c \times (c+1)}{2}$$

- For $c + 1$ which means $k + 2$ we have:

$$C_i = \frac{L_i}{\frac{k_i(k_i-1)}{2}} = \frac{\frac{(c+1)(c+2)}{2}}{\frac{(2c+2)(2c+1)}{2}} = \frac{(c+1)(c+2)}{2(c+1)(2c+1)} = \frac{c+2}{2(2c+1)}$$

- While from the previous formula we have:

$$C = \frac{3(c_{new} - 1)}{2(2c_{new} - 1)} = \frac{3((c+1) - 1)}{2(2(c+1) - 1)} = \frac{c+2}{2(2c+1)}$$

- Both get the same value so the induction is complete. And formula is proved.

2. For $\boldsymbol{\beta = 0}$ prove the below formula:

$$\boldsymbol{< d[\beta = 0] > = \frac{N}{4c}}$$

a. Let's draw the distance table: (in each step we can jump c node to the right or to the left)

2c node at distance	1
2c node at distance	2
…	…
2c node at distance	$\frac{N}{2 \times c}$

b. Now Let's sum this distances

$$2 \times c \times 1 + \cdots + 2 \times c \times \frac{N}{2 \times c}$$
$$= 2 \times c \times \left(1 + 2 + \cdots + \frac{N}{2 \times c}\right)$$
$$= 2 \times c \times \left(\frac{\frac{N}{2 \times c} \times \left(\frac{N}{2 \times c} + 1\right)}{2}\right)$$
$$= 2 \times c \times \frac{\left(\frac{N^2}{2 \times c^2}\right) \times \left(\frac{N}{c}\right)}{4} = \frac{\left(\frac{N^2}{2 \times c}\right) \times (N)}{2}$$

c. Let's take an average of this, we have N nodes, so:

$$\frac{1}{N} \times \frac{\left(\frac{N^2}{2 \times c}\right) \times (N)}{2} = \frac{\left(\frac{N}{2 \times c}\right) \times (1)}{2} = \frac{N}{4 \times c}$$

d. And that's exactly what we wanted to prove.

e. We make a conclusion of this that will be used for Question 3: <d> is relative to the longest distance in the graph.

3. **Average distance after one rewire in the ring lattice?**
   a. As we can see in the result of the part 2: <d> is equal to the longest distance in the graph.
   b. In order to calculate <d> after one rewire we have to calculate the expected number for the node we connect to. We take distance as the value of each node and because lattice is regular we set a source (Node_1) as our base and find out the average number Node_1 will connect to.
   c. Size of each neighborhood is $< k >$ so we have $\frac{N}{<k>}$ neighborhood. But let's just for the sake of simplicity assume neighborhood size as 1.
   d. $N - 1$ node is left and the probability of being chosen is divided between them.

$$E(X) = \sum_{i=1}^{N} P(X_i) \times X_i = \frac{1}{N-1} \times \sum_{i=1}^{N} X_i$$

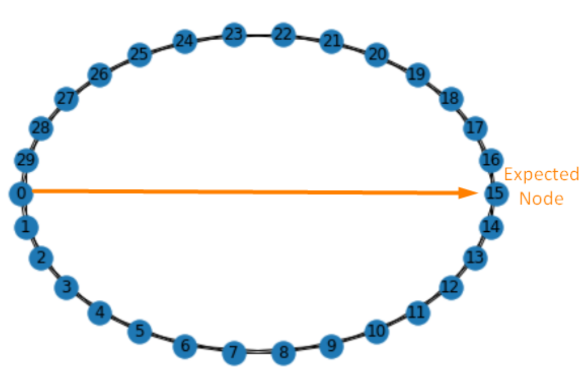
   e. Let's calculate the second part. It's the same as Question 2. And let's do a simulation too:

```python
x=0
for i in range(10000):
  x+=random.randint(1, 1000)
print(x/10000)
```

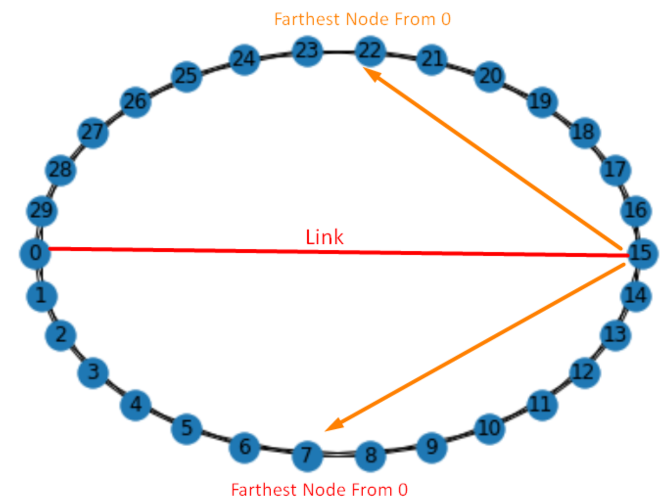
```
497.5582
```

$$E(X) = \frac{N}{2 < k >}$$

   f. $\frac{N}{2<k>}$ means the distance form Base node (Node_1), you can see them in the picture below:



g. Let's show the result after one rewire.



h. Remember this: <d> is equal to the longest distance in the graph.

i. So If the base node is on 0 degree angel then:

- Expected node is on 180 degree angle
- Farthest node is on 90 and 240 degree angle

j. The distance of Farthest node in the graph from zero now is:

$$\frac{N}{4 < k >}$$

k. Let's calculate its $< d >$:

4c node at distance	1
4c node at distance	2
…	…
4c node at distance	$\frac{N}{4 \times c}$

$$sum = \ 4 \times c \times \left(1 + 2 + \cdots + \frac{N}{4 \times c}\right)$$

$$= 4 \times c \times \left(\frac{\frac{N}{4 \times c} \times \left(\frac{N}{4 \times c} + 1\right)}{2}\right)$$

$$= c \times \frac{\left(\frac{N^2}{4 \times c^2}\right) \times \left(\frac{N}{c}\right)}{2} = \frac{\left(\frac{N^2}{4 \times c}\right) \times (N)}{2}$$

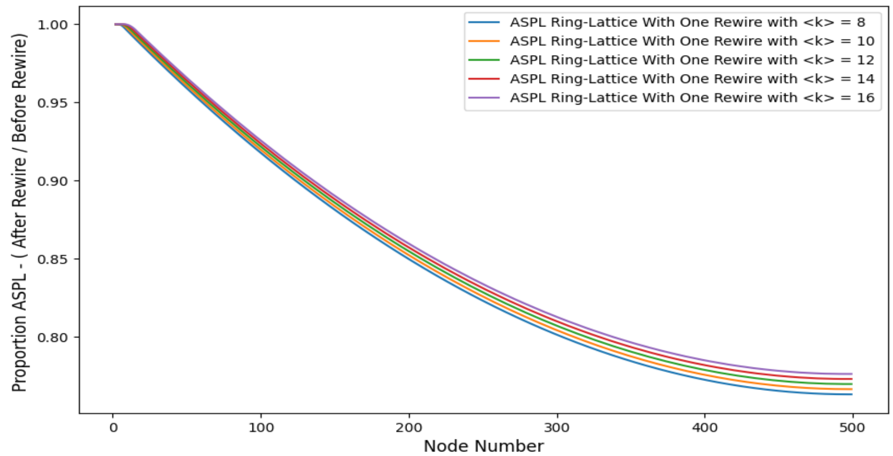
l. Let's take an average of this, we have N nodes, so:

$$\frac{1}{N} \times \frac{\left(\frac{N^2}{4 \times c}\right) \times (N)}{2} = \frac{\left(\frac{N}{4 \times c}\right) \times (1)}{2} = \frac{N}{8 \times c}$$

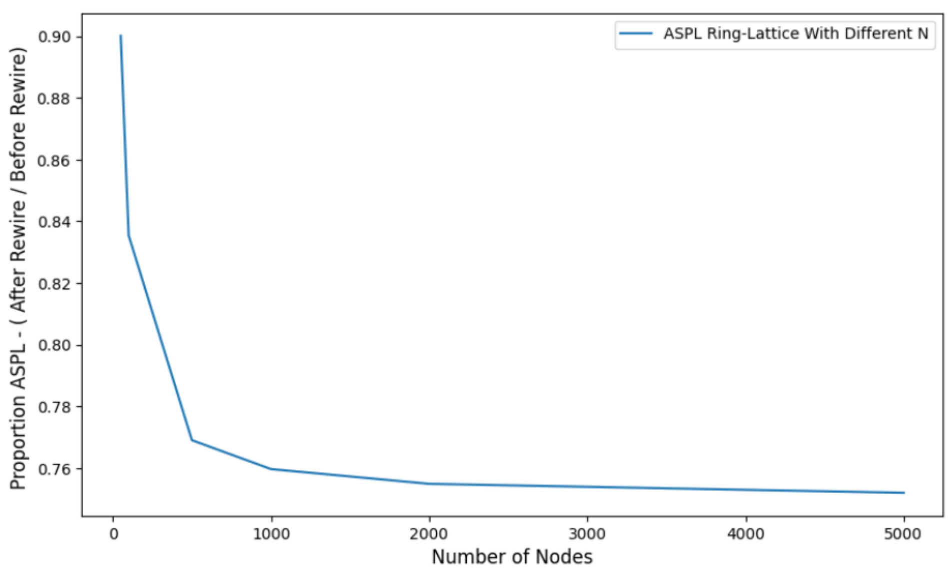
m. So if the $< d > = \frac{N}{4<k>}$ then let's see how much it decreased:

$$\frac{After\ rewire}{Before\ rewire} = \frac{\frac{N}{8 < k >}}{\frac{N}{2 < k >}} = \frac{1}{4}$$

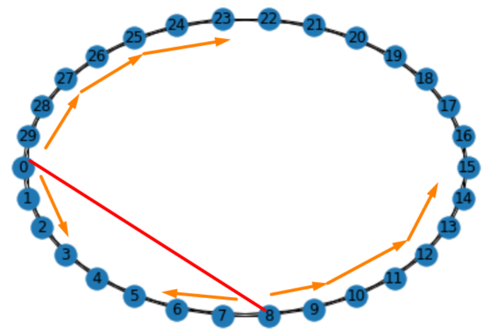
n. But the simulation below is showing something else: (you can find the code in the link below. Please see the code to verify for yourself.) https://colab.research.google.com/drive/17djMO0WIndzeav_QbXadA9sdOMKAeGj?usp=sharing



o. Now if we rewire exactly and without any randomness to node at distance $\frac{N}{2<k>}$ we get the line plot below:



p. But if we assume we connect to a node at distance $= \frac{N}{4<k>}$ like the picture below, we get $< d > = = \frac{3N}{8<k>}$ and then:



$$\frac{After\ rewire}{Before\ rewire} = \frac{\frac{3N}{8<k>}}{\frac{N}{2<k>}} = \frac{3}{4}$$

q. And this is what correlate with the simulation.
r. Maybe I calculated a wrong expected value but that only thing I could find in my short time.