# Exercise two: Network formation models

## Question 6 (3 points):

1. Develop three networks with the same number of vertices (n), but different probability (p); Name them as ER1, ER2, and ER3. Develop the plots of ER1, ER2 and ER3, describe how these three graphs look differently as p increase and explain why.

The value for P indicates how likely a node is connected another node (i.e. its probability). So as the value of P gets higher more connections/edges appear. This also increases the density of the graph, because as we already claimed that probability of the new tie formation in a random graph reflects its density.

|  |  |  |
| --- | --- | --- |
| Shape, circle, polygon  Description automatically generated | Shape, circle  Description automatically generated | Shape, circle, polygon  Description automatically generated |

1. For a large n (e.g., n=1000), study the relation between clustering coefficient of the network and p, and explain the reason for such a relation. (You can use the function of transitivity (graph.object) to calculate clustering coefficient).

Transitivity: 0.2001041 and the chosen P value was 0.2.

As the probability of connection p increases, the transitivity of the network also increases. This is because as more edges are added to the network, nodes become more likely to form triangles, and therefore the transitivity increases.

Question 7 (2 points):

Check the clustering coefficient and average path length of the Regular, SW1, SW2 and SW3. Describe the trend of clustering coefficient and average path length as *p* increase. Which graph does mimic the desirable attributes of a small world network?

SW1\_clustering\_coef: 0.6777076

SW1\_avg\_path\_length: 8.394292

SW2\_clustering\_coef: 0.6387643

SW2\_avg\_path\_length : 4.26903

SW3\_clustering\_coef: 0.3706598

SW3\_avg\_path\_length: 2.919309

The clustering coefficient and average path length decreases as P gets higher. Small world networks are a type of network that have both local clustering and short average path lengths between nodes. They are characterized by a few highly connected hubs that are interconnected to many less connected nodes. So, in this case the SW2 graph represents the small world network the best since the average path length is way shorter than in SW1 and the clustering coefficient did not decrease that much.

## Question 8 (5 points):

You might realize not every value of p can return you a small-world network that you are looking for. Then a question arises as how can one find the range of p. In the Figure 2 of Watts and Strogatz (1998) (<https://www-nature-com.proxy.library.uu.nl/articles/30918>), it explains how can one decide the range of p by looking at the dynamics between path length and clustering coefficient.

Afbeelding met grafiek

Automatisch gegenereerde beschrijving

Figure 1 Figure 2: Characteristic path length L(p) and clustering coefficient C(p) for the family of randomly rewired graphs described in Fig. 1.

Here *L* is defined as the number of edges in the shortest path between two vertices, averaged over all pairs of vertices. The clustering coefficient *C*(*p*) is defined as follows. Suppose that a vertex *v* has *kv* neighbours; then at most *kv*(*kv* − 1)/2 edges can exist between them (this occurs when every neighbour of *v* is connected to every other neighbour of *v*). Let *Cv* denote the fraction of these allowable edges that actually exist. Define *C* as the average of *Cv* over all *v*. For friendship networks, these statistics have intuitive meanings: *L* is the average number of friendships in the shortest chain connecting two people; *Cv* reflects the extent to which friends of *v* are also friends of each other; and thus *C* measures the cliquishness of a typical friendship circle. The data shown in the figure are averages over 20 random realizations of the rewiring process described in [Fig. 1](https://www-nature-com.proxy.library.uu.nl/articles/30918" \l "Fig1), and have been normalized by the values *L*(0), *C*(0) for a regular lattice. All the graphs have *n* = 1,000 vertices and an average degree of *k* = 10 edges per vertex. We note that a logarithmic horizontal scale has been used to resolve the rapid drop in *L*(*p*), corresponding to the onset of the small-world phenomenon. During this drop, *C*(*p*) remains almost constant at its value for the regular lattice, indicating that the transition to a small world is almost undetectable at the local level.

1. Start with a regular network of size=300, nei=6, first reproduce the Figure 2 of Watts and Strogatz (1998). Then provide the range of p which can turn this regular network (size=300, nei=6) into a small-world network.

*Answer:* the range for rewiring probability P is approximately [0.0004, 0.012] see Fig\_ for the picture of the range. As we see on the Fig. \_, if the rewiring probability P goes beyond this range it starts to look like a random graph.

1. Do you need to rewire significant amount of connections to make the network smallworld-like?

*Answer:* No, when we increase the rewiring probability at some point, we see that the clustering coefficient starts to decrease, which starts to make the graph resemble the random graph. So, the desired range is when the clustering coefficient does not decrease, but the diameter drops, which can be seen on the Fig \_.

1. In the paper of Watts and Strogatz (1998), they pointed out that the value of p has two important implications:

“The idealized construction above reveals the key role of short cuts. It suggests that the small-world phenomenon might be common in sparse networks with many vertices, as even a tiny fraction of short cuts would suffice.”

*Answer:* When we worked on the “six-degree separation”, we saw that no matter how big the degree centrality of a node is, it is likely with probability of 90% that it will have the maximum shortest path no bigger than 6. That also is connected to the essence of the small world models, meaning that for a particular node it is not important to have a lot of connections to reach any of other nodes in a network, however, it is crucial to have a connection to a node with a shortcut, or a node which has a connection to a node with a shortcut.

Therefore, in order to make a small world network from a one-dimensional lattice we do not need to introduce that many of shortcuts. We just need to put that many that we balance in the tradeoff between tiring to minimize the diameter and not to reduce the clustering coefficient.

“Thus, infectious diseases are predicted to spread much more easily and quickly in a small world; the alarming and less obvious point is how few short cuts are needed to make the world small.”

Use your own words to explain these two implications. For the second implication, connect it with the spread of COVID.

*Answer:* Here, in case of the virus or information spread, Wattz and Strogatz claim that in the small world network they spread faster than in the regular networks (i.e. with small amount of shortcuts). For the disease to spread two properties are important: how quickly a node falls out of the network(dies) and how infectious it is.

In case the node dies form the disease too fast before it spreads out itself, then the whole network will not be affected. Here comes the importance of the shortcuts and the average diameter of the network. If the diameter is too big, there is a big chance that disease will wipe out the infected nodes or they will recover before it spreads. However, in case of the small world networks, due to the sufficient number of shortcuts, the disease spreads faster and can capture the majority of the nodes.

There is second point, which is the number of shortcuts to be removed for prevention of spreading. In the example of Bearman et. al (See Fig \_) it is enough to remove the bridge between the coloured infected nodes and blank healthy nodes, but such structure is not present in the small world networks. In case of COVID, the ideal and most efficient solution would be to separate everyone and remove all the connections, but it is impossible. The optimal solution from the perspective of the small world models, was to reduce shortcuts to a number, which will restrict the time needed for a virus to spread close to the time after which the disease becomes inactive of not contagious. What the challenge is, how Wattz and Strogatz said it, to find such a number of shortcut reduction.

|  |  |
| --- | --- |
| Shape, circle  Description automatically generated | Shape, circle  Description automatically generated |
|  |  |
| Shape, circle  Description automatically generated | Shape, circle  Description automatically generated |

|  |  |
| --- | --- |
| **Fig \_ The distribution of the clustering coefficient and diameter per rewiring probability** | The bridge structure [Bearman, P. S., Moody, J., & Stovel, K. (2004). Chains of affection: The structure of adolescent romantic and sexual networks. *American journal of sociology*, *110*(1), 44-91.] |
| A picture containing text, sky  Description automatically generated | Chart  Description automatically generated |

Question 9 (3 points):

1. What does the power in the above function mean? How can it govern the structure of the network? (Hint: Change the value of power from 0.05, 0.5, 1, 1.5; See how the plot evolves; if you still fail to see the difference, visualize the vertex size according to the edge number, you can consider the code below.)

*Answer:* Power argument in this function means the alpha in the formula of the scale-free model, which is , where is the probability of the of a new node to connect to node i. The interpretation of this parameter is the preferential attachment mechanism, meaning that if alpha is more or equal than one, then the more degree the node has, the more probable that a new node will connect to it. That is called the super linear probability dependency, and in case of alpha being less than one, it is called sublinear probability dependency; in such case we do not have preferential attachment, but rather get a randomized network

1. For two networks with a power of 0.5 and 1.5, respectively, what will be their resilience for 1) random attack, and 2) targeted attack? (the meanings of ‘random attack’ and ‘targeted attack’ are the same as what is mentioned in Lecture 6, scale-free network)

|  |  |
| --- | --- |
| Chart, scatter chart  Description automatically generated  Diameter = 12 | Chart, scatter chart  Description automatically generated |
| Chart, scatter chart  Description automatically generated  Diameter = 8 | Chart  Description automatically generated |
| Chart, scatter chart  Description automatically generated | Chart, scatter chart  Description automatically generated |
| Diameter = 12, N components = 1 | Diameter = 10, N components = 8 |
| Chart  Description automatically generated | Chart, scatter chart  Description automatically generated |
| Diameter = 8, N components = 2 | Diameter = 5, N components = 27 |
|  |  |

So, when we apply Random Failure, a node is emoved randomly, when we apply targeted Attack, the most connected node is removed. For both the Scale free network (i.e. the one with alpha = 1.5) and the Randomized sublinear the diameter have not changed after random attack. However, after target removal of the most connected nodes, the diameter of Scale free graph dropped by 37,5% from 8 to 5, while the diameter of the Random network the decrease was only 16.5% from 12 to 10.

In contrast to the results of the same procedures from the work of Réka Albert, Hawoong Jeong and Albert-László Barabási(see Fig \_), our models showed decrease of the diameter in case of the target attack, which happened because our networks were split into many components – into 8 and into 27 in random and scale-free networks, respectively(2000). That could be explained with the size and density of the networks from their paper that is much larger than ours. None the less, the number of resulting components is much larger in scale–free networks, which also provides the same idea as the Réka et al.: The scale–free networks are less resilient with respect to target removals.

Fig \_ Diameter after random and target attacks (Réka et al.)

Chart

Description automatically generated with medium confidence

Réka Albert, Hawoong Jeong & Albert-László Barabási., 2000. Error and attack tolerance of complex networks. Nature, 406, pages378–382

Bearman, P. S., Moody, J., & Stovel, K. (2004). Chains of affection: The structure of adolescent romantic and sexual networks. *American journal of sociology*, *110*(1), 44-91.]