CAI Lab 7: Network analysis

03/12/2022
Sonia Castro Paniello, Alba Koehler Borras

Exercice 1. Plot the clustering coefficient and the average shortest-path as a function of the parameter p of the WS model.

We create the probability vector taking into account the fact that we are scaling it logarithmically later for the plot so the values should tend more towards 0 than towards 1. The values that we have chosen are p = [0, 0.0001, 0.0025, 0.005, 0.0075, 0.01, 0.025 , 0.05, 0.075, 0.1, 0.25, 0.5, 0.75, 1]. We normalize both variables as indicated by the statement so that the resulting graph is better visualized.

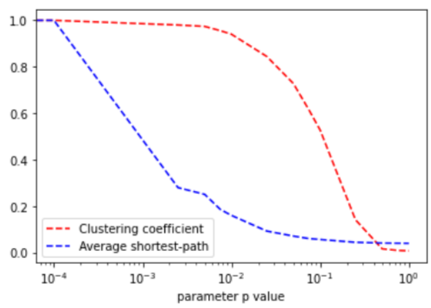

Figure 1: Clustering coeffient and average shortest-path for different p values.

As expected, using this method to create de graph the diameter and the clustering coefficient decrease as the p value increases. We have obtained the result already seen in class that for a value of p $\approx$ 0.01 , the model achieves high clustering and small diameter. With these measurements we are verifying the small-world property. This property says that a friend of a friend is often a friend (a high clustering coefficient would confirm this) and that there are short paths between most nodes (a small diameter would confirm this).

Exercice 2. Plot the average shortest-path length as a function of the network size of the ER model.

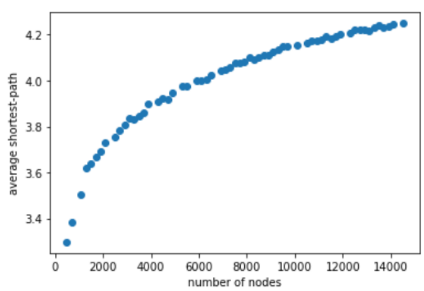

Figure 2: Average shorttest-path length in function of the number of nodes

In this exercise we have encountered the problem of computing time. The values of n chosen have been from 500 to 15000 in steps of 200. Reaching larger n increased the time too much, so we have prefered to only go up to a value of 15000 and decrease the size of the "step" to 200 to have more samples, instead of increasing both the step and the maximum n.

In addition, at first we proposed checking that p was greater than the lower limit of the result indicated in the statement ($p > \frac{(1+\epsilon)\ln n}{n}$), and choosing a random number between that and 1. We realized that since p is the probability that an edge will be created between each pair of nodes, when it was high many edges were created. And therefore with this approach the calculation of the average shortest path consumed too much time.

So, to facilitate obtaining results, taking into account that a p greater than the value of the given limit already indicates that the graph is almost surely connected and that the lower this is, the faster the calculations are, instead of taking a random number between the limit and one decided to add a small amount to that limit. We decided this element to be 1/n, so depending on the number of nodes n, because in this way the more nodes the lower the probability that an edge will be created. This prevents us from creating too much edges when the number of nodes is higher which would slow the computation.

Exercice 3.Plot a histogram of the degree distribution of a BA network. What distribution does this follow? Can you describe it?

The degree distribution of most real-world networks follows a power-law distribution, for creating scale-free grafs we can use the Barabasi-Albert that takes into account preferential attachment. We can confirm with our plots that the histogram follows a power-law. When ploting the loglog plot we can see an aproximated line (Figures 5 and 6).

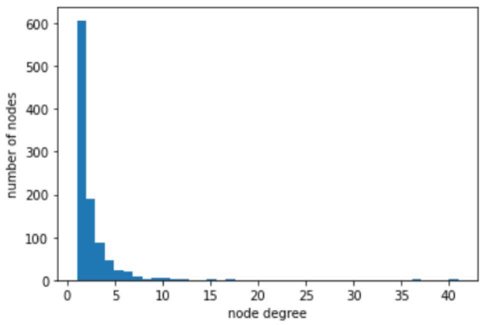

Figure 3: Degree distribution histogram.

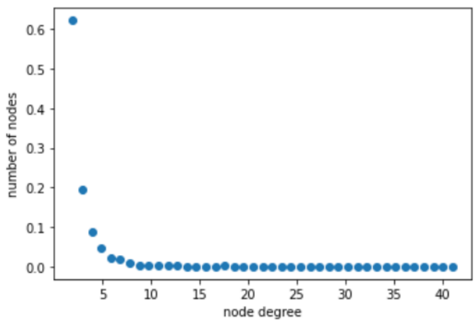

Figure 4: Degree distribution.

Once we know what it follows a power law, we use linear regression to approximate its parameters. We ran into the problem that some of the values in the vector of number of nodes with degree i were 0. Therefore, when calculating the logarithm to be able to do the regression, it gave us infinite values. We have been looking for ways to solve it, in the end we decided to eliminate the samples with 0 and calculate the approximation with the remaining samples. The result if we take into account that $y = cx^{-a}$, is c = 2.40786057 and a = 2.640835

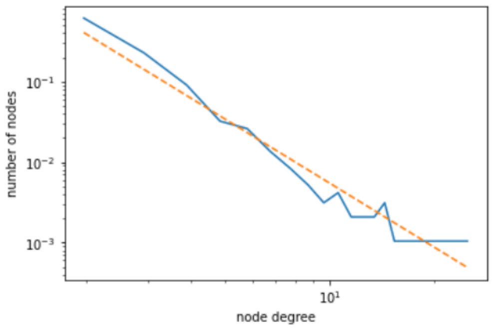

Figure 5: Degree distribution without 0 and its linear regression aproximation.

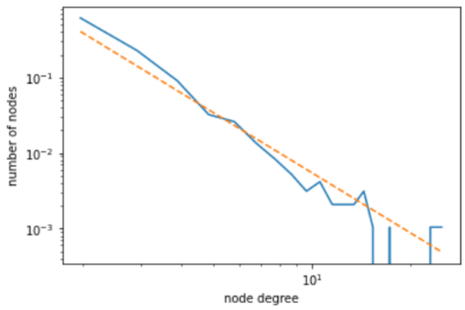

Figure 6: Original degree distribution and the linear regression aproximation.