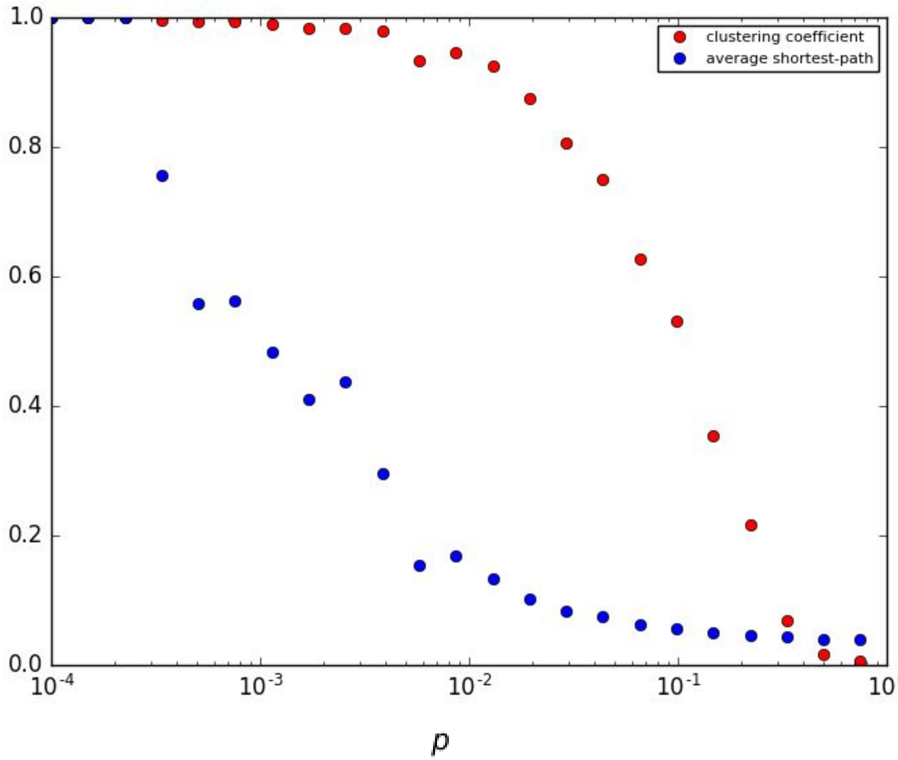
CAIM - Laboratori 7

José Ángel Casado Pérez
Òscar Mañas Sanchez

Task 1

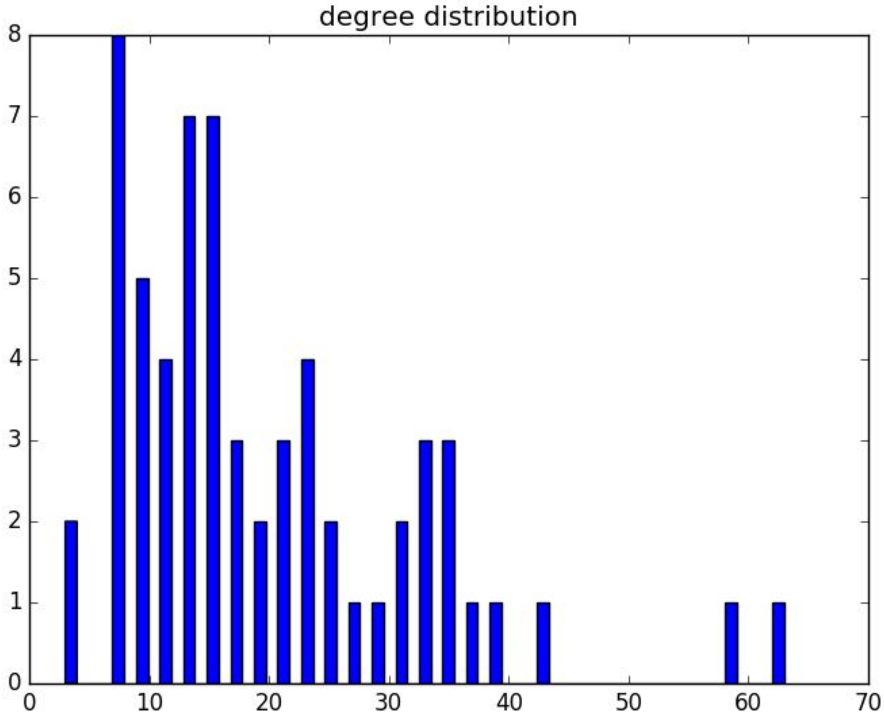
In order to reproduce the chart seen in class, we implemented a function that basically creates different graphs following the Watts-Strogatz model, each of them with a different *p* parameter value - the rewiring probability - ranging from 0.0001 to 1, and keeps the clustering coefficient and the average shortest-path distance for each of them; finally, these values are normalized dividing them by the highest one (with $p=0.0001$).

We obtained, for a specific run of our script, the following plot - with the *x* axis in logarithmic scale -, which resembles to the one seen in class, with the average shortest-path distance and clustering coefficient exponentially decreasing with increasing values of *p*. We observe that $p \approx 10^{-2}$ is the point where the model achieves high clustering and small diameter at the same time, both qualities of real networks.

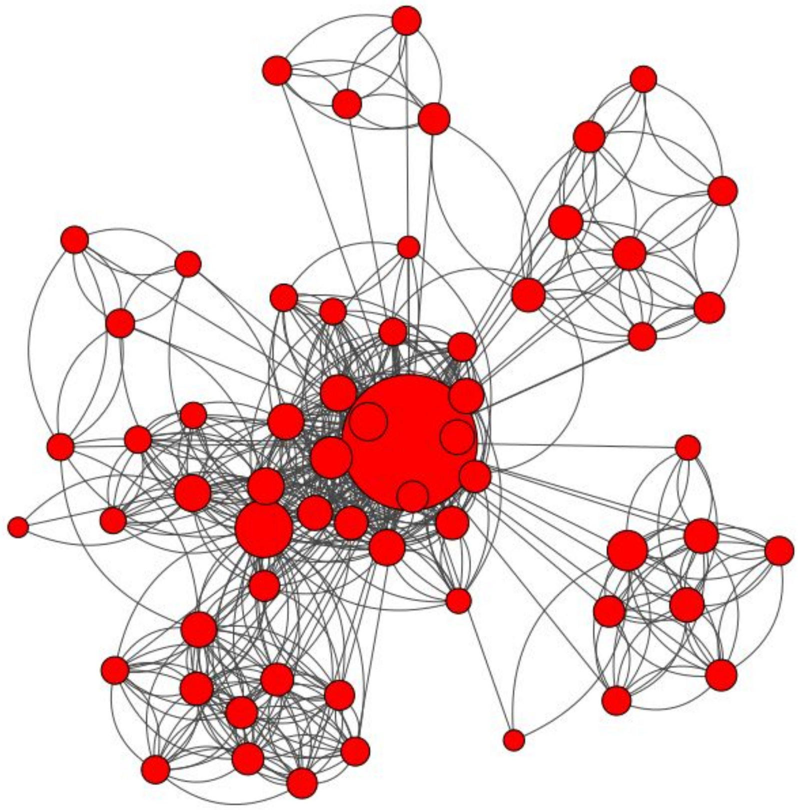


Task 2

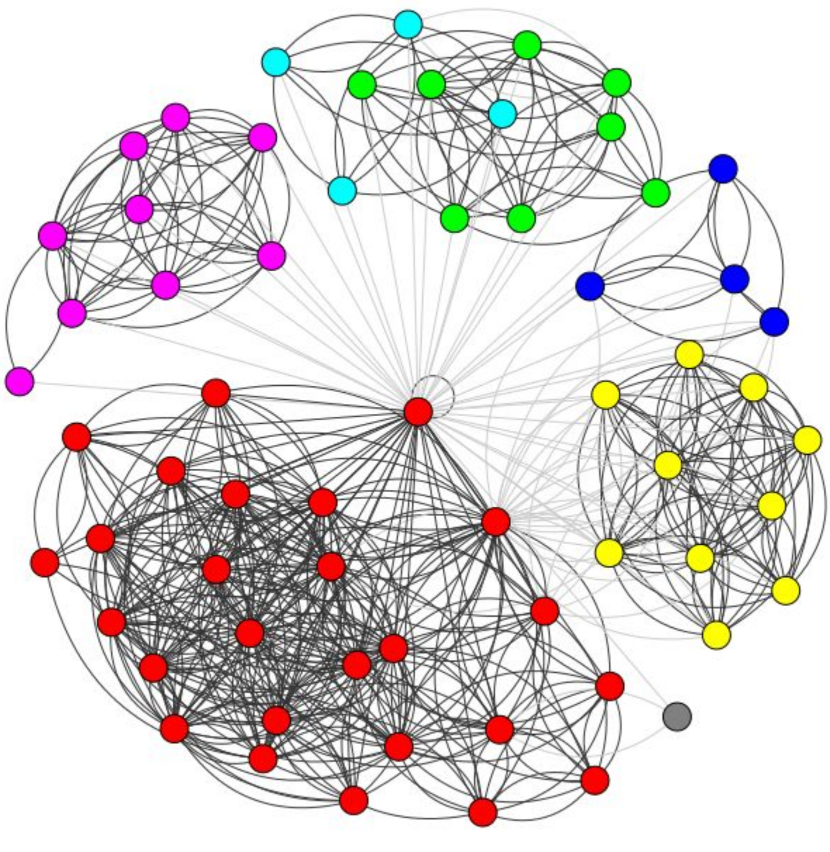
After loading the network from the file, we checked, using a small script, its characteristics. We could see that it was composed of 62 nodes and 602 edges, with a diameter of 2 - quite connected - and a degree distribution of 0.523. To check whether it looked like a random network or not, we plotted in a histogram the distribution of the vertices' degree, as shown in the next figure.



It clearly looks like the distribution of the degrees follows sort of a power-law distribution, where the vertices are more likely to be low-connected (the vertices at the "spikes" of the function, with 10-15 connections), yet there are some that are highly connected (the ones at the tail of the function); this is common for real networks, where this tail implies the existence of hubs. Therefore, the graph looks more like a real network than a random network. In the following figure we can see this degree distribution, where the vertices of the graph have been displayed with a size proportional to their Pagerank value.



Finally, in order to study the possible communities within the graph, we used the Girvan-Newman algorithm, which uses divisive hierarchical clustering to find all the communities in it. It's based on the idea that bridges between communities must have high edge betweenness. We obtained the following graph by coloring every node depending on its community.



We can see that there are 7 communities of different sizes, the red one being the biggest, and the grey one being the smallest. We can see the community size distribution in the next histogram, where there are two communities with four vertices (light and dark blue ones) and a single community for each of the other sizes - one, eight, nine, ten and twenty-six vertices, respectively.