Information Retrieval and Analysis
December 2020

Lab Session 11: Introduction to iGraph

PROJECT REPORT

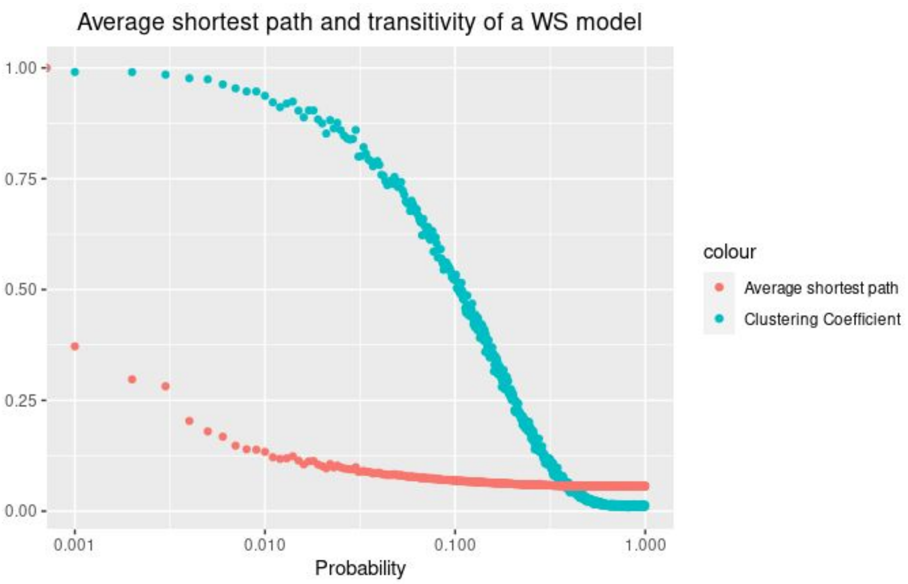
Elías Abad Rocamora
Victor Novelle Moriano
Barcelona, UPC - FIB & FME & ETSETB

Introduction

In this lab session we will analyze two different random graph generators: Watts-Strogatz and Barabasi-Albert models. These two models are designed so that they satisfy some properties of social networks. WS is designed so that the Small World property is satisfied and BA focuses on having a degree distribution that follows a power-law.

In order to analyze these models, we will answer some questions regarding the above-mentioned properties.

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

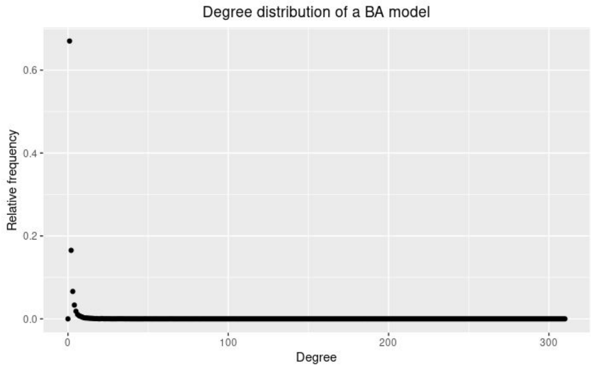


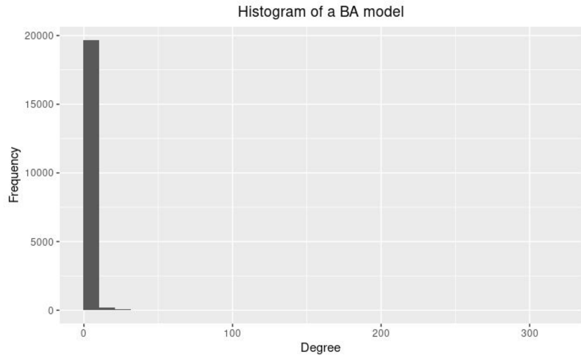
In order to create this plot we used R programming language. We iterated over a sequence of values of p ranging from 0 to 1 and incrementing 0.001 in each step, for each step we created a new graph with 1000 nodes and computed both the average shortest path length and the global clustering coefficient. After the computation of every value, we plotted it with the x axis in log-scale to better see the evolution for small values of p.

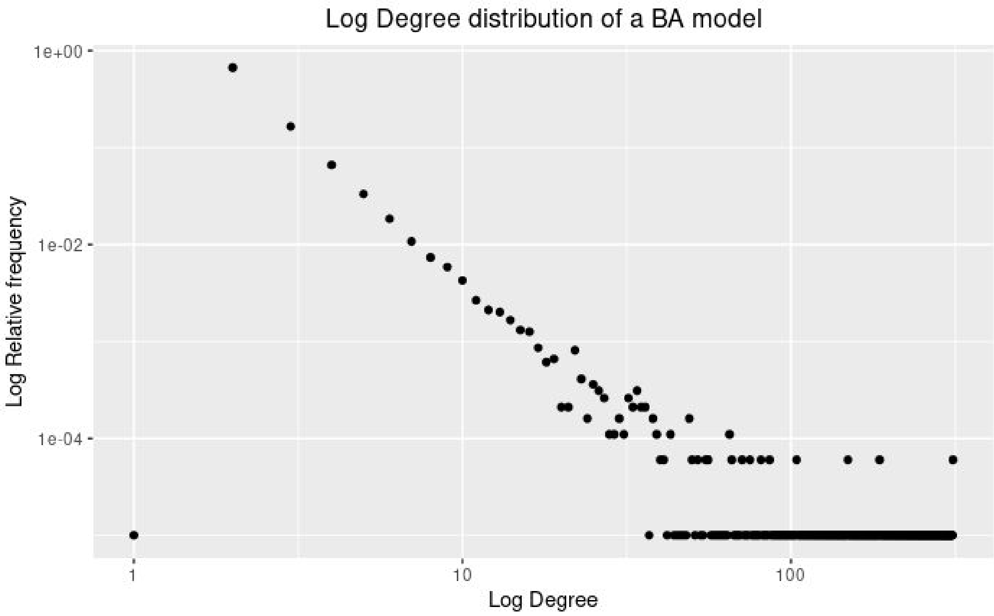
As we can observe, the average shortest path length decreases very rapidly and then stabilizes around 0.05. Alternatively, the clustering coefficient starts decreasing very slowly and then starts to drop from 0.05.

If we wanted to model the behaviour of a social network, we would like to have a high clustering coefficient and a low average shortest path distance (small world phenomenon), So based on these results, with this number of nodes (1000), one would have to choose a value of p ranging from 0.01 to 0.1 in order to model a social network.

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







These plots were again made with the R programming language. A Barabasi-Albert graph with 20000 nodes was created and then the distribution of its nodes' degrees was plotted with and without aggregating in bins.

In order to check that the distribution of the nodes' degrees follows a power-law, we have made a third plot with both axes in log scale. As can be observed, the points follow a straight line exactly as a power-law should do.
For big values of the degree, we see that the points form an extrange pattern. We think that this might be due to numerical errors.

As it can be seen, the frequency distribution is noticeably different from the one we would obtain if we performed the same analysis using a Erdős-Rényi model, which follows a binomial distribution as the mean degree is computed as $p \cdot n$. As a result of that, and looking at the *power-law plot*, we can affirm that we are in front of a scale-free model.