CSN-MIRI Complex and Social Networks - Lab 1
Introduction to `igraph`

presented by
Thomas Maria Frey
Daniel Benedí García

submitted to
Ramon Ferrer-i-Cancho and Argimiro Arratia
at the Universitat Politècnica de Catalunya - Facultat D'informàtica De Barcelona

September, 2023

Introduction

Usage of random graphs is relevant in different areas in order to model several behaviors. During this assignment, we will try to empirically understand the properties of some models in random graphs such as the Erdös-Rényi model (Section 3) and the Watts-Strogatz model (Section 2).

Watts-Strogatz model

The Watts-Strogatz model given the desired number of nodes, N, the mean degree K and a parameter $0 \leq p \leq 1$ will construct an undirected graph without local clustering, in contrast with the Erdös-Rényi model in which there is local clustering.

To comprehensively analyze the influence of the parameter p on both the clustering coefficient and the average shortest path, we will construct plots displaying the normalized progression of these metrics with varying p. Specifically, we will generate 100 random graphs of a specified size and mean degree, varying the parameter p and taking the average. In Figure 2.1, the outcomes of this experiment are presented for a graph comprising 1000 nodes and an average degree of 2.

The depicted results confirm to the anticipated behavior of the random graph model. Initially, the average shortest path remains relatively stable until a specific p value is reached, at which point it dramatically decreases to the normalized value of 0. This phenomenon indicates a highly interconnected network where almost all nodes are connected. Concurrently, the clustering coefficient starts off at a high value and then rapidly declines as the inter-connectivity intensifies, resulting in a large cluster and subsequently, a diminished coefficient.

During our experimental investigations aimed at determining suitable mean degree values, an intriguing anomaly emerged, as depicted in Figure 2.2. Contrary to expectations, the Clustering Coefficient did not approach zero even at significantly high values of p. This behavior can be attributed to the inherent structure of the random graph model, where the presence of local clustering is preserved even at

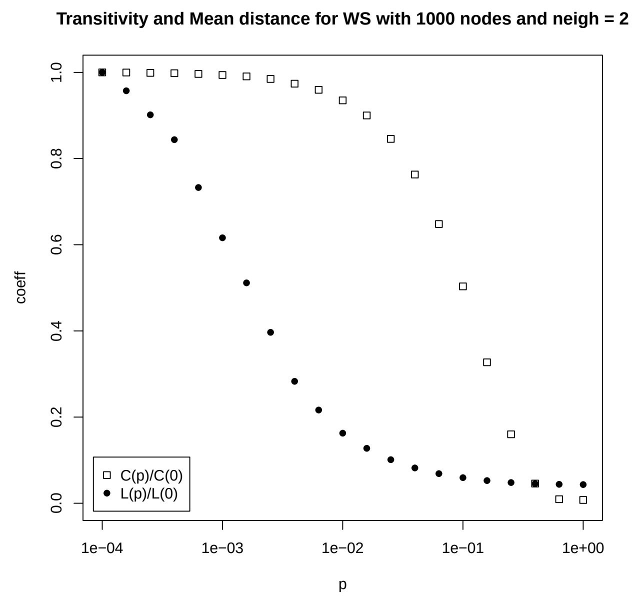

Figure 2.1: Evolution of Clustering Coefficient $L(p)$ and the average shortest-path $C(p)$

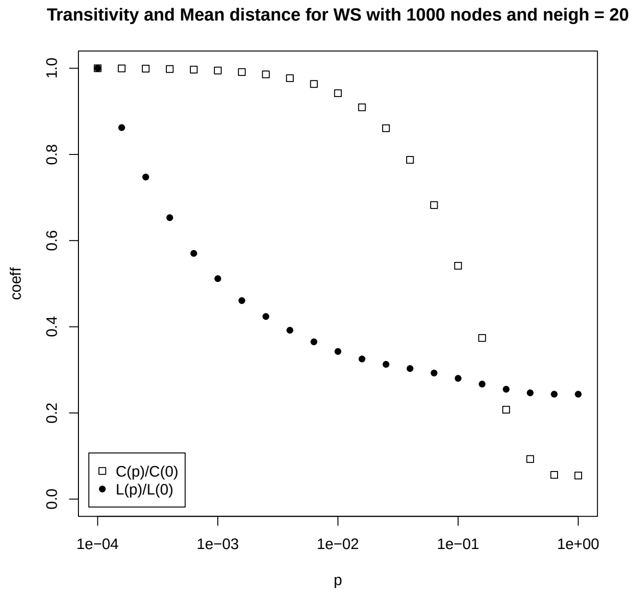

Figure 2.2: Evolution of Clustering Coefficient $L(p)$ and the average shortest-path $C(p)$

higher p values. Additionally, our choice of normalization with respect to $L(0)$ contributes to this outcome, ensuring that the clustering coefficient doesn't deviate significantly from $L(0)$ due to the graph's elevated mean degree.

Erdös-Rényi model

The Erdös-Rényi model is used to generate random graphs. It is a model that allows for the generation of graphs with a small diameter. This is a property that is also found in real networks. Contrary to real networks or the other models discussed in the lectures, it does not posses the property of high clustering nor it is scale-free. To plot the average shortest-path as a function of the network size the two variables for the Erdös-Rényi model G(n,p) have to be chosen accordingly. The graph from the task sheet uses zero to one million nodes and an unknown *p* value. To calculate a suitable *p* value that allows for a connected graph we use the following formulas from [1].

- If $p < \frac{(1-\epsilon)\, ln\, n}{n}$ then a graph in G(n,p) will almost surely contain isolated vertices, and thus be disconnected.
- If $p > \frac{(1-\epsilon)\, ln\, n}{n}$ then a graph in G(n,p) will almost surely be connected.

Because of insufficient hardware we had to reduce the maximum number of nodes *n* that are considered to one hundred thousand instead of one million. Furthermore in the case that the graph is not fully connected we consider only the largest fully connected sub-graph. The graph from the task can not be recreated with a constant *p* value for all values of *n*. For example using a *p* value of 0.1 for all values of *n* results in the graph that can be seen in 3.1. We believe that using a rather large *p* value of 0.1 leads to the creation of large central components in the graph once the *n* value is adequately large. These central components lead to small shortest paths on average. Therefore the value of *p* has to be chosen in accordance of the value of *n* to recreate the graph from the assignment. We have used the p value of $2\, ln(n)/n$, this is also recommended by the relevant literature to this topic. The resulting graph can be seen in 3.2. This graph shows that the average shortest path increases sharply with with the amount of nodes before reaching an stable value.

Bibliography

[1] Paul L. Erdos and Alfréd Rényi. On the evolution of random graphs. *Transactions of the American Mathematical Society*, 286:257–257, 1984.

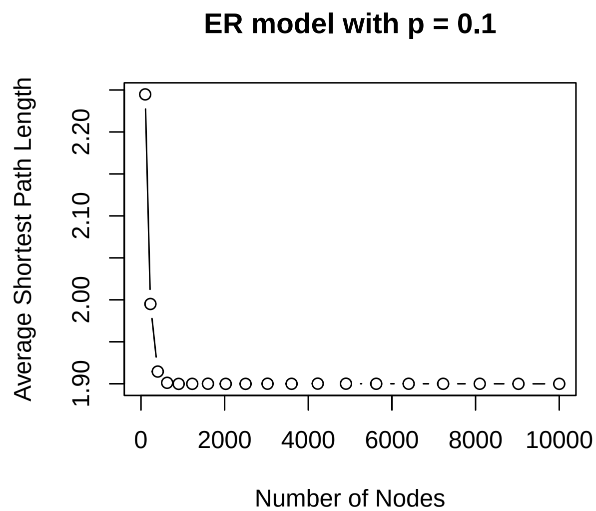

Figure 3.1: ER model with a static *p* of 0.1 and a maximal *n* of 10000

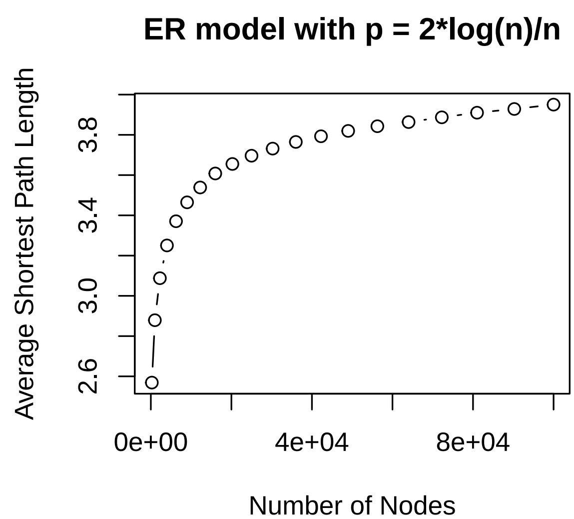

Figure 3.2: ER model with a dynamic *p* of $2\,ln(n)/n$ and a maximal *n* of 100000