Intorduction to igraph - Laboratory 1 Report

Complex and Social Networks (CSN)

José Ángel Vicente Porres, Odysseas Kyparissis

September 25, 2023

Facultat d'Informàtica de Barcelona (FIB)
Universitat Politècnica de Catalunya (UPC) - BarcelonaTech

Contents

List of Figures

1 Introduction

Networks, whether they represent social interactions, transportation systems, or biological pathways, have become a ubiquitous framework for understanding the complex web of connections that underlie the functioning of diverse systems in our world. These networks, often referred to as graphs, have been extensively studied, modeled, and analyzed to uncover the fundamental principles governing their structure and behavior. Two prominent models in the study of networks, the Watts-Strogatz (WS) model and the Erdős-Rényi (ER) model, have played pivotal roles in our quest to comprehend real-world networks and draw meaningful comparisons between them.

1.1 The Significance of Network Models

Network models, such as the Watts-Strogatz and Erdős-Rényi models, serve as essential tools for researchers and scientists seeking to unravel the intricate topology of networks and their dynamic behaviors. These models provide simplified representations that capture key features of real networks while allowing for systematic exploration and hypothesis testing. In this laboratory report, we delve into the analysis of these models with a specific focus on two critical aspects, one for each separate model, explained in the following subsections.

1.1.1 Erdős-Rényi Model

This classic random graph model ($G_{n,p}$), named after mathematicians Paul Erdős and Alfréd Rényi, assumes that each pair of nodes has a fixed probability p of being connected. By plotting the average shortest-path length as a function of network size n, we explore how network connectivity scales with the number of nodes. This analysis provides valuable information about the efficiency of information flow in random networks, enabling comparisons with real-world networks of varying sizes.

1.1.2 Watts-Strogatz Model

This model addresses the balance between local clustering and global connectivity in networks. It introduces a parameter p that controls the rewiring probability of edges in a regular lattice, effectively transitioning between a highly clustered network and a random network. By plotting the clustering coefficient and the average shortest-path length as functions of p, we gain insight into how networks can combine the small diameter of random networks with the high clustering coefficients observed in real-world networks.

1.2 Comparing Models with Real Networks

One of the primary motivations for studying network models like WS and ER is their utility in comparing the properties of synthetic networks with those observed in real-world systems. It is well-established that real networks exhibit intriguing characteristics, often differing from the idealized models. Notably, real networks tend to exhibit small diameters similar to ER or random models while boasting high clustering coefficients akin to WS networks.

These observations raise important questions: To what extent do these models capture the essence of real networks? How well do they mimic the intricate interplay of local and global connectivity that characterizes networks found in nature? By systematically varying parameters and network sizes in these models, we aim to shed light on these questions and draw meaningful conclusions about the quality of model-network comparisons.

2 Methodology

In this chapter, we describe the methodology employed to address the two tasks mentioned in the lab report. The first task involves plotting the clustering coefficient and the average shortest-path as functions of the parameter p in the WS model. The second task focuses on plotting the average shortest-path length as a function of the network size in the ER model.

2.1 Task 1: Watts-Strogatz Model

To explore the WS model, we aim to investigate how the clustering coefficient ($C(p)$) and the average shortest-path length ($L(p)$) change with varying p values. As mentioned before, the WS model starts with a regular lattice structure and rewires edges with probability p.

2.1.1 Clustering Coefficient ($C(p)$)

The clustering coefficient of a network measures the degree to which nodes tend to cluster together, indicating the presence of local connectivity within the network. For the WS model, the clustering coefficient ($C(p)$) can be defined as follows:

$$C(p) = \frac{3 \cdot \text{Number of triangles in the network}}{\text{Number of connected triples of nodes}}$$

In the context of this analysis, $C(p)$ represents the average clustering coefficient across multiple simulations for a given p value.

2.1.2 Average Shortest-Path Length ($L(p)$)

The average shortest-path length in a network measures the average number of edges that must be traversed to go from one node to another. It quantifies the network's efficiency in information or signal propagation. In the context of the WS model, the average shortest-path length ($L(p)$) can be defined as follows:

$$L(p) = \frac{2}{N(N-1)} \sum_{i>j}^{N} d_{ij}(p)$$

Where:

N : Number of nodes in the network

$d_{ij}(p)$: Shortest-path distance between nodes i and j for a given p

$L(p)$ represents the average shortest-path length across multiple simulations for a given p value. We normalize it by dividing by $L(0)$ (the average shortest-path length at $p = 10^{-4}$) for better visualization and comparison.

2.1.3 Steps of the Analysis for Task 1

1. **Parameter Setup:** We fix the number of nodes (N) to 500 and the initial neighborhood size (k) to 4. We vary p from 10^{-4} to 1 in increments of 0.2. For each p value, we conduct 100 simulations to ensure statistical significance.

2. **Simulation:** We create a custom R function, `compute_metrics`, which generates WS graphs with the specified parameters (N, k, p) and calculates the clustering coefficient and average shortest path for each graph.

3. **Data Collection:** We run the simulations and collect the results for each p value. For each simulation, we compute the average clustering coefficient ($C(p)$) and average shortest-path length ($L(p)$).

4. **Normalization:** To facilitate comparisons, we normalize the computed values by dividing them by their initial values at $p = 10^{-4}$.

5. **Plotting:** We generate a plot of p against the normalized clustering coefficient ($C(p)/C(0)$) and the normalized shortest-path length ($L(p)/L(0)$). The plot visualizes how the network's properties change with p.

2.2 Task 2: Erdős-Rényi Model

In the second task, we analyse the ER model, specifically focusing on the average shortest-path length ($L(n)$) as a function of network size (n), since the value of p is being kept fixed. As previously mentioned, the ER model generates random graphs with n nodes, where each pair of nodes is connected with probability p.

2.2.1 Steps of the Analysis for Task 2

1. **Parameter Setup:** We set the simulation parameters as follows: $\epsilon = 0.01$ (a small constant), num_simulations = 50 (simulations for each network size), and max_2_n = 14 (maximum power of 2 for the number of nodes).

2. **Network Size Selection:** We generate a range of network sizes, n, as 2^4 to 2^{14} to explore how average shortest-path length scales with network size.

3. **Simulation:** We create a custom R function, `generate_and_measure_ER`, which generates random ER graphs for given n and p values and calculates the average shortest path length for each graph.

4. **Data Collection:** For each network size n, we perform simulations by varying p such that $p = ((1 + \epsilon) \log(n))/n$. With the usage of the small constant ϵ, it is sure that the graphs that are constructed are connected with high probability. For each of the graphs in the sample we compute the average shortest. We calculate the average shortest-path length ($L(n)$) for each simulation.

5. **Plotting:** We generate a plot of network size (n) against the average shortest-path length ($L(n)$) for the ER model. This plot illustrates how the efficiency of information flow changes with the network size.

The methodology outlined above allows us to explore the behavior of the WS and ER models and draw comparisons between their properties, thereby enhancing our understanding of network structures and dynamics.

3 Results

In this section, we present the results of the analysis, focusing on the two distinct tasks, involving the generated figures.

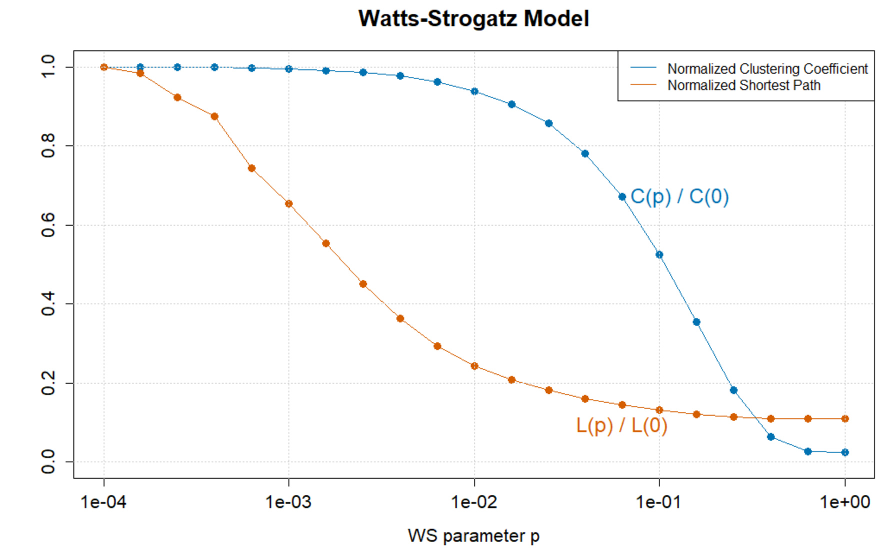

Figure 1: Normalized Clustering Coefficient and Normalized Shortest-Path Length as Functions of p in the WS Model.

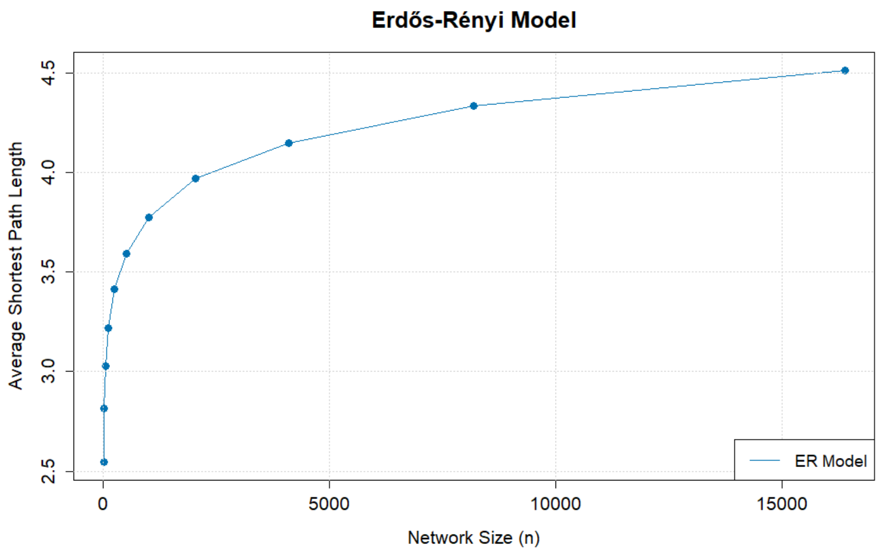

Figure 2: Average Shortest-Path Length as a Function of Network Size (n) in the ER Model.

4 Conclusion

In order to analyze Figure 1, the one associated to the WS model, we have to take into account the x axis is in logarithmic scale. We can see a really high decrease in the average longest path for small values of $p \leq 0.01$, and a really solid stability point around 0.1 times the original value. In the other hand, the clustering coefficient has a slower decrease, going from almost its highest value in $p = 0.01$ to reaching 0 at p around 0.5. However, for an appropriate value of $p \approx 0.01$ (1%), we observe that the model achieves **high clustering and small diameter**.

The WS model has two completely different extremes, if $p = 0$ we have an unrealistic graph, with too large diameter and clustering coefficient. As there is no rewiring, the structure has no kind of randomization and is too strict and symmetric, no valid for real use. However, if $p = 1$ we have the complete opposite situation, really low average shortest path and clustering. In this case we have that every edge is rewired, this kind of randomization is similar to the ER model, and in fact they have the same properties.

In Figure 2, we have the relationship between the shortest average path in a ER model and the size of the graph. As the complexity of these operation increases exponentially with n, and the computational power of the personal computers is quite low, we decided to add more points to the lower part of the spectrum, leading to a better precision in the faster changing area of the curve. Another possibility for getting a better performance during the simulations for large values of network sizes n, could have been to compute the harmonic average instead of the regular one, however this only makes a difference with long enough paths, where in this analysis no more than 8 steps are usually observed.

Moreover, there is a logarithmic growth in the average shortest path for the ER model as the size increases. This may indicate that the density in this kind of graphs gets higher with the size, even though the chances of adding an edge are in fact decreasing, $p = (1 + \epsilon)\frac{\log p}{p}$. For that reason the graphs generated this way have small diameter, and this property also tends to appear in real networks.

The main advantage once we have a plot similar to Figure 1, is that we can tune the parameters in order to get a model similar to the one we can study. The idea is to choose a value of p in a way that both average shortest path and clustering coefficient resembles the ones that we have in reality.