Universidade de Aveiro

Complex Systems Modelling (40780)

Complex networks

Gonçalo Freitas
NºMEC: 98012

Masters In Computational Engineering

Course Instructor:
Rui Américo Ferreira da Costa

July 8, 2023

Introduction

Graph theory plays a crucial role in understanding and analyzing complex systems, ranging from social networks to biological interactions and the internet. Within the realm of random graph models, *Erdős-Rényi* networks emerge as a fundamental framework for studying the properties and dynamics of networks with random connections. These networks belong to a class of statistical graph models, known as statistical ensembles, which are widely used to investigate the statistical properties of graphs.

Introduced by mathematicians Paul Erdős and Alfréd Rényi in the 1950s, *Erdős-Rényi* networks offer a probabilistic approach to constructing random graphs.

The key idea behind these networks lies in their construction process, where each pair of nodes is considered independently(uncorrelated), and an edge is added between them with a specific probability, denoted as p. As p increases, the resulting network becomes denser with a greater number of edges. Conversely, when p is small, the network tends to be sparse with fewer connections.

These random networks have been extensively studied in various fields such as graph theory and statistical physics, often serving as fundamental models for understanding phenomena in complex networks, such as phase transitions.

To generate *Erdős-Rényi* networks we can do as follows. Start with N isolated nodes, and then insert L edges between pairs of randomly chosen nodes. Make sure that each edge connects two distinct nodes (no self loops are allowed) and that there are no multiple edges connecting the same pair of nodes. We can store and represent the network using the adjacency matrix A. Start with a matrix A of 0's, then, when an edge is added between nodes i and j replace the entries A_{ij} and A_{ji} by a 1 (we can also use the adjacency matrix to avoid connecting the same pair of nodes more than once).

With this, the degree of vertex i, i.e, the number of edges connected to this vertex, is defined as

$$q_i = \sum_{j=1}^{N} A_{ij} \tag{1}$$

As for the degree distribution, which represents the probability that a randomly chosen node has degree q, we can calculate it as:

$$P(q) = \frac{N(q)}{N} \tag{2}$$

where N(q) is the number of vertices with degree q.

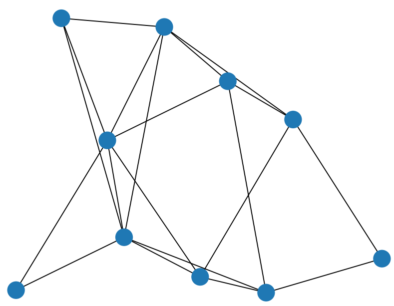


Figure 1: Example of a *Erdős–Rényi* graph with 10 nodes and 20 edges, generated using the NetworkX package [9]

Task 1

The main objective of this task is to calculate the mean degree, $\langle q \rangle$, the branching coefficient, B, and second and third moments of a *Erdős–Rényi* network, while comparing the results obtained with the theoretical results.

The mean degree of a *Erdős-Rényi* network with N nodes and $L = \frac{N}{2}c$ edges is given by Equation 3. [6]

$$\langle q \rangle = \frac{2L}{N} = c \tag{3}$$

For the simulated network, we will consider a number of nodes of $N = 10000$ and a number of edges of $L = 250000$ ($c = 50$), so that $N >> 1$ and $N >> \langle q \rangle$. In this limits, the network degree distribution approaches the Poisson distribution, [6], which is given by the following formula:

$$P(q) = e^{-\lambda}\frac{\lambda^q}{q!} \tag{4}$$

With this, knowing that in a Poisson distribution of parameter λ, the mean value is λ itself, and the mean degree of a *Erdős-Rényi* network is c, we can conclude that, for a large network of this type, $\lambda = c$.

As for the expected values of the second and third moments, due to knowing the degree distribution, we are able to write the equations for this moments, given by Equations 5 and 6, respectively.

$$\langle q^2 \rangle = \langle q \rangle^2 + \langle q \rangle = (50)^2 + 50 = 2550 \tag{5}$$

$$\langle q^3 \rangle = \langle q \rangle^3 + 3\langle q \rangle^2 + \langle q \rangle = (50)^3 + 3 \times (50)^2 + 50 = 132550 \tag{6}$$

Finally, the branching coefficient, B, that represents the mean number of edges emanating from a nearest neighbour, can be calculated by Equation 7. [6]

$$B = \frac{\langle q^2 \rangle}{\langle q \rangle} - 1 = \frac{2550}{50} - 1 = 50 \tag{7}$$

Now that we have presented and discussed the theoretical results we can compared this to the numerical results obtained using the algorithm described in the Appendix (steps 1-11), which implements a *Erdős-Rényi* network.

In Figure 2 we have a blue line that represents the numerical results of the degree distribution, $P(q)$, calculated using Equation 2. In this same figure, we also have red circles that represent a Poisson distribution with mean value of c, which is the theoretical degree distribution. From this figure, we can see that the obtained distribution closely matches the theoretical one.

As for the values of the mean degree $\langle q \rangle$, the branching coefficient B and the second and third moments obtained, these are represented in Table 1. In this table we also

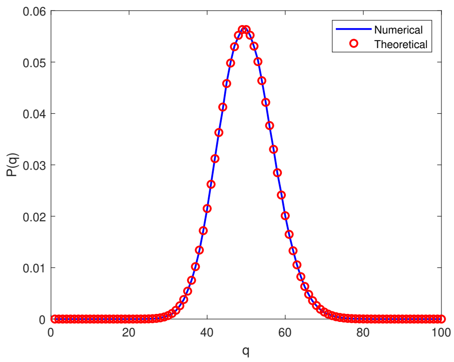


Figure 2: Comparison between the numerical degree distribution obtained for a *Erdős-Rényi* network, with $N = 10000$ nodes and $L = \frac{N}{2}c$ edges (assuming $c = 50$), with a Poisson distribution with mean value of 50

include the value of $B/\langle q \rangle$, that is suppose to be approximately equal 1, as in a *Erdős-Rényi* network $B = \langle q \rangle$, [6]. Note that the numerical values presented in this table represent the average of these values across the m simulations.

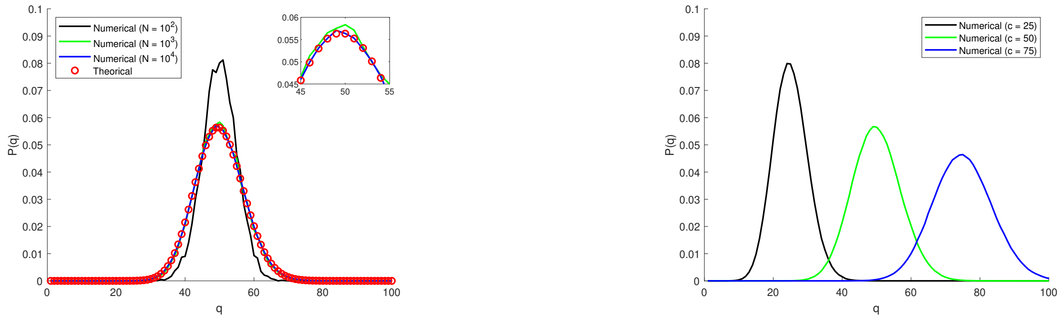
	Numerical	Theoretical
$\langle q \rangle$	50	50
B	49.9957	50
$\langle q^2 \rangle$	2549.7874	2550
$\langle q^3 \rangle$	132516.8928	132550
$B/\langle q \rangle$	0.9999	1

Table 1

As we can see from Table 1, the numerical results presented in this table are very close to their respective theoretical values, which leads us to conclude that the implementation of this *Erdős-Rényi* network is working well.

Two extra studies that can be made is what happens to the degree distribution and to the value of the mean degree, $\langle q \rangle$, when N or c increases. As mentioned above, for large networks ($N \to \infty$), the degree distribution tends to a Poisson distribution with a mean value of c.

With this information we can conclude that, when N increases it's expected that the numerical results get closer to a Poisson distribution with mean value c and that when c increases (assuming $N \to \infty$), the mean value also increases, shifting the curve of the degree distribution to the right. This is exactly what happens in our results, as we can see in Figure 3. With this, we can also conclude that, for an infinite *Erdős-Rényi* network, the mean degree is also infinite.



(a) (b)

Figure 3: Degree distribution of a *Erdős–Rényi* network, varying (a) N or (b) c

We can also check that when N increases, the degree distribution approaches a Poisson distribution by analyzing the result of the quotient between the variance and the mean, which, for this distribution, is suppose to be one, as the Poisson function is a probability distribution function with equal mean and variance, [4]. The numerical results obtained for this division are represented in Table 2, where we see that, as N increases, the result of the quotient approaches 1, as expected.

N	$\langle(q-\langle q\rangle)^2\rangle/\langle q\rangle$
10^2	0.4951
10^3	0.9591
10^4	0.9926

Table 2

The variance was calculated as follows:

$$\langle(q-\langle q\rangle)^2\rangle = \frac{\sum_{i=1}^{N}(q_i-\langle q\rangle)^2}{N-1} \tag{8}$$

Task 2

In this task, the goal is to calculate the clustering coefficient, which quantifies the extent to which nodes in a network tend to form clusters or groups.

This coefficient is given by the equation bellow, where n_{pt} and n_{tr} represent the number of possible and actual triangles in the network, respectively.

$$C = \frac{n_{tr}}{n_{pt}} \tag{9}$$

The number of possible triangles in the network can be given by:

$$n_{pt} = \frac{N}{6}\sum_{q} q(q-1)P(q) = \frac{1}{6}N\langle q(q-1)\rangle \tag{10}$$

As for the total number of triangles in the network, this can be calculated by the following equation:

$$n_{tr} = \frac{1}{6}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N} A_{ij}A_{jk}A_{ki} \tag{11}$$

Note that this calculation is of high computation order, $\mathcal{O}(n^3)$, which would lead to a high execution time. However, n_{tr} can also be written as: [6]

$$n_{tr} = \frac{1}{6}tr(A^3) \tag{12}$$

Where, $tr(A^3)$ represents the trace of the matrix A^3, which we can estimate as following.

$$tr(A^3) = \sum_{i\neq j\neq k} A_{ij}A_{jk}A_{ki} = \sum_{i\neq j\neq k} p^3 = N(N-1)(N-2)p^3 \approx \langle q\rangle^3 \tag{13}$$

Where we used the probability $p = \langle q\rangle/(N-1)$.

Therefore, using this estimation in Equation 12 we can say that, in *Erdős-Rényi* networks, where $\langle q(q-1)\rangle = \langle q\rangle^2$, the clustering coefficient equals:

$$C \approx \frac{\langle q\rangle}{N} = \frac{50}{10000} = 0.005 \tag{14}$$

Using our algorithm and Equations 10, 11 and 9, we obtained the values represented in Table 3, where we can see that the value of the numerical clustering coefficient is in fact approximately to $\langle q\rangle/N$. We can also conclude that the value obtained for n_{tr} is also close to it's estimation, $\langle q\rangle^3/6 \approx 20833.33$. Note that, as in **Task 1**, the values presented in this table represent the average of these values across the m simulations.

	Obtained
n_{pt}	4166115.50
n_{tr}	20839.53
C	0.005002

Table 3

Task 3

In this final part of the project, the goal is to use the results of the second and third moments from **Task 1** to calculate the Pearson coefficient, which quantifies the correlation between degrees of nodes and it's given by the following formula:

$$\rho = \frac{\sum_{i=1}^{N}\sum_{j=1}^{N} A_{ij}(q_i - Q)(q_j - Q)}{N\langle q\rangle\sigma^2} \tag{15}$$

Where Q is given by Equation 16 and σ^2 by Equation 17.

$$Q = \frac{\langle q^2\rangle}{\langle q\rangle} \tag{16}$$

$$\sigma^2 = \frac{\langle q^3\rangle}{\langle q\rangle} - \frac{\langle q^2\rangle^2}{\langle q\rangle^2} \tag{17}$$

The Pearson coefficient can take values between -1 and 1, where -1 indicates a perfect negative correlation, 1 indicates a perfect positive correlation, and 0 indicates no correlation.

As in a *Erdős-Rényi* graph the edges are formed randomly, without any preference or bias, there is no inherent correlation between the degrees of nodes, which leads to a Pearson coefficient equal to 0. With our algorithm we got $\rho = -0.00036653$, which is very close to the theoretical value. As in the previous tasks, the value obtained represent the average of this value across the m simulations.

Assumptions on the network

Based on the degree distribution (Poisson distribution with mean value of 50), the clustering coefficient (0.005002), and the Pearson coefficient (-0.00036653), we can make assumptions about our network.

As seen in **Task 1**, only when the number of nodes tends to infinity, the degree distribution of the simulated network tends to a Poisson distribution, which aligns with what happens in *Erdős-Rényi* networks.

A clustering coefficient of 0.005002 indicates that the network has relatively low local clustering. This suggests that the nodes in the network have a lower tendency to form tightly connected groups or clusters, which could be indicative of a network where connections are distributed more randomly rather than forming cohesive local communities.

Finally, a Pearson coefficient close to zero (-0.00036653) indicates that there is no strong tendency for nodes with similar or different degrees to be preferentially connected. Due to this, the network can be considered approximately uncorrelated in terms of degree-degree relationships.

Therefore, based on the low clustering coefficient, that indicates a network with relatively low local clustering or community formation, and in the Pearson coefficient close to zero, that suggests an approximately uncorrelated network in terms of degree-degree relationships, we can conclude that the simulated network exhibits characteristics of an uncorrelated random graph. These findings, along with the degree distribution following a Poisson distribution for large networks, align with the fundamental properties of *Erdős-Rényi* networks, mentioned throughout this project.

Conclusion

In Task 1, we calculated the mean degree $\langle q \rangle$, the branching coefficient B, and the second and third moments of the degree distribution. The numerical results closely matched the theoretical values, indicating that the implementation of the *Erdős-Rényi* network was accurate.

In Task 2, we explored the number of triangles and the clustering coefficient of the network. The clustering coefficient was found to be approximately equal to $\langle q \rangle / N$. This suggests a low level of clustering in the simulated *Erdős-Rényi* network, indicating that the connections in the network are distributed more randomly.

As for Task 3, we calculated the Pearson coefficient, which measures the correlation between node degrees in the network. The Pearson coefficient was close to zero, indicating a weak/non-existing correlation between node degrees.

Finally, the analysis of the network properties (degree distribution, clustering coefficient and Pearson coefficient) revealed that our simulated *Erdős-Rényi* network exhibits characteristics of a uncorrelated random graph, as expected as it aligns with the fundamental properties of the type of networks.

Bibliography

[1] Erdős, P. & Rényi, A. (1959). On random graphs. Publicationes Mathematicae Debrecen, 6(290-297), 17.

[2] Erdös P, Rényi A. On the evolution of random graphs. In: The Structure and Dynamics of Networks. Princeton: Princeton University Press; 2006. p.38-82. https://doi.org/10.1515/9781400841356.38

[3] Bollobás, B. (2001). Random graphs (Vol. 73). Cambridge University Press.

[4] Americo, Rui. "Lecture 2", Modelação de Sistemas Complexos 2022/23.

[5] Americo, Rui. "Lecture 12", Modelação de Sistemas Complexos 2022/23.

[6] Americo, Rui. "Lecture 13", Modelação de Sistemas Complexos 2022/23.

[7] Dorogovtsev, Sergey N., and Jose FF Mendes. The Nature of Complex Networks. Oxford University Press, 2022.

[8] MathWorks. rng - Control random number generator, (2021) https://www.mathworks.com/help/matlab/ref/rng.html.

[9] NetworkX Examples - Erdos Renyi, https://networkx.org/documentation/networkx-2.4/auto_examples/graph/plot_erdos_renyi.html

Appendix

Algorithm

1. Initialize a NxN matrix A with all entries set to 0.
2. For each edge generate two integers i and j uniformly at random between 1 and N.
3. If $i = j$ or $A_{ij} = 1$ repeat step 2, otherwise proceed to the next step.
4. Update matrix A_{ij} with the new edge by setting $A_{ij} = A_{ji} = 1$.
5. Repeat steps 2-4 L times, with $L = \frac{N}{2}c$.
6. Calculate degrees, $q_i = \sum_{j=1}^{N} A_{ij}$.
7. Calculate the number of vertices with degree q, i.e., N(q), and then find degree distribution $P(q)$, $P(q) = \frac{N(q)}{N}$.
8. Repeat steps 1-7 m times and average the degree distribution $P(q)$ over the m realizations of the network.
9. Calculate the mean degree $\langle q \rangle = \sum_{q=1}^{N-1} qP(q)$.
10. Calculate the branching coefficient,

$$B = \frac{1}{\langle q \rangle} \sum_{q=0}^{N-1} P(q)q(q-1)$$

11. Calculate second and third moments, $\langle q^2 \rangle$ and $\langle q^3 \rangle$
12. Calculate the clustering coefficient $C = \frac{n_{tr}}{n_{pt}}$.