

Social Networks: Assignment #1

Due on Monday, December 15, 2025

Dr. Masoud Asadpour 15:00am

Ali Dashtbozorg 810104302

Contents

1	4
(a) Consider the small-world model of Watts and Strogatz with rewiring probability p . The model starts with a regular ring lattice where each node is connected to its $\langle k \rangle / 2$ neighbors on either side for a total degree of $\langle k \rangle$.1 Show that when $p = 0$, the overall clustering coefficient of this graph is given by:	4
(b) Show that when $p > 0$, the overall clustering coefficient is given by:	5
2	5
(a) For a very large, snobbish network where $p \gg q$, determine the minimal values for p and q (each in terms of N) required to ensure global connectivity.	5
(b) What is the expected scaling of the average shortest path length $\langle d_{same} \rangle$ between two nodes of the same color?	5
(c) What is the expected scaling of the average shortest path length $\langle d_{diff} \rangle$ between two nodes of different colors?	5
(d) Based on your answers, does this network model exhibit the small-world property?1 Justify your answer by explaining how this model does (or does not) satisfy the two key characteristics of small-world networks (path length and clustering). .	6
3	6
(a) Network Construction and Simulation	6
(b) Scaling Analysis	7
(c) Explain why the Random Network follows a logarithmic scaling while lattices follow a polynomial scaling. What structural feature of the Random Network (often referred to as "shortcuts") is responsible for this drastic reduction in path length? .	7
4	8
(a) Deterministic Construction of Scale-Free Networks	8
(b) The Random Genetic (RG) Model & Parameter Space	11
5	12
(a) Simulating Network Evolution	12
(b) Analyzing the Critical Threshold	13
(c) Finite Size Effects	14
(d) The Critical State	15

List of Figures

1	linear plot and log-log plot scaling of average path length $\langle d \rangle$ vs network size N for Lattices and Random Networks..	6
2	comparing simulation results with the theoretical predictions shown in the lecture slide.	7
3	the adjacency matrix of the network showing how rules effect the distribution of degrees.	9
4	the out degree plot for the generated network showing both the data and how a line with slope -0.95 perfectly fits the data.	10
5	the in degree plot for the generated network showing both the data and how a line with slope -0.95 perfectly fits the data.	11
6	The heatmap plot for network density for both theoretical density and simulated density which shows they are aligned,similar and both saturate	12
7	The order parameter S and The average size of isolated clusters $\langle s \rangle$ versus $\langle k \rangle$. .	13
8	S versus $\langle k \rangle$ Plot for $N = 100$, $N = 1,000$, and $N = 10,000$	14
9	$P(s)$ vs s for our data,theory and fitted line on a Log-Log scale here $P(s)$ is the pmf not pdf	15

Question 1

The Watts-Strogatz Model

(a) Consider the small-world model of Watts and Strogatz with rewiring probability p . The model starts with a regular ring lattice where each node is connected to its $\langle k \rangle / 2$ neighbors on either side for a total degree of $\langle k \rangle$. Show that when $p = 0$, the overall clustering coefficient of this graph is given by:

$$C(0) = \frac{3(k-2)}{4(k-1)} \quad (1)$$

We know that the global clustering coefficient is given by equation 2:

$$C = \frac{3 \times \text{number of triangles}}{\text{number of connected triples of vertices}} \quad (2)$$

we know that for the WS model the number of connected triples of vertices is just N times the number of triplets each node gives which is given by equation 3:

$$N \binom{\langle k \rangle}{2} \quad (3)$$

for the number of triangles for each node we need to consider 2 cases:

1. where the node has triangles with only one side of its connections so the number of triangles is given by 4

$$\binom{\frac{\langle k \rangle}{2}}{2} + \binom{\frac{\langle k \rangle}{2}}{2} \quad (4)$$

2. where each node has triangles with one node on one side and the other on another side the number of triangles is given by 5

$$0 + 1 + 2 + 3 + 4 + \dots + \frac{\langle k \rangle}{2} - 1 = \frac{(1 + \frac{\langle k \rangle}{2} - 1)(\frac{\langle k \rangle}{2} - 1)}{2} = \binom{\frac{\langle k \rangle}{2}}{2} \quad (5)$$

When we sum the results from equation 4,5 we are given 3 times the number of triangles so for the actual number of triangles we need to divide by 3 putting all of this together with equation 3 we get equation 6

$$\begin{aligned} C(0) &= \frac{3 \left(\frac{N \binom{\langle k \rangle}{2}}{3} \right)}{N \binom{\langle k \rangle}{2}} = \frac{3 \left(\frac{\binom{\langle k \rangle}{2}(\frac{\langle k \rangle}{2} - 1)}{2} \right)}{\frac{\langle k \rangle(\langle k \rangle - 1)}{2}} = \frac{3 \left(\frac{\langle k \rangle}{2} \right) \left(\frac{\langle k \rangle}{2} - 1 \right)}{\langle k \rangle(\langle k \rangle - 1)} = \frac{\frac{3}{4} \langle k \rangle (\langle k \rangle - 2)}{\langle k \rangle(\langle k \rangle - 1)} \\ &= \frac{3(\langle k \rangle - 2)}{4(\langle k \rangle - 1)} \end{aligned} \quad (6)$$

(b) Show that when $p > 0$, the overall clustering coefficient is given by:

$$C(p) \approx \frac{3(k-2)}{4(k-1)}(1-p)^3 \quad (7)$$

In this question we refer back to equation 2 since the degree of nodes isn't dependent on p the number of connected triples will not change with changing p the only thing that will change is the number of triangles for that case we can make an induction that the number of triangles we have at $p = 0$ called $t(0)$ will survive in $p = p_1$ with probability that all 3 triangles edges survive thus $t(p) = t(0)(1-p)^3$ and for the global clustering coefficient we reach equation 8:

$$C(p) \approx C(0)(1-p)^3 = \frac{3(k-2)}{4(k-1)}(1-p)^3 \quad (8)$$

Question 2

Snobbish Network

(a) For a very large, snobbish network where $p \gg q$, determine the minimal values for p and q (each in terms of N) required to ensure global connectivity.

Since $p \gg q$ we have to first connect each community individually and then make at least 1 connection between the 2 communities for the first part we know from lecture notes that $p > \frac{\ln N}{N}$ is sufficient condition for connectivity in each of the communities for there to be at least 1 connection between the 2 communities we follow equation 9

$$\begin{aligned} \text{possible edges} &= N^2 \\ p(\text{No Edges}) &= (1-q)^{N^2} = [(1-q)^{\frac{1}{q}}]^{qN^2} \xrightarrow{p \gg q} \approx [e^{-1}]^{qN^2} = e^{-qN^2} \\ qN^2 &\geq \ln N \rightarrow q \geq \frac{\ln N}{N^2}, p \geq \frac{\ln N}{N} \end{aligned} \quad (9)$$

(b) What is the expected scaling of the average shortest path length $\langle d_{\text{same}} \rangle$ between two nodes of the same color?

from the lecture notes we know that equation 10 holds:

$$\langle d_{\text{same}} \rangle = \frac{\ln N}{\ln \langle k \rangle} = \frac{\ln N}{\ln N p} = \frac{\ln N}{\ln N \frac{\ln N}{N}} = \frac{\ln N}{\ln \ln N} \quad (10)$$

(c) What is the expected scaling of the average shortest path length $\langle d_{\text{diff}} \rangle$ between two nodes of different colors?

from equation 10 we know the average distances between 2 nodes of the same color so for the distance between nodes of different colors we first need to get to bridge nodes and

from there go to the destination node the total distance is given by equation 11

$$\langle d_{diff} \rangle = \langle d_{same} \rangle + 1 + \langle d_{same} \rangle = 2\langle d_{same} \rangle + 1 = \frac{2 \ln N + \ln \ln N}{\ln \ln N} = O\left(\frac{\ln N}{\ln \ln N}\right) \quad (11)$$

(d) Based on your answers, does this network model exhibit the small-world property? Justify your answer by explaining how this model does (or does not) satisfy the two key characteristics of small-world networks (path length and clustering).

The average distance between nodes scales logarithmically so it has the short path property of small world networks but since it is also a random network it does not have a high clustering coefficient but by definition and lecture notes small world networks don't need to have a high clustering coefficient (unless we are looking for ultra small world property) so this network is a small world network given that it is connected

Question 3

Small World Phenomena in Random Networks

The code for generating the graphs and calculating the shortest path is provided in Helpers section and uses igraph package

(a) Network Construction and Simulation

in figure 1 we have visualized the results for average shortest path which correspond to our theory where 1d lattice have significantly higher shortest path compared to the rest and random networks which have small world property have lower shortest path compared to the rest

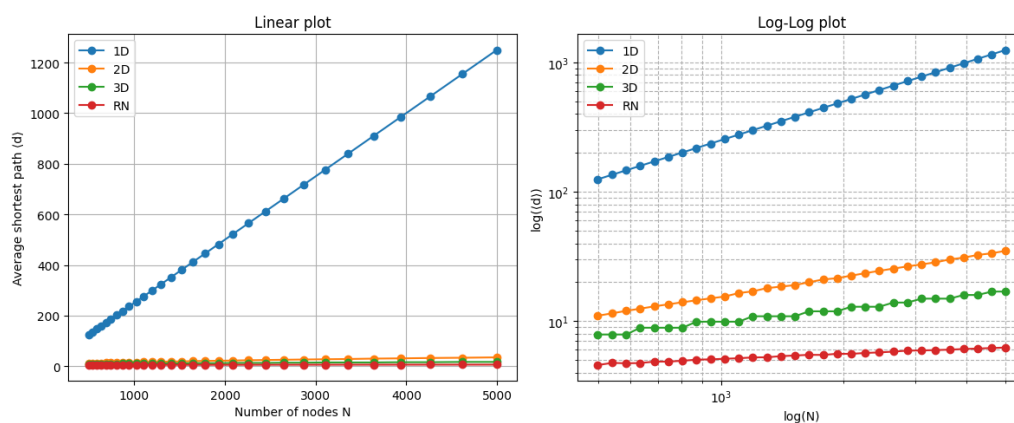


Figure 1: linear plot and log-log plot scaling of average path length $\langle d \rangle$ vs network size N for Lattices and Random Networks..

(b) Scaling Analysis

in figure 2 we computed the slope derived from our data and it almost is perfect when compared to our theory the only reason theory differs from our data is because when we are talking about scale of the shortest path we don't include coefficient behind it so actually for a 2d lattice it would be $\langle d \rangle \sim aN^{1/2}$ if we were to include a than our theory and simulations would align perfectly and as shown on the graph our simulated exponents match the theoretical dimensionality

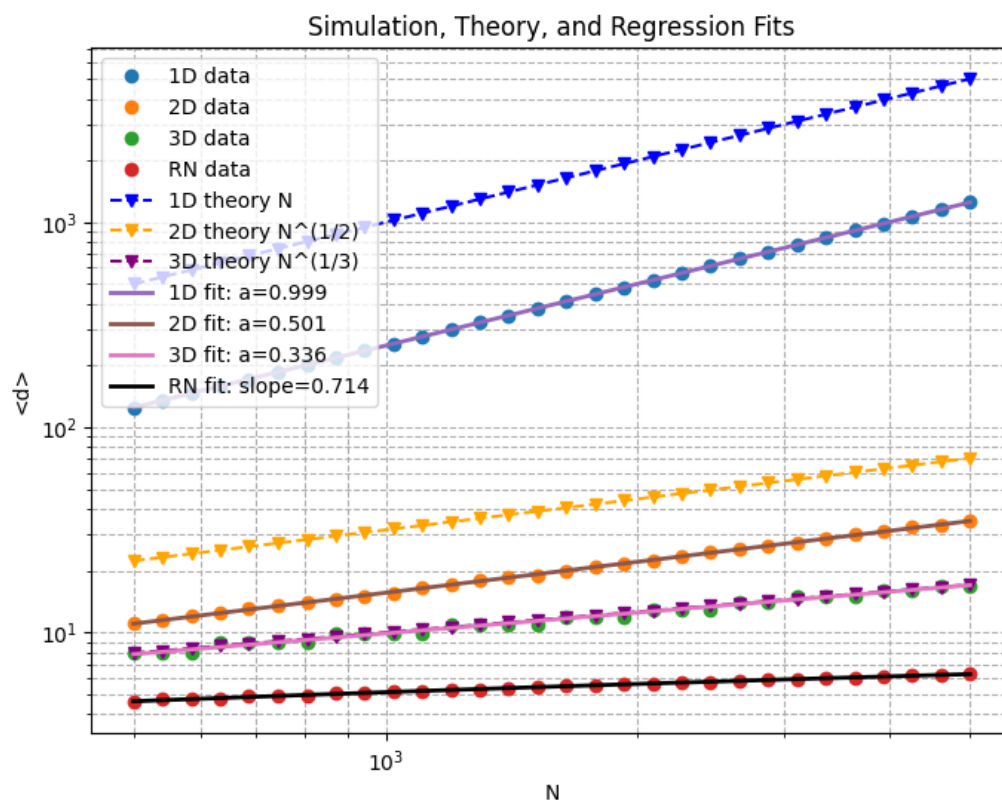


Figure 2: comparing simulation results with the theoretical predictions shown in the lecture slide.

(c) Explain why the Random Network follows a logarithmic scaling while lattices follow a polynomial scaling. What structural feature of the Random Network (often referred to as "shortcuts") is responsible for this drastic reduction in path length?

in a regular lattice nodes are connected only to nearby neighbors so reaching distant nodes requires traversing space step by step causing the typical shortest path length to grow as a power of the system size N (specifically $\ell \sim N^{1/d}$ in d dimensions) In contrast a random network contains long range connections often called shortcuts that link nodes which are far apart in the underlying structure. These shortcuts allow the number of reachable nodes to grow exponentially with the number of steps, so the entire network can be reached in a number of steps that scales logarithmically with N ($\langle d \rangle \sim \log N$). The presence of these shortcuts is therefore the key structural feature responsible for the drastic reduction in

average path length in random networks.

Question 4

Generative Graph Models & Genetic Encodings

(a) *Deterministic Construction of Scale-Free Networks*

1. Examine the Adjacency Matrix plot. Explain how the recursive, block-like structure visually represents the hierarchical nature of the hubs.

the adjacency matrix spy plot given in figure 3 for $b = 10$ reveals a highly structured non random pattern dominated by pronounced vertical stripes concentrated near the top rows which directly reflects the hierarchical hub structure induced by the genetic wiring rules. The dense vertical columns correspond to destination nodes whose binary addresses end with long suffixes of ones causing them to receive edges from exponentially many source patterns and thus become high in degree hubs. The varying stripe depths and their recursive spacing encode the successive application of wiring rules at different scales visually demonstrating how hubs emerge deterministically and hierarchically from shared genetic prefixes and suffixes rather than from random attachment.

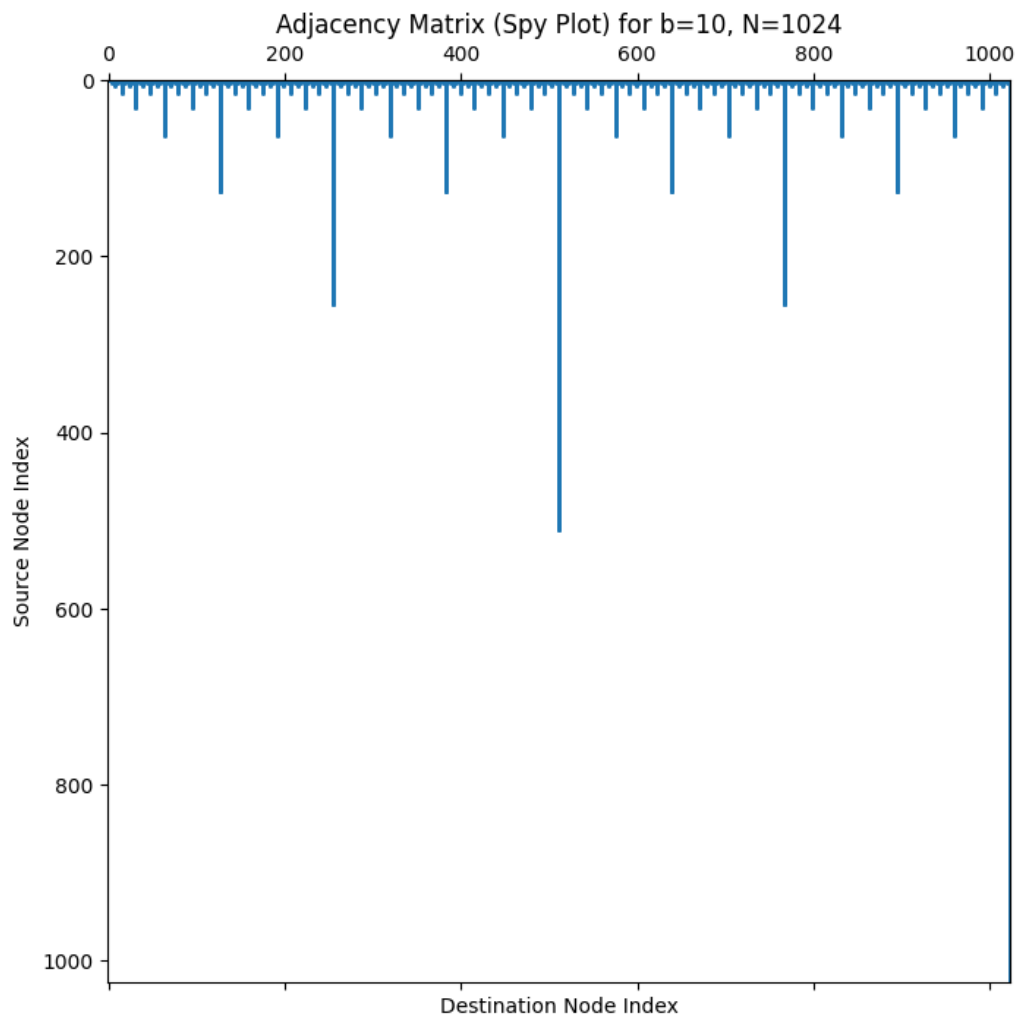


Figure 3: the adjacency matrix of the network showing how rules effect the distribution of degrees.

2. In standard Barabasi–Albert models, $\gamma \approx 3$. However, this deterministic method produces a different exponent. Does your estimated γ converge closer to $\gamma \approx 1$ (slope of -1)? Explain why this specific bit-matching heuristic creates such an extreme “winner-takes-all” topology.

in this deterministic genetic construction, the estimated power-law exponent (γ) indeed converges toward values close to $\gamma \approx 1$ (corresponding to a slope of approximately -1 on a log–log plot) rather than the $\gamma \approx 3$ characteristic of the Barabási–Albert model. This occurs because the bit-matching heuristic creates a highly uneven allocation of edges: nodes whose binary addresses end with long runs of ones match a large number of destination patterns (D_i), while nodes with many leading zeros participate in multiple source patterns (S_i). As a result a small subset of nodes is deterministically selected by many wiring rules across all hierarchical scales, accumulating edges exponentially faster than the rest. Unlike preferential attachment, where degree grows proportionally to existing degree through stochastic growth, this model encodes preference directly into the address space, effectively collapsing attachment probability onto a few genetically privileged nodes. This produces an ex-

treme “winner-takes-all” topology, where hubs dominate the network and the degree distribution becomes much heavier-tailed, driving the exponent toward $\gamma \approx 1$. Empirical evidence for this behavior is shown in two loglog plots. Figure 4 presents the out-degree distribution together with its fitted line, while Figure 5 shows the corresponding in-degree distribution and fit. In both cases, the fitted slope is $\gamma = -0.95$, and the linear fit is effectively perfect, confirming that both the in-degree and out-degree distributions follow the same heavy-tailed scaling with an exponent very close to -1 .

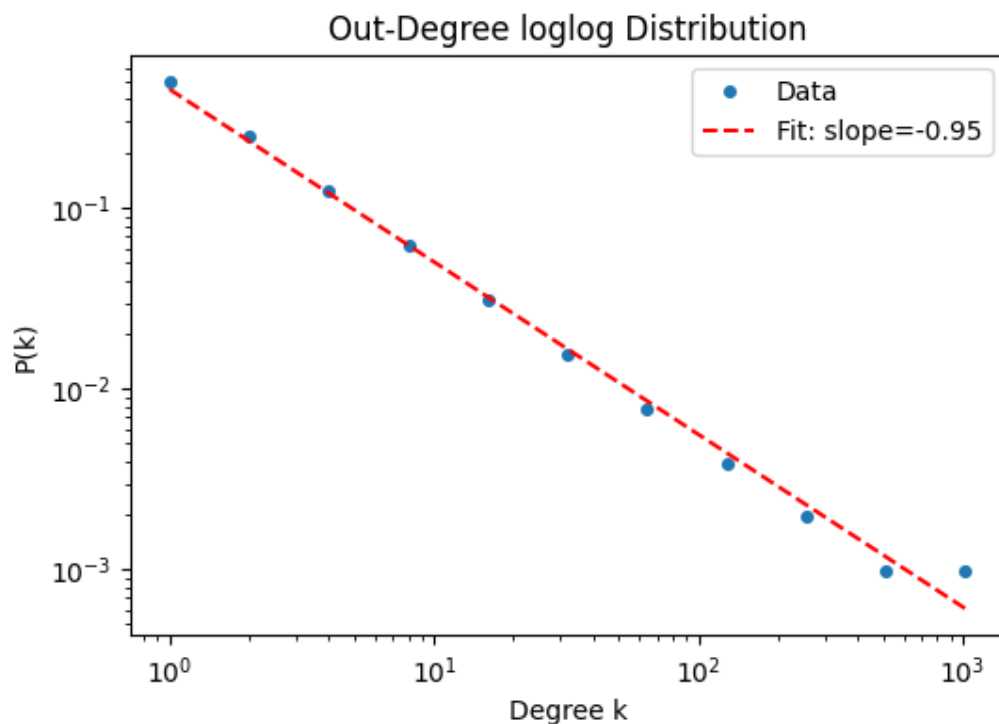


Figure 4: the out degree plot for the generated network showing both the data and how a line with slope -0.95 perfectly fits the data.

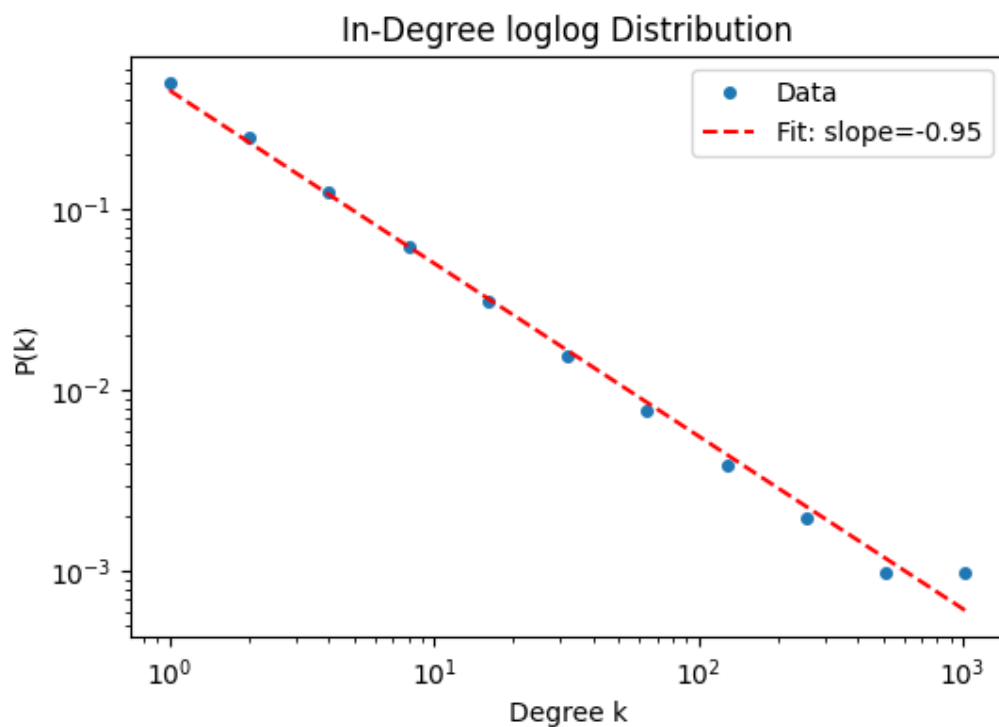


Figure 5: the in degree plot for the generated network showing both the data and how a line with slope -0.95 perfectly fits the data.

(b) *The Random Genetic (RG) Model & Parameter Space*

1. The theoretical density is given by $\rho = 1 - (1 - \pi^2)^{2r}$, with $\pi = 2^{x-b}$. Compare your simulation results with this formula. Does the density grow linearly with r , or does it saturate?

The simulation results closely follow the theoretical prediction

$$\rho = 1 - (1 - \pi^2)^{2r}, \quad \text{with } \pi = 2^{x-b},$$

confirming that edge formation in the RG model is well captured by independent rule coverage probabilities. This agreement is visually reinforced by the heatmap plot in Figure 6, which compares simulated versus theoretical densities and shows that the two are almost identical and exhibit the same behavior. For small r , the density initially grows approximately linearly, since each newly added rule is unlikely to overlap with previous ones and thus contributes many new edges. However, as r increases, the growth becomes sublinear and eventually saturates toward $\rho \rightarrow 1$. This saturation occurs because the probability that a potential edge has already been created by at least one previous rule increases rapidly with r , especially for larger x where each rule already covers a large fraction of the node space. Therefore, the density does not grow indefinitely with the number of rules but instead exhibits a clear saturation regime, in strong agreement with the theoretical formula.

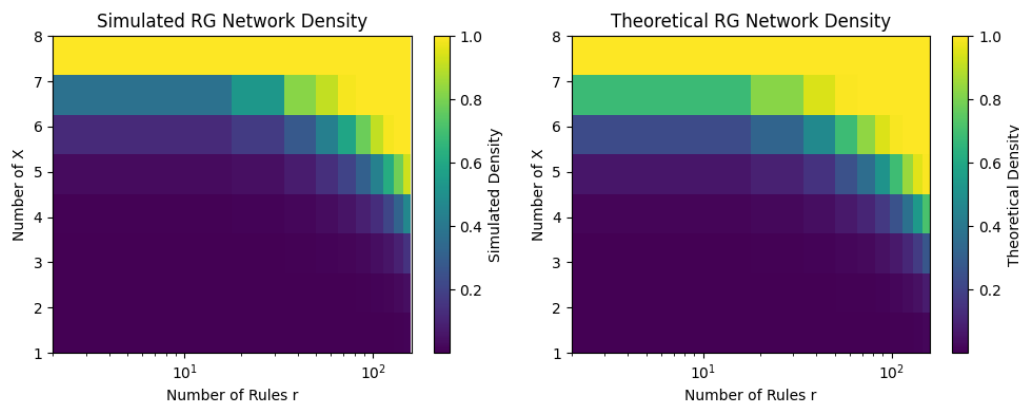


Figure 6: The heatmap plot for network density for both theoretical density and simulated density which shows they are aligned, similar and both saturate

2. Explain the concept of "Rule Overlap." Why does adding more rules eventually yield diminishing returns in terms of new edges created?

Rule overlap refers to the phenomenon where multiple genetic wiring rules generate the same edges between node pairs as the number of rules increases, especially when the number of wildcards x is large, each rule matches a substantial fraction of nodes, making overlaps increasingly likely. Consequently, newly added rules tend to recreate edges that already exist rather than introduce novel connections. This leads to diminishing returns although rules are added linearly the number of new edges added per rule decreases rapidly. In the extreme case once most node pairs are already connected, additional rules contribute almost no new edges at all. Rule overlap is therefore the fundamental mechanism driving density saturation in the RG model and explains why increasing r beyond a certain point no longer significantly alters the network structure.

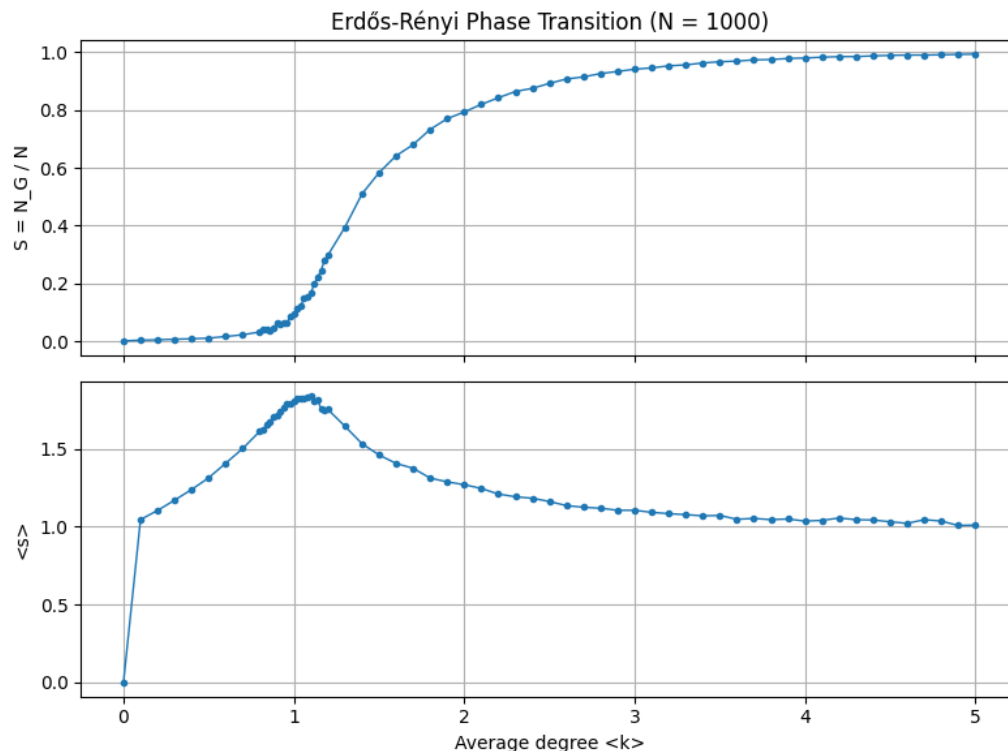
Question 5

The Evolution of Random Networks & Phase Transitions

(a) Simulating Network Evolution

The code is provided in q3.ipynb file the logic is implemented in the general helpers section since i don't want this report to get too long i will skip the explanation

(b) Analyzing the Critical Threshold

Figure 7: The order parameter S and The average size of isolated clusters $\langle s \rangle$ versus $\langle k \rangle$

1. Identify the critical point $\langle k \rangle_c$ from your simulations. Does the transition occur at the theoretical prediction $\langle k \rangle = 1$? Explain any deviations (consider finite-size effects).

From the simulations shown in Figure 7, the critical point $\langle k \rangle_c$ can be identified as the value of $\langle k \rangle$ where the order parameter $S = N_G/N$ starts rising sharply and the average size of small clusters $\langle s \rangle$ reaches its maximum which occurs around $\langle k \rangle_c \approx 1$ in a finite network of $N = 1000$ nodes the transition does not occur as abruptly as the theoretical prediction for an infinite network and small deviations from $\langle k \rangle = 1$ are observed. These deviations arise due to finite size effects, stochastic fluctuations in edge formation, limited total nodes, and the smoothing of the phase transition cause the giant component to emerge slightly earlier or later than the ideal threshold, and the peak of $\langle s \rangle$ is less pronounced than in an infinite system.

2. Examine the plot of $\langle s \rangle$. Why does the average cluster size diverge (peak) near the critical point and then decrease in the supercritical regime ($\langle k \rangle > 1$)? Explain the physical meaning of this peak.

The average cluster size $\langle s \rangle$ diverges and reaches a peak near the critical point because at $\langle k \rangle \approx \langle k \rangle_c$ many small clusters begin to merge forming larger clusters without yet creating a single giant component. This results in a maximal average size of the remaining small clusters. Beyond the critical point ($\langle k \rangle > 1$) a giant component dominates the network, absorbing most nodes, and the remaining clusters are small.

and isolated causing $\langle s \rangle$ to decrease physically the peak in $\langle s \rangle$ reflects the network's transition from a fragmented state to a cohesive system: it signals the point where medium-sized clusters are most abundant just before coalescing into the giant component. This result is also evident in Figure 7

(c) Finite Size Effects

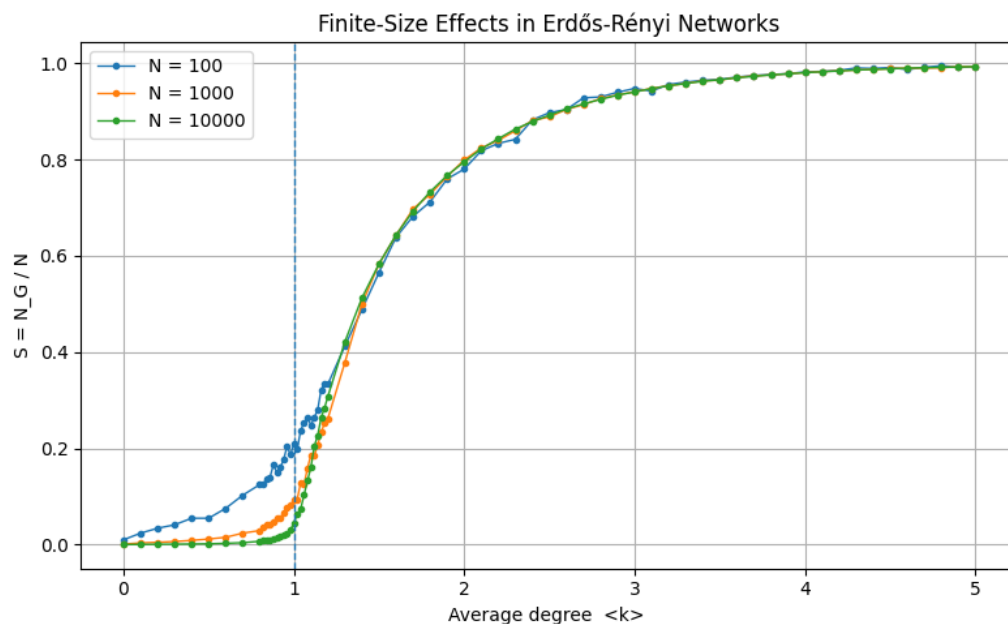


Figure 8: S versus $\langle k \rangle$ Plot for $N = 100$, $N = 1,000$, and $N = 10,000$

1. Observe the slope of the curve near $\langle k \rangle = 1$. How does the "sharpness" of the transition change as N increases?

From the Figure 8, it is evident that the sharpness of the phase transition increases with network size. For the smallest network ($N = 100$), the rise of S near $\langle k \rangle = 1$ is gradual, indicating a smooth transition. As N increases to 1000 and 10,000, the slope of the curve becomes steeper, approaching a nearly vertical jump for the largest network. This behavior illustrates the thermodynamic principle that phase transitions become sharper in the limit $N \rightarrow \infty$, while finite networks exhibit smoothed transitions due to stochastic fluctuations in edge formation.

2. At $\langle k \rangle = 1$, does the relative size of the giant component S tend towards zero or a finite value as N increases? Relate this to the theoretical prediction $N_G \approx N^{2/3}$ at criticality

At the critical point $\langle k \rangle = 1$ as shown in Figure 8, the relative size of the giant component $S = N_G/N$ decreases with increasing network size. This is consistent with the theoretical prediction that the size of the largest component at criticality scales as $N_G \sim N^{2/3}$, so the relative size $S \sim N^{-1/3}$ tends toward zero as N grows. In smaller networks, the giant component occupies a noticeable fraction of the nodes

at $\langle k \rangle = 1$, whereas in very large networks, it contains many nodes but represents only a vanishing fraction of the total network, reflecting the finite-size effects at criticality.

(d) The Critical State

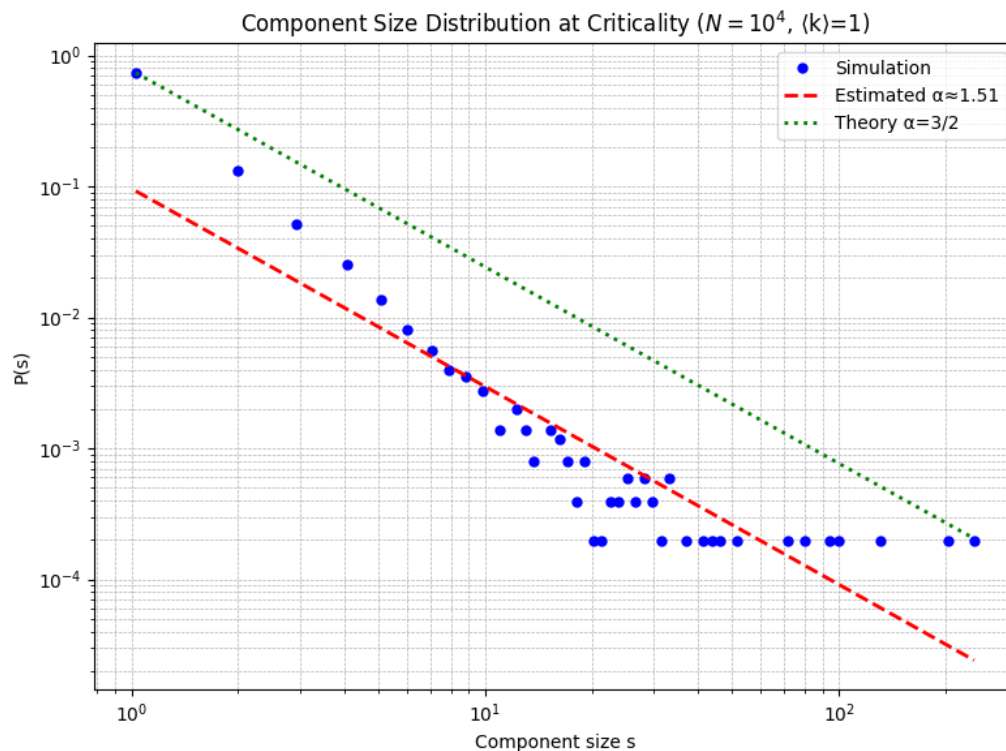


Figure 9: $P(s)$ vs s for our data, theory and fitted line on a Log-Log scale here $P(s)$ is the pmf not pdf

1. Does the component size distribution follow a power law $P(s) \approx s^{-\alpha}$? If so, estimate the exponent α .

The component size distribution $P(s)$ for a large Erdos Renyi network at the critical point $\langle k \rangle = 1$ exhibits a clear power law behavior, as seen in the log-log plot of Figure 9. Small clusters are most frequent, while larger clusters occur less often, forming a straight-line tail in log-log space. Using logarithmic binning to reduce noise, a linear fit to the tail of the distribution yields an estimated exponent $\alpha \approx 1.51$, indicating that $P(s) \sim s^{-\alpha}$ over several orders of magnitude.

2. Compare your estimated exponent with the theoretical prediction for random graphs at criticality ($\alpha = \frac{3}{2}$).

The estimated exponent $\alpha \approx 1.51$ is in excellent agreement with the theoretical prediction for random graphs at criticality $\alpha = 3/2$. This confirms that at the critical point the network is composed of trees of various sizes following a scale free distribution.

bution. The power law tail reflects the high variability of cluster sizes where medium and large clusters although less frequent, still contribute significantly to the network's structure, highlighting the unique critical properties of the Erdos-Renyi network.

References

- [1] ChatGPT, *Chat session used for assignment guidance*, <https://chatgpt.com/share/69409aea-8c1c-800d-af8b-10e54b089cb6>, accessed December 16, 2025.