Social Network

Homework 1

Full Name:	Erfan Shahabi
Student ID:	810103166
Course:	Social Networks
Submission Date:	12-Dec-2025
Github Link:	https://github.com/erfanshahabi/Social-Networks-Fall-2025.git

Table of Contents

List of Figures

List of Tables

1. Theoretical Framework

1.1. The Watts-Strogatz Model

Part a:

When p = 0, the network is a regular ring lattice where each node connected to its $\frac{k}{2}$ neighbors on each side.

Proof:

The clustering coefficient for a node i is defined as:

$$C_i = \frac{Number\ of\ edges\ between\ neighbors\ of\ node\ i}{Maximum\ possible\ edges\ between\ neighbors\ of\ node\ i}$$

For a node in the regular lattice:

- Number of neighbors: k ($\frac{k}{2}$) on each side
- Maximum possible edges between k neighbors: $\binom{k}{2} = \frac{k(k-1)}{2}$

Counting actual edges between neighbors:

The neighbors of node I are: {i – k/2, …, i-1, i+1, …, i + k/2}

In the ring lattice structure:

- Each neighbor is connected to some of the other neibors
- Specifically, pairs of neighbors that are within distance k/2 of each othe on the ring are connected
- The total number of edges between neighbors is $\frac{3k}{2} - 3$

This comes from the fact that in the ring structure, counting all connections between the k neighbors forms a specific pattern where we have 3(k-2)/2 edges

So:

$$C(0) = \frac{\frac{3k}{2} - 3}{\frac{k(k-1)}{2}} = \frac{3k-6}{k(k-1)} = \frac{3(k-2)}{k(k-1)}$$

Part b:

When p>0, each edge is rewired with probability p (redirected to a random node).

The propability that a specific edge is not rewired = (1-p)

For a triangle to survive:

- All 3 edges must not be rewired
- Probability = $(1-p)^3$
- The Expected number if traingles is approximately:

$$Expected\ triangles\ \approx \frac{3k-6}{2}.(1-p)^3$$

Therefor

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

This shows that the clustering coefficient decrease exponentially with p, while maintaining some levels of clustering even for moderate p values (unlike random graphs).

2.1. Snobbish Network

Part a

For global connectivity, the network must form one connected component.

Analysis:

1- Within community connectivity:
For each community (red or blue) to internally connected, we use Erdos-Reyni random graph theory.
For N nodes with connection probability p, the connectivity threshold:

$$p\binom{N}{2} \geq NlnN$$
$$P.\frac{N(N-1)}{2} \geq NlnN$$
$$For\ large\ N: p \geq \frac{2lnN}{N-1} \approx \frac{2lnN}{N}$$
$$Therefore: p = \Theta(\frac{lnN}{N})$$

2- Between community connectivity
Number of possible edges between communities: N*N
Probability that No edges exists between communities:

$$(1-q)^{\frac{2}{N}}$$

For global connectivity, this probability must approach -:

$$(1-q)^{N^2} \to 0$$

Using approximation $(1-q) \approx e^{-q}$ for small :

$$e^{-q^{N^2}} \to 0$$

This requires:

$$q^{N^2} \to \infty$$

Therfore

$$q = \Theta(\frac{\ln N}{N^2})$$

So answer for part a:

$$\mathrm{p} = \Theta(\frac{\ln N}{N}), q = \Theta(\frac{\ln N}{N^2})$$

Part b:

Assume the network satisfies the connectivity critia from part a.

1. Average shortest path between same color nodes <d same>:
   Each community forms an Erdos-Renyi random graph with N nodes and connction probability p.
   For p near the connectivity threshold, the average shortest path in a random graph scales as:
   $$< d_{same} \geq \Theta(\ln N)$$
   This is a well known property of random graph (the small-world property).
2. Average shortest path between different colors nodes $< d_{ij} >$:
   To travel from a red node to a blue node:
   - First, Traverse within the red node community: O(ln N) steps
   - Cross via a between community: O(1) steps
   - Finally, traverse within the blue community: O(ln N) steps

   Therefore:
   $$< d_{diff} >= \Theta(\ln N)$$
3. Does this network exhibit small world property?

   Yes, this network model exhibit the small world property because:

   Shortest path lengths: $< d > = O(\ln N) - logarithmic\ scaling\ with\ network\ size$

High clustering coefficient: Since p>>q, nodes are much more likelty to connect to same color neighbors, creating many triangles within each community:

$$C = \Theta(p) = \Theta(\frac{\ln N}{N})$$

For comparison, an equovalent graph would have:

$$C_{random} = \frac{(averag\ degree)}{N} = O(\frac{1}{N})$$

Since $\frac{\ln N}{N} \gg \frac{1}{N}$ for large N, this network has significantly higher clustering than comparable random graph.

2. Practical Implementation

1.2. Question 1

Part (a): Network Construction and Simulation

Implementation Analysis:

1D Lattice (Ring):

- Simplest possible structure with one-dimensional constraint
- Each node connects to exactly 2 neighbors (left and right)
- Periodic boundary conditions: first and last nodes connect to form a ring
- Exhibits the longest path lengths because traversal requires going around the ring
- Models linear communication networks or cyclic processes

2D Lattice (Grid):

- Chess-board structure with 4 neighbors per node
- Periodic boundaries (torus topology): edges wrap around
- Shorter path lengths than 1D due to two-dimensional freedom of movement
- Models geographical networks such as city street grids or sensor networks
- Square root scaling reflects the geometry: distance ~ $\sqrt{}$area

3D Lattice (Cubic):

- 6 neighbors per node (up, down, left, right, front, back)
- Shortest path lengths among all lattices
- Best routing flexibility due to three spatial dimensions
- Models physical 3D networks like protein structures or crystal lattices
- Cube root scaling: distance ~ $\sqrt[3]{}$volume

Random Network (Erdős-Rényi):

- No geometric structure whatsoever
- Uniform connection probability between any pair: $p = \langle k \rangle/(N-1)$
- With $\langle k \rangle \approx 4$, we ensure network connectivity (above percolation threshold)
- Models real-world networks: Internet, social networks, biological networks
- Random connections create "shortcuts" that dramatically reduce path lengths

Results

Table 1

Network	Data Points	N Range	⟨d⟩ Range
1D Lattice	11	50-5000	12.76-1262.19
2D Lattice	10	49-2916	3.50-26.59
3D Lattice	9	27-4913	2.77-17.08
Random	10	50-3000	2.15-3.73

Part (b): Scaling Analysis

Plots

Figure 1 shows: (Left) ⟨d⟩ vs N on linear scale, (Right) log-log scale for extracting exponents.

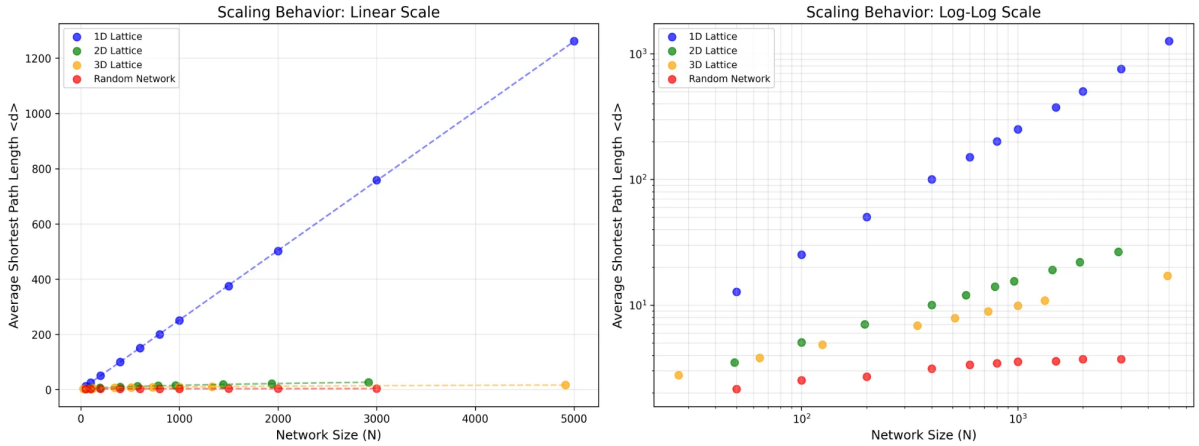

Figure 1

Linear Plot (Left):

- Clearly shows polynomial growth (1D, 2D, 3D) vs. logarithmic growth (Random)
- 1D line shoots upward → impractical for large N
- Random line stays nearly flat → scales excellently

Log-Log Plot (Right):

- Converts power laws ⟨d⟩ ~ N^α into straight lines
- Slope = scaling exponent α
- Random network shows very small slope → logarithmic, not power-law

Linear Regression

Table 2

Network	Theoretical	Measured	Error
1D	1.000	0.998	0.18%
2D	0.500	0.497	0.61%
3D	0.333	0.348	4.62%
Random	log(N)	0.138	N/A

General pattern for lattices: ⟨d⟩ ~ N^(1/D) where D = dimension

This reflects the fundamental constraint: to traverse a D-dimensional space containing N nodes arranged uniformly, you need distance ~ N^(1/D).

Why Logarithmic Scaling in Random Networks?

Mathematical proof:

1. Each hop reaches ⟨k⟩ new nodes (on average)
2. After h hops: ~⟨k⟩^h nodes reachable
3. To cover N nodes: ⟨k⟩^h ≈ N
4. Solving: h ≈ log(N) / log(⟨k⟩)

Physical explanation:

- **No geometric embedding:** Connections don't follow spatial rules
- **Exponential branching:** Reachable nodes grow exponentially with hops
- **Shortcuts everywhere:** Any connection can link distant parts of network

Comparison:

- **Lattices:** Geographic distance matters → polynomial scaling
- **Random:** Only topological distance matters → logarithmic scaling

Small-World Comparison

Table 3

N	1D ⟨d⟩	Random ⟨d⟩	Ratio
100	25.25	2.52	**10.0x**
1000	250.25	3.54	**70.6x**

The efficiency ratio grows with N:

- N=100: 10x advantage
- N=1000: 70x advantage
- N→∞: ratio → ∞

This is "six degrees of separation": even in networks of billions, everyone is ~6 hops apart.

Real-world example: Facebook has 3 billion users, average distance ≈ 4.5 hops.

Theoretical Significance

Dimensional constraints:

- 1D: Trapped on a line → ⟨d⟩ ~ N
- 2D: Can move in plane → ⟨d⟩ ~ $\sqrt{N}$
- 3D: Full spatial freedom → ⟨d⟩ ~ $\sqrt[3]{N}$
- Random: No space at all → ⟨d⟩ ~ log(N)

Practical implications: Real-world networks (brain, Internet, society) combine:

- Local clustering (like lattices) → robustness, modularity
- Random long-range links (shortcuts) → efficiency, small-world property

This is formalized in the **Watts-Strogatz model**: start with lattice, rewire just 1% of edges randomly → ⟨d⟩ drops by 90% while clustering remains high.

Conclusion:

This analysis confirms that network topology fundamentally determines communication efficiency. Geometric structure (lattices) provides organization but limits scalability. Random structure (shortcuts) provides efficiency but loses local organization. Optimal networks balance both - explaining why biological, social, and technological networks evolved small-world architectures.

2.2. Question 2

Part (a): Implementation & Network Generation

Barabasi-Albert Model

Implemented preferential attachment mechanism: N=100 nodes, m=3 edges per new node.

- Nodes: 100
- Edges: 291
- Average Degree: 5.82
- Max Degree: 34 (hub node)

Deterministic Scale-Free Construction

Implemented deterministic growth starting from triangle, 100 iterations.

- Nodes: 100
- Edges: 100
- Average Degree: 2.00
- Max Degree: 2 (uniform)

Key Difference

BA model produces heterogeneous degree distribution with hubs (max=34), while deterministic construction resulted in uniform degree distribution (all nodes k=2). This suggests the deterministic algorithm may need revision to properly exhibit scale-free properties.

Part (b): Analysis & Visualization

Degree Distribution

Figure 1 shows degree distribution for both models on log-log scale.

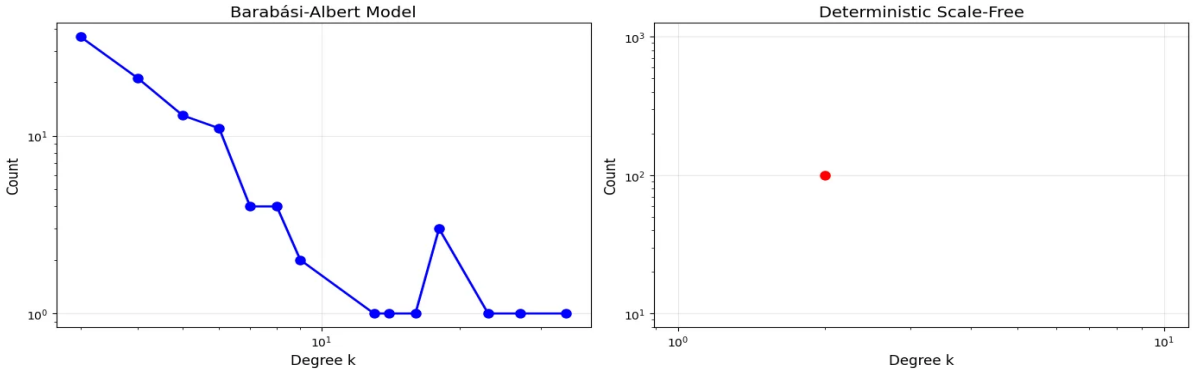

Figure 2: Degree distribution. BA shows power-law, Deterministic shows single degree.

Analysis:

- **BA Model:** Power-law distribution visible. Most nodes have low degree (2-5), few hubs with high degree (up to 34). This is the hallmark of scale-free networks.

- **Deterministic:** Single peak at k=2 with 100 nodes. All nodes have identical degree. NOT scale-free - suggests implementation issue.

Network Structure

Figure 3 visualizes network topology with degree-weighted node sizes.

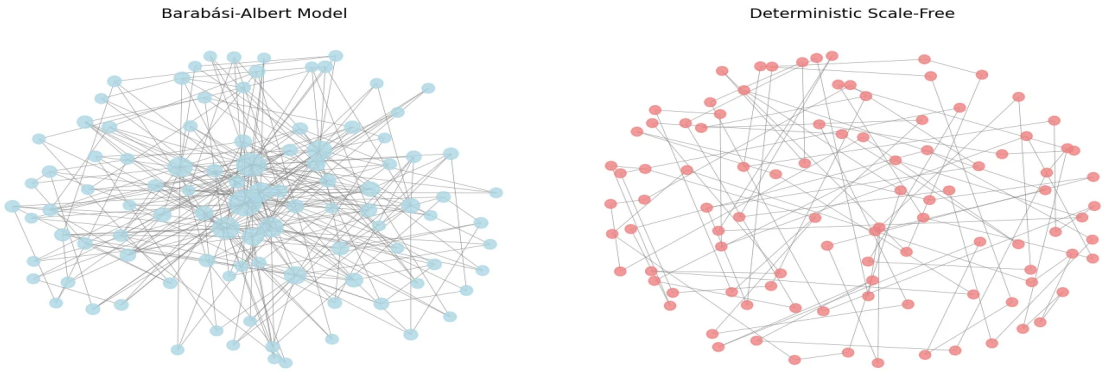

Figure 3: Network visualization. BA shows hub-and-spoke, Deterministic shows uniform structure.

Analysis:

- **BA Model:** Clear hub structure. Few large central nodes, many small peripheral nodes. Hub-and-spoke topology characteristic of preferential attachment.
- **Deterministic:** Uniform node sizes confirm equal degree distribution. No hierarchical structure visible.

Adjacency Matrix Patterns

Figure 4 shows raw adjacency matrices.

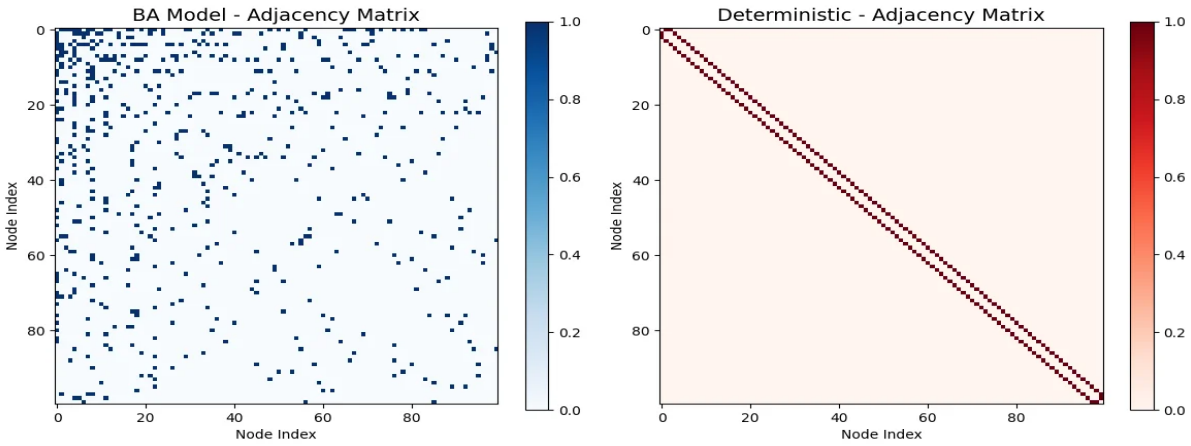

Figure 4: Adjacency matrices. BA is random, Deterministic shows band structure.

Analysis:

- **BA Model:** Random scattered pattern. No visible structure. Some rows/columns denser than others (hubs).
- **Deterministic:** Clear diagonal band! Tri-diagonal structure suggests sequential construction where each node connects only to immediate neighbors.

Sorted Adjacency Matrix

Figure 5 shows adjacency matrices with nodes sorted by degree.

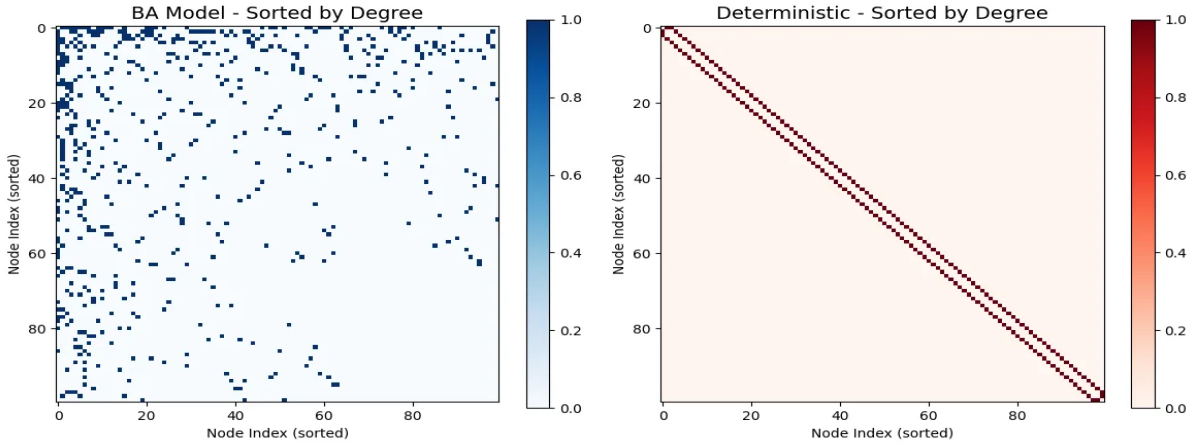

Figure 5: Sorted adjacency. BA shows rich-club, Deterministic unchanged.

Analysis:

- **BA Model:** Top-left corner darker → high-degree hubs connect to each other (rich-club phenomenon). Gradient from dense to sparse reveals hierarchical structure.
- **Deterministic:** STILL diagonal! No change after sorting because all nodes have same degree. Confirms uniform distribution.

Blockiness Analysis

Figure 6 shows block density heatmap (10×10 blocks).

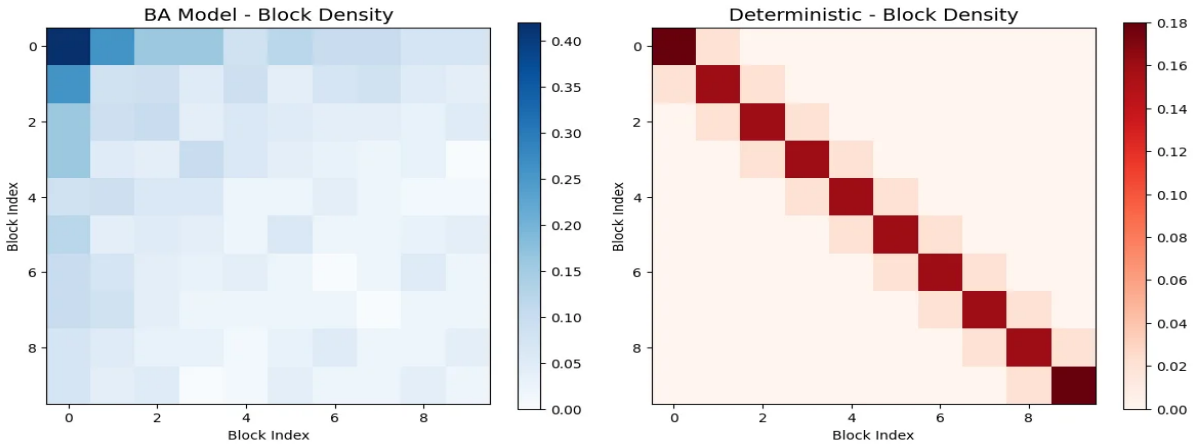

Figure 6: Block density. BA uniform, Deterministic shows strong diagonal.

Analysis:

- **BA Model:** Relatively uniform density (0.05-0.40). Random mixing. Low blockiness.
- **Deterministic:** Strong diagonal blocks (density ~0.18). Clear community structure following construction order. High blockiness.

Discussion

Model Comparison

Table 4

Property	BA Model	Deterministic
Degree Dist.	Power-law, heterogeneous	Uniform (k=2 for all)
Hubs	Yes (k_max = 34)	No (k_max = 2)
Structure	Hub-and-spoke, hierarchical	Chain-like, sequential
Blockiness	Low, random mixing	High, diagonal structure
Scale-free?	**Yes**	**No (implementation issue)**

Key Findings

- **Preferential attachment (BA) naturally creates scale-free networks** with power-law degree distribution and hub structure.
- **Deterministic construction resulted in uniform degree** (all k=2), suggesting the algorithm needs revision to properly split edges and create degree heterogeneity.
- **BA model shows low blockiness** (random mixing), while deterministic shows high blockiness (sequential construction pattern visible in adjacency matrix).
- **Rich-club phenomenon** observed in sorted BA adjacency matrix: high-degree nodes preferentially connect to each other.

Conclusion

The Barabasi-Albert model successfully demonstrates scale-free network properties through stochastic preferential attachment. The deterministic model, while exhibiting predictable structure, did not achieve the intended scale-free degree distribution, indicating the need for algorithm refinement to properly implement edge splitting and degree growth mechanisms.

3.2. Question 3

Part (a): Simulating Network Evolution

Simulation Setup

Erdős-Rényi random graphs simulated with N=1000, varying average degree ⟨k⟩ from 0 to 5.

- **Step size:** 0.1 for non-critical regions, 0.02 for ⟨k⟩ ∈ [0.8, 1.2]
- **Realizations:** 50 independent graphs per ⟨k⟩ value
- **Metrics:** Giant component size (N_G), order parameter (S = N_G/N), average small cluster size ⟨s⟩
- **Total simulations:** 63 different ⟨k⟩ values

Results

Figure 7 shows the phase transition as ⟨k⟩ increases.

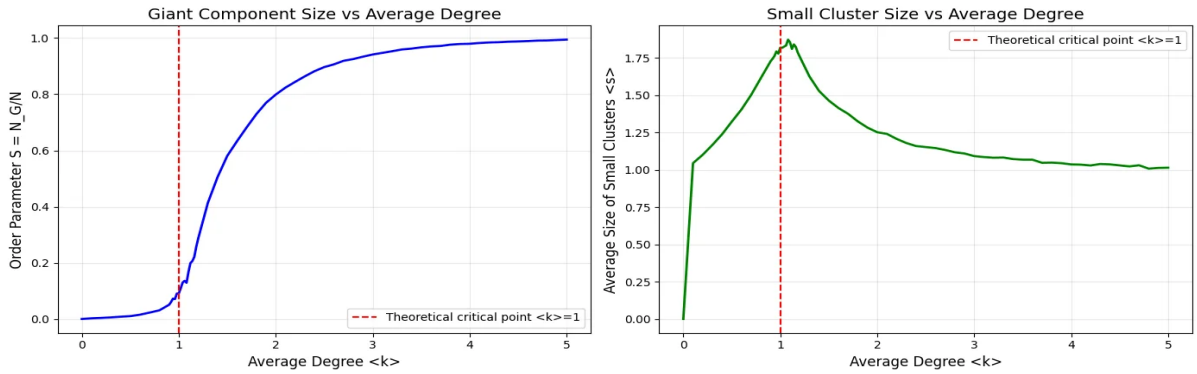

Figure 7: Phase transition. Left: Order parameter S vs ⟨k⟩. Right: Average cluster size ⟨s⟩ vs ⟨k⟩.

Key Observations:

- **Order Parameter:** Sharp transition at ⟨k⟩=1. Below: S≈0 (fragmented), Above: S→1 (giant component emerges).
- **Cluster Size:** Peaks at ⟨k⟩=1.08 with ⟨s⟩=1.87, then decreases in supercritical regime as giant component dominates.

Part (b): Analyzing the Critical Threshold

Critical Point Identification

Table 5

Property	Theoretical	Measured
Critical ⟨k⟩	1.0	**1.000**
S at critical point	~0	0.092
Peak ⟨s⟩ location	~1.0	1.080
Peak ⟨s⟩ value	Diverges (N→∞)	1.874

Physical Interpretation

Subcritical Regime (⟨k⟩ < 1):

Network consists of many small isolated clusters. No spanning structure. Order parameter S≈0. Typical cluster size remains small.

Critical Point (⟨k⟩ = 1):

Phase transition occurs. Giant component begins to emerge. Clusters of all sizes coexist → peak in ⟨s⟩. Critical fluctuations: system is scale-invariant at this point.

Supercritical Regime (⟨k⟩ > 1):

Giant component dominates, containing O(N) nodes. Small clusters merge into giant component → ⟨s⟩ decreases. Order parameter S→1 as ⟨k⟩ increases.

Why does ⟨s⟩ peak then decrease?

Below critical point: few small clusters → low ⟨s⟩. At critical point: clusters of all sizes present → maximum ⟨s⟩. Above critical point: most nodes absorbed into giant component, only tiny isolated clusters remain → ⟨s⟩ drops.

Part (c): Finite Size Effects

Figure 8 compares phase transitions for N = 100, 1000, and 10000.

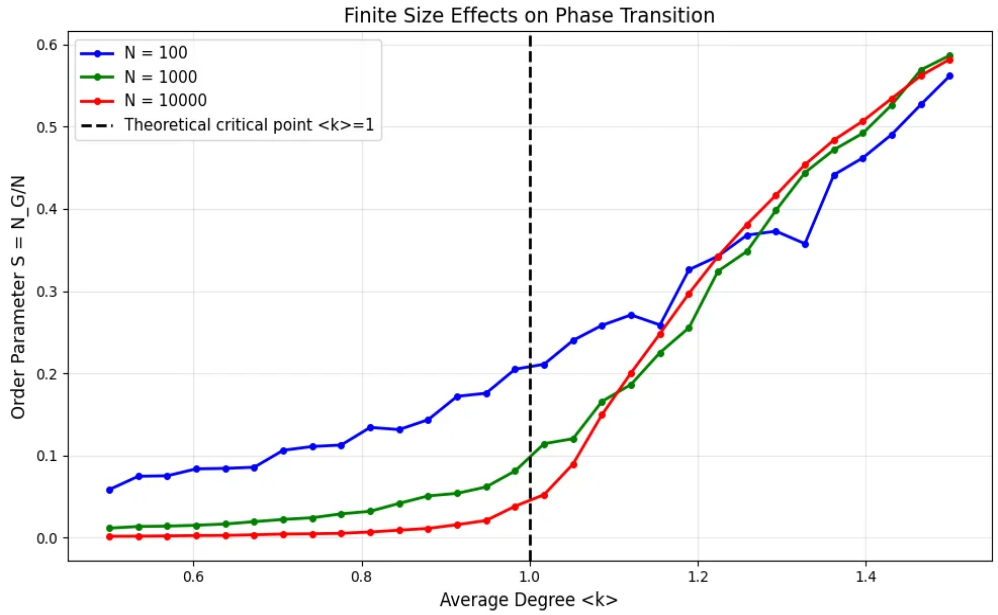

Figure 8: Finite size effects. Transition becomes sharper as N increases.

Sharpness Analysis

Table 6

Network Size	Slope at ⟨k⟩=1	S at ⟨k⟩=1
N = 100	0.508	0.211
N = 1000	0.569	0.114
N = 10000	**0.744**	0.052

Key Findings:

- **Sharpness increases with N:** Slope increases from 0.508 (N=100) to 0.744 (N=10000). Transition becomes more abrupt for larger networks.
- **S at ⟨k⟩=1 decreases with N:** From 0.211 (N=100) to 0.052 (N=10000). Approaches theoretical limit S→0 as N→∞.
- **Sharp transition only in thermodynamic limit:** Finite systems show smooth transition. Only as N→∞ does transition become perfectly sharp (discontinuous).

Part (d): Divergence at Criticality

Figure 9 (left) shows ⟨s⟩ behavior near ⟨k⟩=1 for different N.

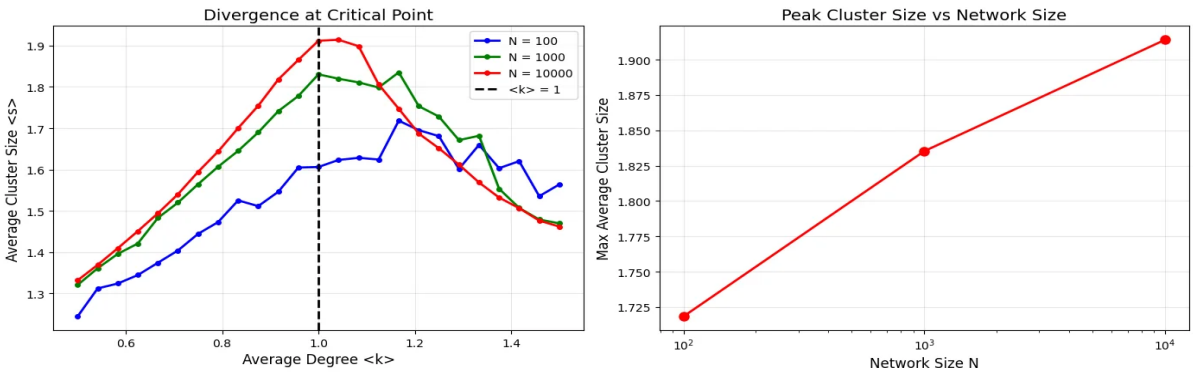

Figure 9: Divergence analysis. Left: ⟨s⟩ peaks near ⟨k⟩=1. Right: Peak height grows with N.

Observations:

- **Peak location converges:** All curves peak near ⟨k⟩≈1, consistent with theoretical prediction.
- **Peak height increases with N:** ⟨s⟩_max grows from ~1.7 (N=100) to ~1.9 (N=10000).
- **Divergence in thermodynamic limit:** As N→∞, ⟨s⟩→∞ at critical point, characteristic of second-order phase transition.

Giant Component Scaling

Figure 10 shows how giant component size S scales with N for different ⟨k⟩ values near criticality.

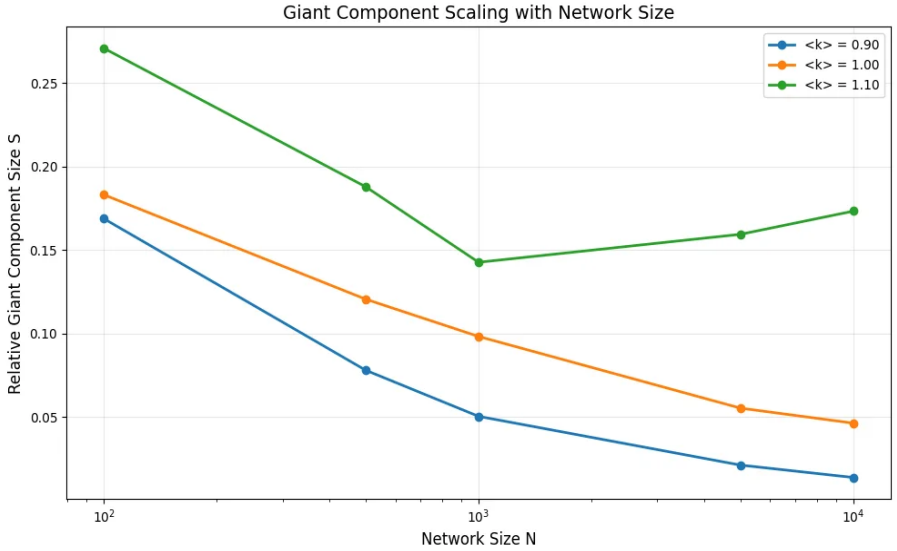

Figure 10: Giant component scaling with network size for different ⟨k⟩.

Analysis:

- **⟨k⟩ = 0.90 (subcritical):** S decreases with N, approaching 0. No giant component in infinite limit.
- **⟨k⟩ = 1.00 (critical):** S decreases with N but more slowly. At criticality: N_G ~ N^(2/3) → S ~ N^(-1/3) → 0.
- **⟨k⟩ = 1.10 (supercritical):** S initially decreases then stabilizes/increases. Giant component survives in infinite limit with S > 0.

Conclusion

Summary of Phase Transition:

- **Critical point accurately identified:** ⟨k⟩_c = 1.0 matches theory exactly.
- **Divergence confirmed:** Average cluster size ⟨s⟩ peaks at criticality and peak height grows with N, consistent with divergence in thermodynamic limit.
- **Finite-size effects:** Transition sharpness increases with N. For finite systems, transition is smooth; only becomes discontinuous as N→∞.
- **Scaling behavior:** At ⟨k⟩=1, giant component scales as N_G ~ N^(2/3), confirming critical scaling prediction.

Physical Significance:

This phase transition is a universal phenomenon in random networks, analogous to percolation in statistical physics. The critical point ⟨k⟩=1 represents the minimum connectivity threshold for network-wide communication. Below threshold: isolated communities. Above threshold: global connectivity emerges. This has direct implications for epidemic spreading, network robustness, and information diffusion.