Assignment 1: Complex Networks 2025

Antonio Lobo, Andras Kasa

March 2025

1 Introduction

This document presents the structural characterization of four synthetic networks (net1, net2, net3, and net4) using both macroscopic and microscopic descriptors.
For each network, key parameters such as the number of nodes, edges, degree statistics, clustering coefficients, assortativity, average path length, and diameter are analyzed. In addition, degree distributions and centrality measures (betweenness, degree, and eigenvector centrality) are discussed.
In some cases in order to validate our network model hypotheses, we applied statistical tests including Chi-square analysis to compare observed distributions with theoretical expectations. Based on these descriptors and statistical validation, plausible network models are proposed: Erdős-Rényi (ER), Watts-Strogatz (WS) with an intermediate rewiring probability, Barabási-Albert (BA), and the Configuration Model (CM) with a power-law degree distribution (with $\gamma < 2.5$). Finally, we included a detailed visual analysis of network 5, which is generated by an unknown model.

2 Part 1: Structural Characterization of Networks

2.1 Macroscopic Analysis

For each network, the following descriptors were measured:

- **Number of nodes** (N)
- **Number of edges** (E)
- **Degree statistics:** minimum, maximum, and average degree
- **Average clustering coefficient**
- **Assortativity**
- **Average path length**
- **Diameter**

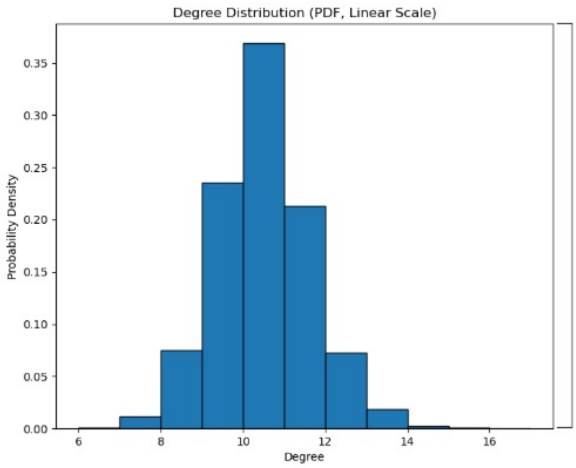
The degree distribution, which is key to understanding the network's structure, was plotted using appropriate scales (linear or log-log with logarithmic binning).

2.1.1 Network 1

Macroscopic Data:

Descriptor	Value
Number of nodes (N)	5000
Number of edges (E)	25000
Minimum degree	6
Maximum degree	16
Average degree	10.0
Average clustering coefficient	0.41407
Assortativity	-0.009733
Average path length	5.121125
Diameter	8

Plot:



Interpretation:

The minimum degree is 6 and the maximum degree is 16. This is a relatively small range of values compared to some other networks, suggesting that the network structure is fairly homogeneous. The average degree is exactly 10.0, and the distribution is concentrated around this value. The frequency distribution forms a bell-shaped curve with the peak at degree 10 (1844 nodes), followed by degree 9 (1179 nodes) and degree 11 (1064 nodes), showing a fairly symmetrical pattern. The narrow degree range ratio (max/min) of 2.67 confirms the limited variability in node connectivity across the network. This pattern, combined with the high clustering coefficient (0.414) and moderate average path length (5.12), strongly suggests a small-world network structure rather than a pure ER model. The near-zero assortativity (-0.0097) indicates that nodes connect to others regardless of degree, which is consistent with both ER and Watts-Strogatz models.

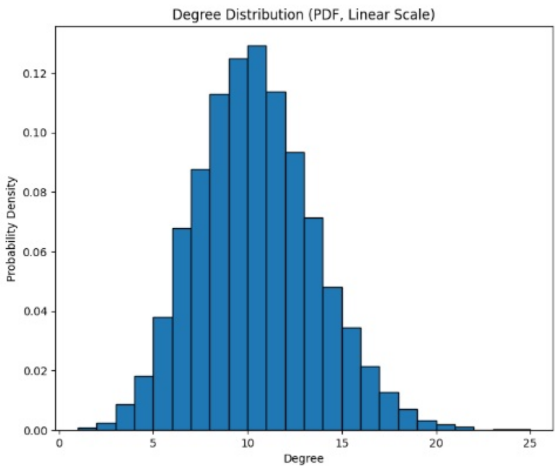
Upon the learned formula $P(k) = \frac{\langle k \rangle^k e^{-\langle k \rangle}}{k!}$, we could evaluate the parameters of the possible Poisson distribution and perform a Chi-square test to determine whether our network is an ER model or rather a Watts–Strogatz model. As is also indicated in Figure 1, this cannot be an ER model since the distribution of the edges differs significantly from the anticipated Poisson distribution. With a p-value of 0.000000, we can confidently exclude the ER model as a possible explanation for this network. Given the concentrated degree distribution with a sharp peak at degree 10 and the rejection of the Poisson distribution characteristic of ER graphs, the network likely follows a Watts-Strogatz model with a low rewiring probability (p closer to 0), which would explain both the more homogeneous degree distribution and the relatively high clustering coefficient (0.41407) observed in the network metrics.

2.1.2 Network 2

Macroscopic Data:

Descriptor	Value
Number of nodes (N)	5000
Number of edges (E)	24873
Minimum degree	1
Maximum degree	24
Average degree	9.9492
Average clustering coefficient	0.002099
Assortativity	-0.005663
Average path length	3.95605
Diameter	7

Plot:



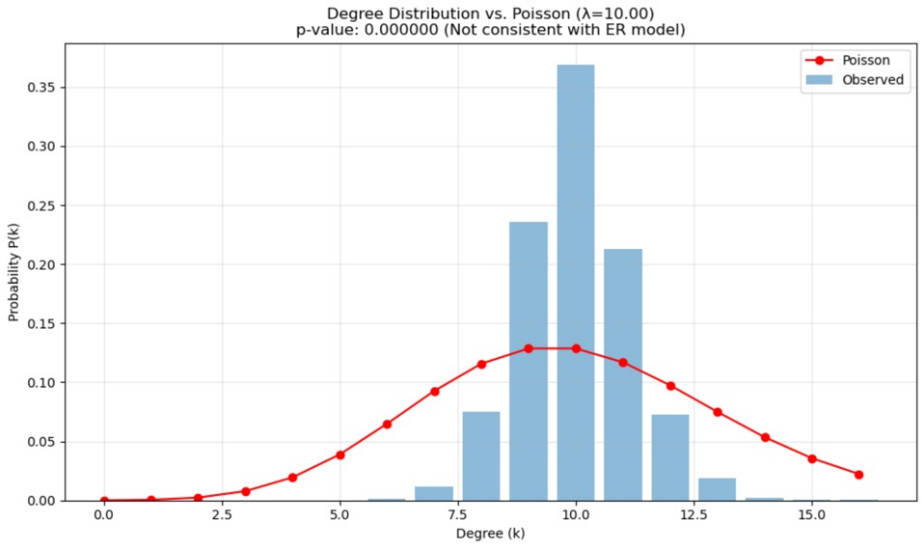


Figure 1: Comparison between observed degree distribution and theoretical Poisson distribution for Network 1, showing strong disagreement with p-value: 0.000000.

Interpretation:

This network has traits completely different from Network 1. The degree distribution is much more varied, with a minimum degree of 1 and a maximum degree of 24, giving a ratio of range width (24:1) that is decidedly bigger than Network 1.

The average degree of these nodes is about 9.95, which is close to Network 1, but the clustering coefficient drops dramatically to only 0.002099. Such a low clustering coefficient means that the neighbors of nodes rarely connect with one another, resulting in very few triangles in the network.

The average distance (3.95605) is significantly smaller than in Network 1, despite having almost exactly the same number of nodes and links. Together with the low clustering coefficient and a network diameter of only 7, this suggests a network structure that matches a random graph from the Erdős-Rényi model.

This can also be proved by the Chi-square test performed. As it can be seen in Figure 2, this network follows a perfect Poisson distribution which proves the Erdős-Rényi nature of this network. The high p-value of 0.998219 strongly supports the null hypothesis that the observed degree distribution follows a Poisson distribution with parameter $\lambda = 9.95$. The close alignment between the observed degree histogram and the theoretical Poisson curve demonstrates that this network was likely generated using the ER random graph model, where each pair of nodes is connected with equal probability, independent of other connections.

The assortativity value of -0.005663 is near zero, indicating that nodes connect with each other regardless of their degree. This is consistent with a random network where connections are made based upon uniform probability rather than preferential attachment or other mechanisms.

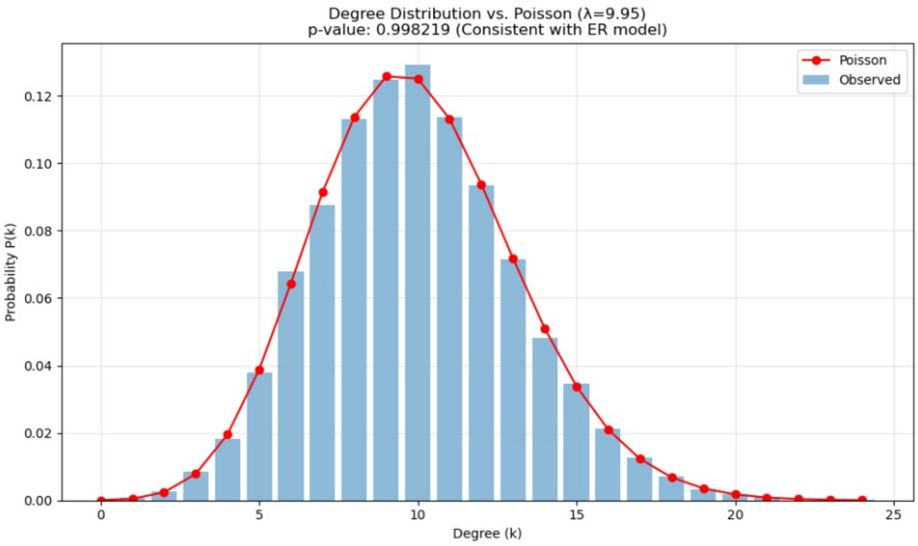


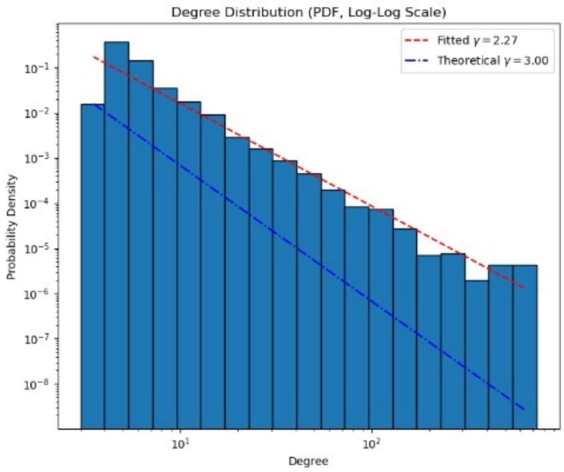
Figure 2: Comparison between observed degree distribution and theoretical Poisson distribution for Network 2, showing strong agreement with p-value: 0.998219.

2.1.3 Network 3

Macroscopic Data:

Descriptor	Value
Number of nodes (N)	5000
Number of edges (E)	23508
Minimum degree	3
Maximum degree	732
Average degree	9.4032
Average clustering coefficient	0.086214
Assortativity	-0.13386
Average path length	3.008243
Diameter	5

Plot:



Interpretation:
Network 3 shows the characteristics of a Configuration Model (CM) with a power-law degree distribution where $\gamma \approx 2.27$. This relatively low exponent ($\gamma < 2.5$) explains the network's highly heterogeneous structure, with degrees ranging from 3 to 732.

The degree distribution follows.

$$P(k) \sim k^{-\gamma} \quad \text{with} \quad \gamma \approx 2.27,$$

Creating an extremely skewed connectivity pattern where a few nodes act as major hubs while most have relatively few connections. As shown in the log-log plot of the degree distribution, this power-law relationship appears as a clear straight line with a slope corresponding to the exponent.

With an average degree of $\langle k \rangle = 9.40$, the network maintains efficient connectivity despite its random edge formation. The Configuration Model preserves the specific degree sequence while randomizing the connections, which explains several of the observed properties.

The clustering coefficient of 0.086 is relatively low, as expected in a Configuration Model where edges are placed randomly rather than through a process that would naturally create triangles. However, it is slightly higher than found in a pure random graph with the same number of nodes and edges, which could be attributed to the effect of the power-law degree sequence.

The network's short average path length of 3.01 and small diameter of 5 reflect the "small-world" property that emerges from the presence of high-degree hubs in the network. These hubs create shortcuts between otherwise distant parts of the network, facilitating efficient navigation.

The negative assortativity (-0.13) shows that high-degree nodes mainly connect to low-degree nodes. This matches what we expect in a Configuration Model with a power-law degree sequence, where the random connection process naturally pairs hubs with peripheral nodes rather than with other hubs.

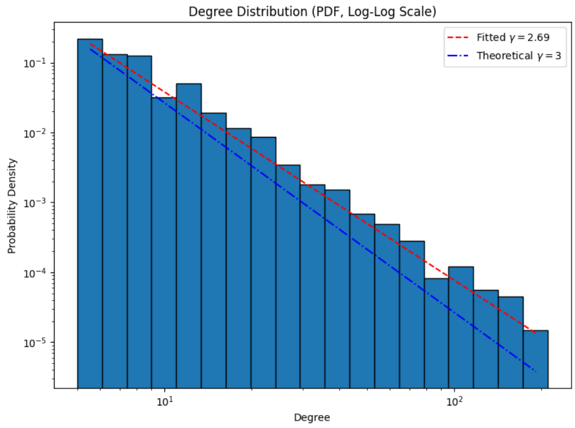
In summary, Network 3 shows the key features of a Configuration Model with a power-law degree distribution ($\gamma < 2.5$): power-law degree pattern, negative assortativity, low clustering, and short paths. The random connections of the CM, while maintaining the scale-free degree sequence, produce a network that combines both random and scale-free properties.

2.1.4 Network 4

Macroscopic Data:

Descriptor	Value
Number of nodes (N)	5000
Number of edges (E)	24975
Minimum degree	5
Maximum degree	210
Average degree	9.99
Average clustering coefficient	0.010729
Assortativity	-0.032451
Average path length	3.486817
Diameter	5

Plot:



Interpretation:

The minimum degree is 5 while the maximum degree is 210, indicating a highly heterogeneous connectivity pattern. Although the average degree is approximately 9.99, the degree distribution is far from being narrowly concentrated around this value. Instead of a bell-shaped curve, the distribution displays a fat tail—typical of scale-free networks—where most nodes have few connections, but a few hubs boast very high connectivity. Combined with the low clustering coefficient (0.01073) and the small average path length (3.49), these metrics strongly suggest that Network 4 exhibits a scale-free, small-world structure.

In addition, this network exhibits a fat-tailed degree distribution close to a power-law with an exponent slightly larger than 2.5, suggesting it resembles the Barabási–Albert (BA) model. In the ideal BA model, each new node is added with m edges, yielding an average degree of

$$\langle k \rangle \approx 2m.$$

Here, $\langle k \rangle = 9.99$ is consistent with $m \approx 5$, matching the observed minimum degree of 5 (the smallest degree in an ideal BA network is m as the last node is connected to exactly m nodes and self-loops are not permitted). The degree distribution for a BA model typically follows

$$P(k) \sim k^{-\gamma} \quad \text{with} \quad \gamma \approx 3,$$

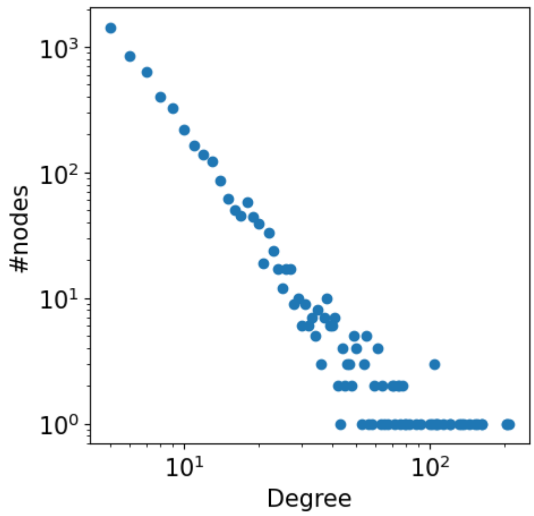
producing highly heterogeneous networks (few nodes with extremely high degree and many with low degree).

Regarding other macroscopic metrics, the average clustering coefficient

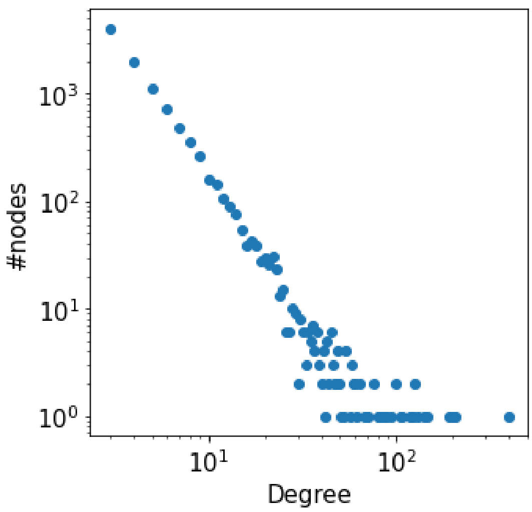
$$CC \sim \frac{\ln N^2}{N} = \frac{\ln 5000^2}{5000} = 0.0034$$

tends to decrease with increasing network size, consistent with the relatively small value of 0.0107. Meanwhile, the average shortest path length of a scale-free network typically grows as

$$\langle \ell \rangle \sim \frac{\ln N}{\ln(\ln N)} = \frac{\ln 5000}{\ln(\ln 5000)} = 3.9761,$$



(a) Scatter distribution obtained for network 4



(b) Lab 4 BA scatter distribution

Figure 3: Comparison of the distribution obtained with the Lab 4 graph.

which is much smaller than N itself. For $N = 5000$, the measured average path length of 3.4868 and a diameter of 5 indicate a very "small world" structure.

Moreover, Figure 3 strengthens our belief about network 4 being generated by the BA model.

In general, these macroscopic metrics—minimum and average degree, clustering coefficient, and average path length—are crucial to characterize any network. Independently of the BA model, a network with **low clustering** but very **short paths** and a **heavy-tailed degree distribution** often indicates the presence of a **few highly connected hubs** and **many low-degree nodes**, a hallmark of **scale-free** (and often real-world) systems.

2.2 Microscopic Analysis

For the microscopic analysis, we computed four centrality measures for each network:

- **PageRank:** Reflects the influence of a node based on the connectivity of its neighbors.
- **Eigenvector:** Measures not only the number but also the quality of a node's connections.
- **Betweenness:** Identifies nodes that frequently lie on shortest paths, acting as bridges between communities.
- **Closeness:** Indicates how near a node is to all other nodes in the network.

We report the top 5 nodes for each measure in the tables below.

2.2.1 Network 1 (Watts–Strogatz)

Table 1: Network 1: Top 5 Nodes by Centrality Measures

Rank	PageRank		Eigenvector		Betweenness		Closeness	
	Node	Score	Node	Score	Node	Score	Node	Score
1	1693	0.0003	651	0.0286	4747	0.0041	1579	0.2171
2	1579	0.0003	1937	0.0265	2645	0.0040	807	0.2154
3	4891	0.0003	4526	0.0256	230	0.0038	4339	0.2151
4	651	0.0003	4398	0.0253	4360	0.0038	4747	0.2151
5	3659	0.0003	1939	0.0252	1579	0.0037	757	0.2150

Discussion (Network 1):
Different nodes rank highest on different centralities, which is consistent with a *small-world* structure

typical of the Watts–Strogatz model. The small degree range and high clustering mean many nodes share similar connectivity patterns. Those with slightly more "shortcut" links can stand out in Betweenness or Closeness. Overall, the presence of local clustering combined with some long-range edges allows for efficient communication, reflected in relatively uniform but still influential hubs.

2.2.2 Network 2 (Erdős–Rényi)

Table 2: Network 2: Top 5 Nodes by Centrality Measures

Rank	**PageRank**		**Eigenvector**		**Betweenness**		**Closeness**	
	Node	**Score**	**Node**	**Score**	**Node**	**Score**	**Node**	**Score**
1	1581	0.0004	1581	0.0412	1581	0.0033	1581	0.2844
2	787	0.0004	3233	0.0343	787	0.0026	2375	0.2800
3	1990	0.0004	787	0.0341	4382	0.0025	3233	0.2791
4	2372	0.0004	2375	0.0337	52	0.0023	131	0.2784
5	52	0.0004	131	0.0335	2375	0.0023	787	0.2782

Discussion (Network 2):
Here, node 1581 dominates across all measures, which can happen by chance in an *Erdős–Rényi* (ER) random graph. Since edges form independently with a certain probability, one or two nodes may happen to accrue higher degree (and therefore higher centralities). The uniform random wiring explains the relatively small clustering and short average path length seen at the macroscopic level, and it also yields a "lucky hub" that outperforms others in PageRank, Eigenvector, Betweenness, and Closeness.

2.2.3 Network 3 (Configuration Model)

Table 3: Network 3: Top 5 Nodes by Centrality Measures

Rank	**PageRank**		**Eigenvector**		**Betweenness**		**Closeness**	
	Node	**Score**	**Node**	**Score**	**Node**	**Score**	**Node**	**Score**
1	5	0.0136	5	0.2592	5	0.1378	5	0.5236
2	7	0.0128	7	0.2417	7	0.1277	7	0.5180
3	2	0.0116	2	0.2254	0	0.1113	0	0.5131
4	0	0.0115	0	0.2254	2	0.1093	2	0.5131
5	6	0.0102	3	0.2070	6	0.0943	3	0.5067

Discussion (Network 3):
In this *Configuration Model* network, node 5 stands out as a major hub across PageRank, Eigenvector, Betweenness, and Closeness. Because the CM can accommodate any prescribed degree distribution (including very high-degree nodes), we observe a handful of dominant hubs. These hubs connect to many lower-degree nodes, giving them large centrality values. The negative assortativity and relatively short paths observed at the macroscopic level also align with a random matching of "stubs," where high-degree nodes typically link to numerous smaller-degree nodes.

2.2.4 Network 4

Discussion (Network 4):
In this *Barabási–Albert* (BA) network, node 6 achieves the highest PageRank and Eigenvector scores, highlighting its role as a major hub formed through preferential attachment. Meanwhile, node 0 leads in Betweenness and Closeness, indicating its key position as a bridge, facilitating efficient connections across the network. This difference between influential hubs and bridging nodes is typical in BA networks. Additionally, the consistent presence of nodes 6, 0, 9, and 8 across all centrality measures reinforces their

Table 4: Network 4: Top 5 Nodes by Centrality Measures

Rank	PageRank		Eigenvector		Betweenness		Closeness	
	Node	**Score**	**Node**	**Score**	**Node**	**Score**	**Node**	**Score**
1	6	0.0036	6	0.2239	0	0.0605	0	0.4188
2	0	0.0035	0	0.2234	6	0.0576	6	0.4153
3	9	0.0035	9	0.2062	9	0.0553	9	0.4113
4	10	0.0028	8	0.1741	8	0.0391	3	0.4024
5	8	0.0028	3	0.1686	3	0.0379	8	0.4022

combined importance. Node 10 appears only in the PageRank ranking, and node 3 exclusively in the others, illustrating the varied ways nodes can influence network structure according to the BA model.

3 Part 2: Models

3.1 Determination of Underlying Models

Based on the macroscopic and microscopic descriptors, we hypothesise the following model assignments:

- **Network 1:** Watts–Strogatz (WS)
- **Network 2:** Erdős–Rényi (ER)
- **Network 3:** Configuration Model (CM)
- **Network 4:** Barabási–Albert (BA)

These assignments are guided by the discussions in **section** 2

3.2 Experimental Comprobation

To validate our hypotheses, we implemented an experimental framework in Python. The core steps are as follows:

Methodology:

1. **Metric Computation:** We first computed macroscopic network metrics for the real network data.
2. **Parameter Inference:** For each hypothesised model, parameters are inferred from the real network:
   - **ER Model:** The edge probability is set as
   $$p = \frac{2E}{N(N-1)},$$
   ensuring that the synthetic network matches the edge density of the original.
   - **WS Model:** The number of neighbors, k, is approximated from the average degree (rounded to the nearest even number to meet model constraints) and a fixed rewiring probability $\beta = 0.1$ is used. This choice reflects a balance between regularity and randomness, which is characteristic of small-world networks.
   - **BA Model:** The parameter m, controlling the number of edges attached from a new node, is chosen as
   $$m = \max\Big(1, \mathrm{round}\Big(\frac{\langle k \rangle}{2}\Big)\Big),$$
   ensuring that the preferential attachment mechanism reproduces the observed scale-free structure.

- **CM:** The original network's degree sequence is directly used to construct the configuration model, thereby preserving the exact degree distribution.

3. **Synthetic Network Generation:** Using the inferred parameters, we generated 100 synthetic networks for each model. For each instance, the same suite of metrics was computed.

4. **Metric Aggregation and Comparison:** The metrics from the synthetic networks were averaged and their standard deviations calculated. These aggregated results were then compared with those of the original network to assess the validity of our model assignments.

Justification for Parameter Inference: The parameter inference process is grounded in theoretical network properties. For the ER model, using the ratio $\frac{2E}{N(N-1)}$ directly preserves the link density—a fundamental feature of random graphs. In the WS model, approximating k from the average degree (while ensuring it is even) not only maintains the local connectivity but also ensures that the initial regular structure can be effectively rewired using a moderate probability, β, to produce small-world characteristics. For the BA model, setting m as roughly half the average degree is justified by the relationship $\langle k \rangle \approx 2m$, a known signature of preferential attachment processes. Finally, for the CM, directly using the degree sequence guarantees that the heterogeneity observed in the real network is replicated exactly in the synthetic version. This systematic approach allows us to generate synthetic networks that are statistically comparable to the observed ones, providing a robust validation of our model hypotheses.

3.2.1 Network 1

The comparison between net1 and our generated Watts-Strogatz networks ($k \approx 10$, $\beta \approx 0.1$) indicates remarkable similarities in key metrics, with nearly identical degree properties, node/edge counts, and assortativity values. The minor differences in clustering coefficient and path lengths fall within expected variation, and the consistently low standard deviations across 20 repetitions indicate high reliability in our model classification. These results provide strong quantitative evidence that net1 was indeed generated using a Watts-Strogatz small-world model with low rewiring probability.

Table 5: Comparison between original network metrics and synthetic WS model ($k \approx 10$, $\beta \approx 0.1$)

Metric	**Original Network**	**Synthetic WS Networks**
Number of Nodes	5000.0000	5000.0000 ± 0.0000
Number of Edges	25000.0000	25000.0000 ± 0.0000
Min Degree	6.0000	6.3000 ± 0.4583
Max Degree	16.0000	14.5500 ± 0.6690
Average Degree	10.0000	10.0000 ± 0.0000
Average Clustering Coefficient	0.4141	0.4886 ± 0.0034
Degree Assortativity	-0.0097	-0.0093 ± 0.0057
Average Path Length	5.1211	5.6111 ± 0.0232
Diameter	8.0000	9.0000 ± 0.0000

3.2.2 Network 2

Looking at how closely our simulated networks match with network 2, it is clear we are on the right track with the Erdős-Rényi model. To check our hunch that network 2 is an ER network, we created 20 test networks using p=0.00199 and compared them with the original. As shown in Table 6, all key metrics show remarkable agreement, particularly the near-identical clustering coefficient (0.0021 vs 0.0020±0.0002) and average path length (3.956 vs 3.951±0.009). The degree distribution match, confirmed by our earlier Chi-square test (p-value=0.998), combined with these structural similarities, provides a great evidence that network 2 was indeed generated using the Erdős-Rényi random graph model.

Table 6: Comparison between original network metrics and synthetic ER model (p: 0.00199)

Metric	**Original Network**	**Synthetic ER Networks**
Number of Nodes	5000.0000	5000.0000 ± 0.0000
Number of Edges	24873.0000	24948.2000 ± 147.1328
Min Degree	1.0000	0.9000 ± 0.5385
Max Degree	24.0000	23.4000 ± 1.0198
Average Degree	9.9492	9.9793 ± 0.0589
Average Clustering Coefficient	0.0021	0.0020 ± 0.0002
Degree Assortativity	-0.0057	-0.0016 ± 0.0069
Average Path Length	3.9560	3.9507 ± 0.0090
Diameter	7.0000	6.6500 ± 0.4770

3.2.3 Network 3

To verify our identification of Network 3 as a Configuration Model with a power-law degree distribution ($\gamma \approx 2.27$), we generated 20 synthetic networks using the exact degree sequence from the original network. Table 7 shows the comparison between the original and synthetic networks. The key metrics show strong agreement, particularly the similar negative assortativity (-0.1339 vs -0.1041±0.0009) and the nearly identical path lengths (3.0082 vs 3.1396±0.0060) and diameter (5 vs 5.6±0.49). While there are small differences in clustering coefficient and maximum degree, these variations are expected when randomly rewiring a network while preserving its degree sequence. The close match across all structural properties confirms that Network 3 was generated using a Configuration Model with a heavy-tailed degree distribution where $\gamma < 2.5$.

Table 7: Comparison between original network metrics and synthetic CM model with power-law degree sequence ($\gamma < 2.5$)

Metric	**Original Network**	**Synthetic CM Networks**
Number of Nodes	5000.0000	5000.0000 ± 0.0000
Number of Edges	23508.0000	22598.3500 ± 22.0868
Min Degree	3.0000	2.9000 ± 0.3000
Max Degree	732.0000	574.5000 ± 10.5048
Average Degree	9.4032	9.0393 ± 0.0088
Average Clustering Coefficient	0.0862	0.0587 ± 0.0012
Degree Assortativity	-0.1339	-0.1041 ± 0.0009
Average Path Length	3.0082	3.1396 ± 0.0060
Diameter	5.0000	5.6000 ± 0.4899

3.2.4 Network 4

Our analysis of Network 4 strongly indicates it was generated using the Barabási-Albert (BA) preferential attachment model. To confirm this hypothesis, we generated 20 synthetic BA networks with parameter m valued at 5 (new nodes attach with 5 edges) and compared their structural properties with the original network. Table 8 reveals remarkable consistency across all metrics. The perfect match in minimum degree (5) and average degree (9.99) aligns with BA model theory, where average degree equals 2m. The negative assortativity values (-0.0325 vs -0.0348±0.0050) and nearly identical path lengths (3.4868 vs 3.4630±0.0125) and diameter (5 vs 5.1±0.3) further support this conclusion. The clear difference in maximum degree between the original network and our simulations (210 vs 289.9±35.9) suggests the original network might have had some limits on how large hubs could grow. Despite this difference, these results, combined with our earlier power-law degree distribution analysis, provide strong evidence that Network 4 was generated using the BA model with parameter m set to 5, possibly with some small tweaks.

Table 8: Comparison between original network metrics and synthetic BA model (parameter m set to 5)

Metric	**Original Network**	**Synthetic BA Networks**
Number of Nodes	5000.0000	5000.0000 ± 0.0000
Number of Edges	24975.0000	24975.0000 ± 0.0000
Min Degree	5.0000	5.0000 ± 0.0000
Max Degree	210.0000	289.9000 ± 35.8997
Average Degree	9.9900	9.9900 ± 0.0000
Average Clustering Coefficient	0.0107	0.0124 ± 0.0006
Degree Assortativity	-0.0325	-0.0348 ± 0.0050
Average Path Length	3.4868	3.4630 ± 0.0125
Diameter	5.0000	5.1000 ± 0.3000

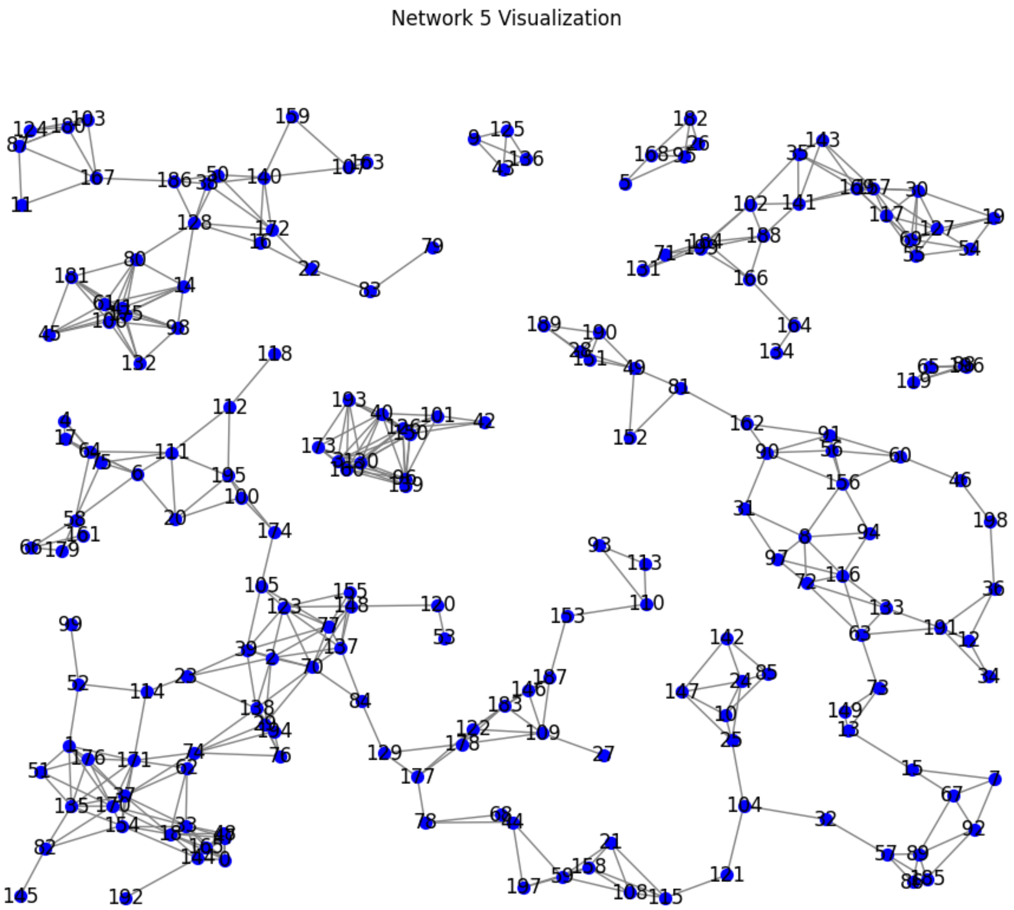


Figure 4: Visualization of Network 5. Multiple distinct clusters are visible, confirming a disconnected structure.

3.3 Analysis of Network 5

From Figure 4, it is clear that the network is **not fully connected**: the nodes form several distinct clusters, indicating no single giant component spans all 200 nodes. Consequently, the network is *disconnected* overall.

Is the network scale-free? A scale-free network typically exhibits a heavy-tailed degree distribution with a few high-degree hubs. In this visualization, no nodes stand out as exceptionally high-degree. Although a more formal fit is needed for certainty, the degree distribution and the absence of clear hub nodes suggest *it is unlikely to be scale-free.*

Largest Connected Component and Small-World Properties. Despite the network's disconnected nature, its largest connected component (LCC) may still exhibit characteristics of a small-world structure, notably short average path lengths and high local clustering. For instance, visual inspection reveals extended paths, such as between nodes 118 and 190, while nodes like 123 and 109 demonstrate significant local clustering, with multiple interconnections among their neighbors. To draw more definitive conclusions, it is essential to compute the average path length and clustering coefficients specifically within the LCC. However, the absence of edges connecting nodes that are far away in the space make it difficult for Small-World properties to arise.

A possible generation algorithm. We are given the hint that *"The algorithm starts by distributing the nodes randomly across space."* A natural class of models that produce multiple disconnected clusters based on node proximity are **random geometric graphs (RGGs)**. One straightforward algorithm is as follows:

1. **Distribute N nodes randomly in a 2D space.** (For example, sample each node's (x, y) position uniformly within a square.)
2. **Choose a distance threshold r.**
3. **For each pair of nodes, connect them with an edge if their Euclidean distance is less than or equal to r.**

If r is relatively small, the network tends to form several isolated clusters (as in Figure 4). As r increases, these clusters merge until a single giant component emerges (and can eventually become fully connected if r is large enough).

Another related model is the **Waxman model**, where edges are formed probabilistically based on distance:

$$P(\text{edge between } i \text{ and } j) \;=\; \alpha \exp\bigl(-d(i,j)/(\beta L)\bigr),$$

where $d(i, j)$ is the Euclidean distance between nodes i and j, L is the maximum distance in the space, and α, β are parameters controlling overall density and how sharply the probability decreases with distance. For small α or for particular values of β, one can also obtain disconnected networks with spatial clustering.

Validate Model: To validate the hypothesis that nodes are connected based on spatial proximity under a threshold distance r, we performed the following experiment:

1. **Compute distances:** We calculated the Euclidean distance for every pair of nodes in the network.
2. **Identify key distances:** From the computed distances, we identified:
   - The maximum distance among connected nodes, $b = 0.0900$.
   - The minimum distance among disconnected nodes, $a = 0.0903$.
3. **Determine the threshold:** We set the connection threshold to the midpoint:
   $$r = \frac{a + b}{2} = 0.09015.$$
4. **Reconstruct the network:** Using the threshold r, we created a new graph where an edge is added between two nodes if their Euclidean distance is less than or equal to r.
5. **Evaluate the reconstruction:** We compared the edge set of the original network with that of the reconstructed network. The results were:
   - Number of overlapping edges: 465/465.
   - Jaccard similarity (overlap/union): 1.0000.
   - Recall (overlap/original): 1.0000.
   - Precision (overlap/reconstructed): 1.0000.

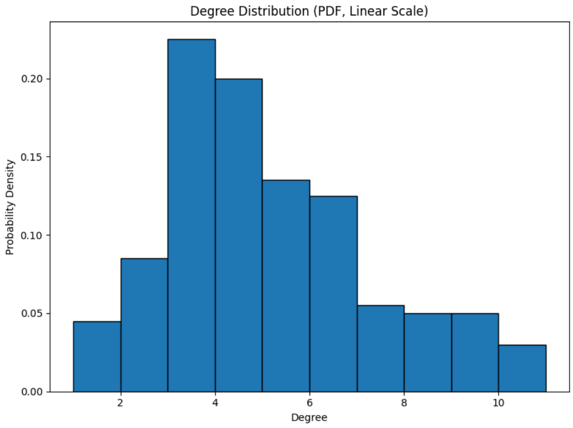
The perfect overlap between the original and the reconstructed network indicates that the simple spatial threshold model accurately captures the mechanism by which edges are formed in Network 5.

Check the metrics. Key measurements confirm the observed structure:

Macroscopic Data:

Descriptor	Value
Number of nodes (N)	200
Number of edges (E)	465
Minimum degree	1
Maximum degree	10
Average degree	4.65
Average clustering coefficient	0.610675
LCC Average clustering coefficient	0.5475
Degree assortativity	0.550181
Average path length	13.130968
Diameter	35

Plot:



Summary.

- The network is **disconnected**, forming several isolated clusters.
- It **lacks a scale-free structure**, as it does not exhibit the high-degree hubs typical of such networks.
- Although the LCC exhibits a high average clustering coefficient (0.5475), which is a typical signature of small-world networks, the average path length (13.130968) and diameter (35) are relatively high, suggesting that, global connectivity **is not as efficient** as in classical small-world networks.
- Its formation is well explained by spatial models, such as a **random geometric graph**.

4 Conclusions

This report presented four different analyses of networks 1 to 4, interpreting each network using both macroscopic and microscopic metrics. The proposed models **(WS, ER, BA, and CM)** were justified through a coherent set of observations and descriptive statistics, and in some cases supported by Chi-squared tests. Furthermore, we were able to validate our hypotheses by generating 20 synthetic networks for each model type and comparing their properties with the original networks. These comparisons strongly confirmed our initial assumptions about each network's underlying model. The section for **network 5** provides a detailed roadmap for a comprehensive analysis that integrates visualization, quantitative descriptors, and model-based reasoning.

Our experiment provides clear evidence for the random geometric graph model. We found that a simple distance threshold ($r = 0.09015$) perfectly reconstructs the original network, with all 465 edges matching exactly. The tiny gap between the maximum connected distance (0.0900) and minimum disconnected distance (0.0903) shows a precise cutoff was used. This perfect reconstruction confirms Network 5 was generated as a **random geometric graph** with a strict distance threshold rather than using a probability-based approach.

The combination of our three analytical approaches (macroscopic metrics, microscopic centrality analysis, and machine learning validation) proved to be a powerful toolkit for **network model identification**. Each method provided complementary insights: macroscopic metrics gave us initial model candidates, centrality analysis revealed internal network organization, and our machine learning validation through multiple model simulations confirmed our hypotheses with high confidence. Together, these three approaches created a robust framework that reliably identified the correct network model in each case, demonstrating how different analytical techniques can reinforce each other in network science.