Network Models and Measurements

Yash Dani[*][†]
York University
Toronto, Canada
yashdani@my.yorku.ca

Abhinav Syal[‡][†]
York University
Toronto, Canada
syal28@my.yorku.ca

ABSTRACT

In this assignment, we generate and analyze networks using three distinct models: Erdős-Rényi, Watts-Strogatz, and Barabási-Albert. We report on the data, models, methods, and measurements, focusing on properties such as node degree distribution, clustering coefficients, and shortest path lengths. This analysis aims to highlight the structural differences and similarities across the generated networks.

KEYWORDS

Graph Theory, Network Analysis, Erdős-Rényi Model, Watts-Strogatz Model, Barabási-Albert Model, Node Degree Distribution, Clustering Coefficient, Shortest Path Lengths

1 INTRODUCTION

In this assignment, we focus on the generation and analysis of networks using three distinct graph models: Erdős-Rényi, Watts-Strogatz, and Barabási-Albert. These models represent different methodologies for constructing networks, each with unique characteristics and applications.

1.1 Motivation and Objective

Understanding the properties of various network models is essential in numerous fields, including social network analysis, biology, and computer science. By generating networks and measuring their key properties, we aim to gain insights into their structural differences and similarities.

1.2 Graph Models Overview

(1) **Erdős-Rényi Model**: This model generates random graphs by connecting nodes with a fixed probability p. It is one of the simplest and most well-studied models in network theory, useful for understanding random processes and percolation phenomena.
(2) **Watts-Strogatz Model**: Known for creating small-world networks, this model introduces a high clustering coefficient and short average path lengths, mimicking many real-world networks. It is particularly relevant for understanding social networks and natural phenomena.
(3) **Barabási-Albert Model**: This model generates scale-free networks through preferential attachment, where new nodes are more likely to connect to existing nodes with high degrees. It is used to explain the emergence of hubs and the power-law degree distribution observed in many complex networks.

[*]Yash's Student ID: 218465231
[†]Both authors contributed equally to this assignment.
[‡]Abhinav's Student ID: 218820330

1.3 Methodology

For each model, we generate three undirected, unweighted graphs with approximately 1,000 nodes and 10,000 edges. We report the parameter values used for generation and analyze the giant connected components of these graphs.

1.4 Measurements and Analysis

We perform several measurements on the generated graphs, focusing on their giant connected components. The key properties analyzed include:

- **Node Degree Distribution**: This metric helps us understand the connectivity of nodes within the network.
- **Local Clustering Coefficient Distribution**: This measures the tendency of nodes to form tightly knit groups.
- **Global Clustering Coefficient**: A single value representing the overall clustering in the network.
- **Shortest Path Length Distribution**: This provides insights into the efficiency of information or signal transmission across the network.
- **Average Shortest Path Length and Diameter**: These metrics help gauge the overall size and connectivity of the network.

1.5 Structure of the Report

This report is organized into several sections:

(1) **Graph Model Generators**: Descriptions and parameter values of the generated graphs.
(2) **Graph Measurements**: Detailed analysis and plots of the measured properties.
(3) **Discussion**: Comparison of properties within the same model and between different models.
(4) **Conclusion**: Summary of findings and potential future work.

By following this structured approach, we aim to provide a comprehensive analysis of the generated networks and their properties.

2 GRAPH MODEL GENERATORS

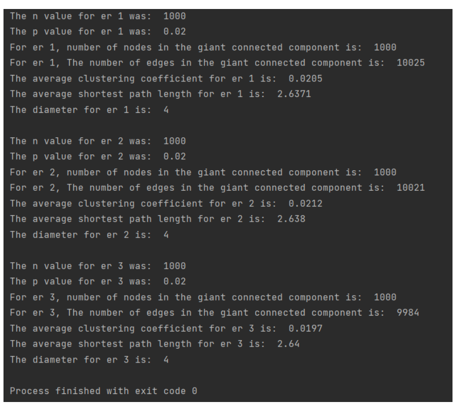
In this section, we generate undirected unweighted graphs using three different models: Erdős-Rényi, Watts-Strogatz, and Barabási-Albert. Each model uses specific parameters to construct the graphs, which are listed in the tables. The columns appearing before "Nodes" and "Edges" represent these parameters, which are used in the graph generator functions. For each model, we generate three graphs, ensuring each has approximately 1,000 nodes and 10,000 edges.

2.1 Erdős-Rényi Model

The Erdős-Rényi model is generated using the function `gnp_random_graph(n, p)` from the NetworkX library. Here, `n` represents the number of nodes, and `p` is the probability of an edge being present between any pair of nodes.

Table 1: Erdős-Rényi Parameter Values and Number of Nodes and Edges

Graph Name	n	p	Nodes	Edges
er1	1000	0.02	1000	10025
er2	1000	0.02	1000	10021
er3	1000	0.02	1000	9984



```
The n value for er 1 was:  1000
The p value for er 1 was:  0.02
For er 1, number of nodes in the giant connected component is:  1000
For er 1, The number of edges in the giant connected component is:  10025
The average clustering coefficient for er 1 is:  0.0205
The average shortest path length for er 1 is:  2.6371
The diameter for er 1 is:  4

The n value for er 2 was:  1000
The p value for er 2 was:  0.02
For er 2, number of nodes in the giant connected component is:  1000
For er 2, The number of edges in the giant connected component is:  10021
The average clustering coefficient for er 2 is:  0.0212
The average shortest path length for er 2 is:  2.638
The diameter for er 2 is:  4

The n value for er 3 was:  1000
The p value for er 3 was:  0.02
For er 3, number of nodes in the giant connected component is:  1000
For er 3, The number of edges in the giant connected component is:  9984
The average clustering coefficient for er 3 is:  0.0197
The average shortest path length for er 3 is:  2.64
The diameter for er 3 is:  4

Process finished with exit code 0
```

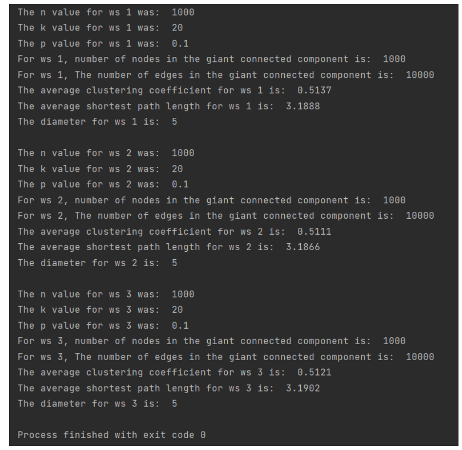
Figure 1: Erdős-Rényi Console Output

2.2 Watts-Strogatz Model

The Watts-Strogatz model is generated using the function `watts_strogatz_graph(n, k, p)` from the NetworkX library. Here, `n` represents the number of nodes, `k` is the number of nearest neighbors in a ring topology, and `p` is the probability of rewiring each edge.

Table 2: Watts-Strogatz Parameter Values and Number of Nodes and Edges

Graph Name	n	k	p	Nodes	Edges
ws1	1000	20	0.1	1000	10000
ws2	1000	20	0.1	1000	10000
ws3	1000	20	0.1	1000	10000



```
The n value for ws 1 was:  1000
The k value for ws 1 was:  20
The p value for ws 1 was:  0.1
For ws 1, number of nodes in the giant connected component is:  1000
For ws 1, The number of edges in the giant connected component is:  10000
The average clustering coefficient for ws 1 is:  0.5137
The average shortest path length for ws 1 is:  3.1888
The diameter for ws 1 is:  5

The n value for ws 2 was:  1000
The k value for ws 2 was:  20
The p value for ws 2 was:  0.1
For ws 2, number of nodes in the giant connected component is:  1000
For ws 2, The number of edges in the giant connected component is:  10000
The average clustering coefficient for ws 2 is:  0.5111
The average shortest path length for ws 2 is:  3.1866
The diameter for ws 2 is:  5

The n value for ws 3 was:  1000
The k value for ws 3 was:  20
The p value for ws 3 was:  0.1
For ws 3, number of nodes in the giant connected component is:  1000
For ws 3, The number of edges in the giant connected component is:  10000
The average clustering coefficient for ws 3 is:  0.5121
The average shortest path length for ws 3 is:  3.1902
The diameter for ws 3 is:  5

Process finished with exit code 0
```

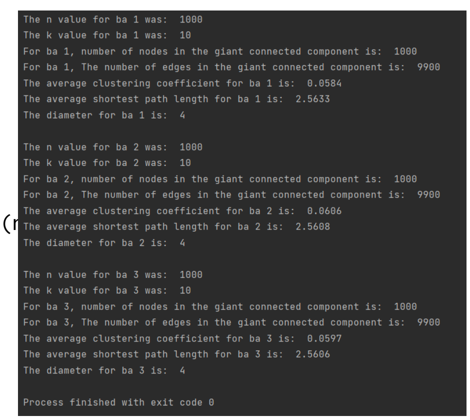
Figure 2: Watts-Strogatz Console Output

2.3 Barabási-Albert Model

The Barabási-Albert model is generated using the function `barabasi_albert_grap` `m)` from the NetworkX library. Here, `n` represents the number of nodes, and `m` is the number of edges to attach from a new node to existing nodes.

Table 3: Barabasi-Albert Parameter Values and Number of Nodes and Edges

Graph Name	n	m	Nodes	Edges
ba1	1000	10	1000	9900
ba2	1000	10	1000	9900
ba3	1000	10	1000	9900



```
The n value for ba 1 was:  1000
The k value for ba 1 was:  10
For ba 1, number of nodes in the giant connected component is:  1000
For ba 1, The number of edges in the giant connected component is:  9900
The average clustering coefficient for ba 1 is:  0.0584
The average shortest path length for ba 1 is:  2.5633
The diameter for ba 1 is:  4

The n value for ba 2 was:  1000
The k value for ba 2 was:  10
For ba 2, number of nodes in the giant connected component is:  1000
For ba 2, The number of edges in the giant connected component is:  9900
The average clustering coefficient for ba 2 is:  0.0606
The average shortest path length for ba 2 is:  2.5608
The diameter for ba 2 is:  4

The n value for ba 3 was:  1000
The k value for ba 3 was:  10
For ba 3, number of nodes in the giant connected component is:  1000
For ba 3, The number of edges in the giant connected component is:  9900
The average clustering coefficient for ba 3 is:  0.0597
The average shortest path length for ba 3 is:  2.5606
The diameter for ba 3 is:  4

Process finished with exit code 0
```

Figure 3: Barabasi-Albert Console Output

3 GRAPH MEASUREMENTS

In this section, we measure and analyze several key properties of the generated graphs, focusing on their derived giant connected components. The specific measurements and their importance are outlined in the introduction section.

3.1 Node Degree Distribution

The plots in this subsection display the node degree distribution for each generated graph. The x-axis represents the degree of the nodes, and the y-axis represents the probability $P(k)$ of nodes having that degree.

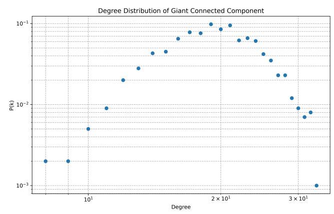


Figure 4: Node Degree Distribution Plot for Graph: er1

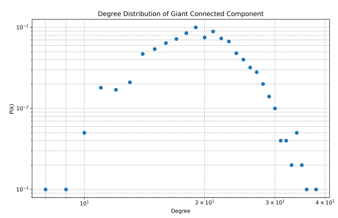


Figure 5: Node Degree Distribution Plot for Graph: er2

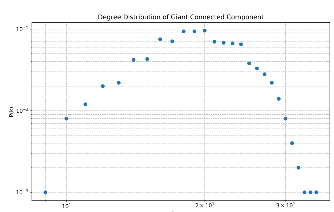


Figure 6: Node Degree Distribution Plot for Graph: er3

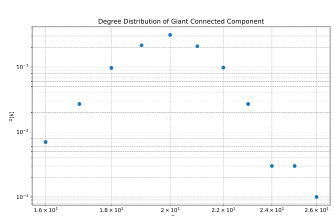


Figure 7: Node Degree Distribution Plot for Graph: ws1

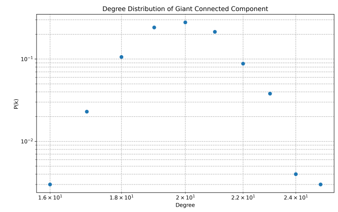


Figure 8: Node Degree Distribution Plot for Graph: ws2

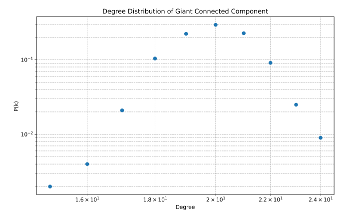


Figure 9: Node Degree Distribution Plot for Graph: ws3

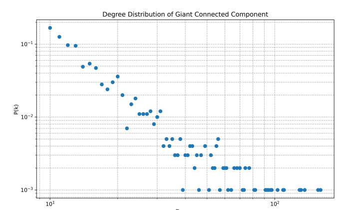


Figure 10: Node Degree Distribution Plot for Graph: ba1

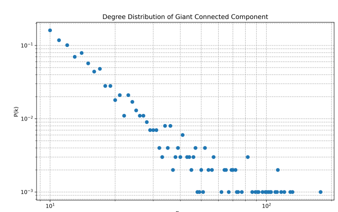


Figure 11: Node Degree Distribution Plot for Graph: ba2

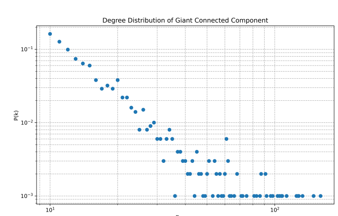


Figure 12: Node Degree Distribution Plot for Graph: ba3

3.2 Local Clustering Coefficient Distribution

The plots in this subsection display the distribution of local clustering coefficients for each generated graph. The x-axis represents the node degree, and the y-axis represents C_k, the average clustering coefficient of all nodes with degree k.

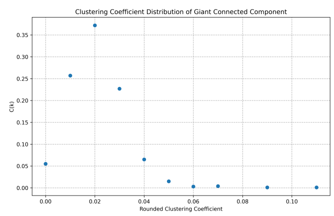


Figure 13: Local Clustering Coefficient Distribution Plot for Graph: er1

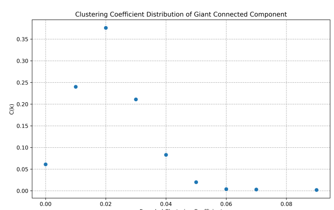


Figure 14: Local Clustering Coefficient Distribution Plot for Graph: er2

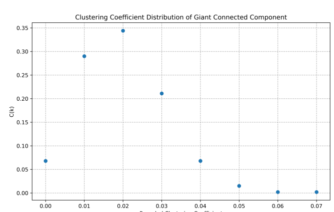


Figure 15: Local Clustering Coefficient Distribution Plot for Graph: er3

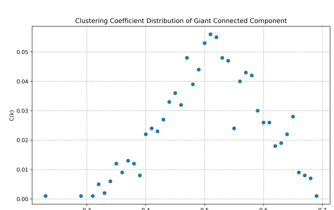


Figure 16: Local Clustering Coefficient Distribution Plot for Graph: ws1

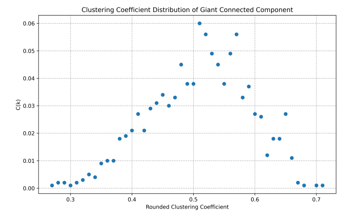


Figure 17: Local Clustering Coefficient Distribution Plot for Graph: ws2

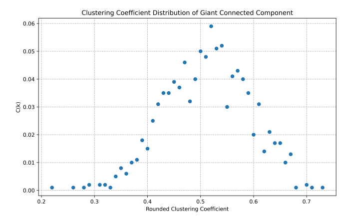


Figure 18: Local Clustering Coefficient Distribution Plot for Graph: ws3

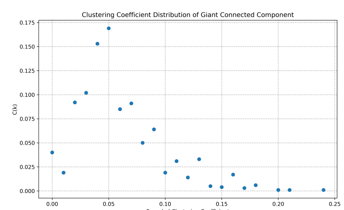


Figure 19: Local Clustering Coefficient Distribution Plot for Graph: ba1

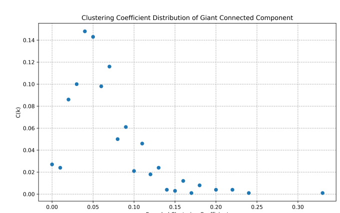


Figure 20: Local Clustering Coefficient Distribution Plot for Graph: ba2

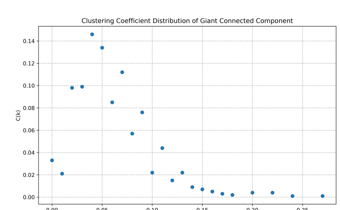


Figure 21: Local Clustering Coefficient Distribution Plot for Graph: ba3

3.3 Global Clustering Coefficient

The global clustering coefficient values for each graph are as follows (refer to Figures 1, 2, and 3):

- Erdős-Rényi Model:
  - er1: 0.0205
  - er2: 0.0212
  - er3: 0.0197
- Watts-Strogatz Model:
  - ws1: 0.5137
  - ws2: 0.5111
  - ws3: 0.5121
- Barabási-Albert Model:
  - ba1: 0.0584
  - ba2: 0.0606
  - ba3: 0.0597

3.4 Shortest Path Length Distribution

The plots in this subsection display the distribution of shortest path lengths for each generated graph. The x-axis represents the distance, and the y-axis represents the probability P(path length).

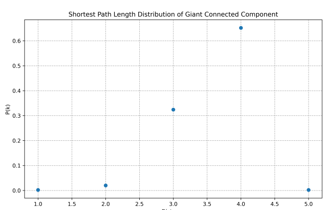


Figure 22: Shortest Path Length Distribution Plot for Graph: er1

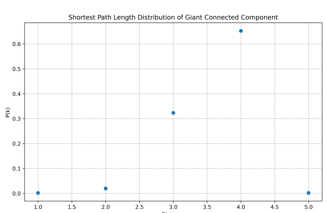


Figure 23: Shortest Path Length Distribution Plot for Graph: er2

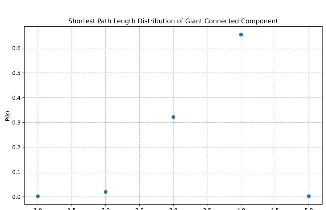


Figure 24: Shortest Path Length Distribution Plot for Graph: er3

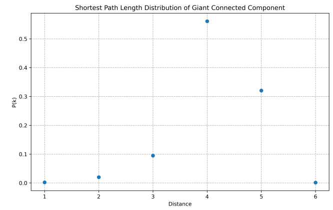


Figure 25: Shortest Path Length Distribution Plot for Graph: ws1

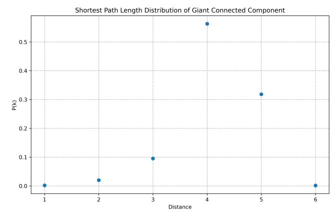


Figure 26: Shortest Path Length Distribution Plot for Graph: ws2

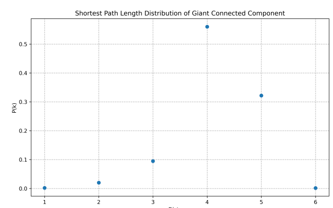


Figure 27: Shortest Path Length Distribution Plot for Graph: ws3

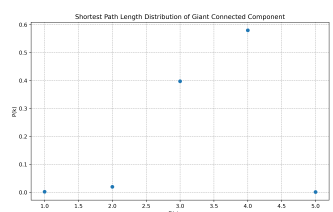


Figure 28: Shortest Path Length Distribution Plot for Graph: ba1

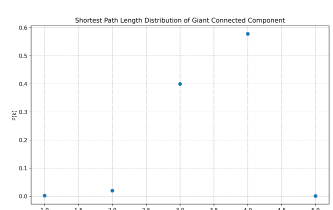


Figure 29: Shortest Path Length Distribution Plot for Graph: ba2

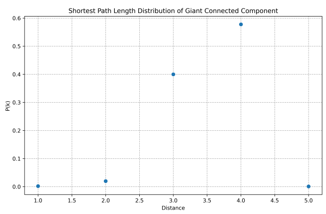


Figure 30: Shortest Path Length Distribution Plot for Graph: ba3

3.5 Average Shortest Path Length

The average shortest path length values for each graph are as follows (refer to Figures 1, 2, and 3):

- Erdős-Rényi Model:
  - er1: 2.6371
  - er2: 2.638
  - er3: 2.64
- Watts-Strogatz Model:
  - ws1: 3.1888
  - ws2: 3.1866
  - ws3: 3.1902
- Barabási-Albert Model:
  - ba1: 2.5633
  - ba2: 2.5608
  - ba3: 2.5666

3.6 Graph Diameter

The diameter values for each graph are as follows (refer to Figures 1, 2, and 3):

- Erdős-Rényi Model:
  - er1: 4
  - er2: 4
  - er3: 4
- Watts-Strogatz Model:
  - ws1: 5
  - ws2: 5
  - ws3: 5
- Barabási-Albert Model:
  - ba1: 4
  - ba2: 4
  - ba3: 4

4 DISCUSSION

Note: Although the parameters are the same, different networks are produced because the functions themselves generate random networks even with identical parameters.

4.1 Comparison within the Same Model

The graphs generated for each model exhibit consistent properties. The Erdős-Rényi model consistently showed very low clustering and low diameter across all graphs. The Watts-Strogatz model had the highest diameter among the models and showed no variation in the number of edges, highlighting its small-world characteristics. The Barabási-Albert model had the smallest average shortest path length and consistent edge counts, demonstrating its scale-free nature. These consistent results across repeated experiments indicate the reliability of the generation and analysis methods used.

4.2 Comparison between Different Models

The three different models exhibit both similarities and differences. All models had low diameters of four or five, indicating compact network structures. Despite having the same number of nodes and edges, the visual appearance of the graphs varied. The Watts-Strogatz model had the highest average shortest path length, emphasizing its small-world properties. In contrast, the Barabási-Albert model displayed a distinct node degree distribution, following a power-law distribution due to preferential attachment, leading to the emergence of hubs. This difference is crucial for understanding the structural diversity and potential applications of each model in real-world scenarios.

5 CONCLUSION

In this assignment, we generated and analyzed networks using three different models: Erdős-Rényi, Watts-Strogatz, and Barabási-Albert. By examining key properties such as node degree distribution, clustering coefficients, shortest path lengths, and graph diameter, we gained insights into the structural characteristics of each model. Our findings highlighted the unique features of each graph model and their potential applications in various domains.

The Erdős-Rényi model, with its random connections, provided a baseline for understanding simple random networks. The Watts-Strogatz model, known for its small-world properties, demonstrated high clustering and short path lengths, making it suitable for modeling social networks. The Barabási-Albert model, which generates scale-free networks, illustrated the presence of hubs and the power-law degree distribution, reflecting many real-world networks' behavior.

For the generation and analysis of the networks, we utilized the NetworkX library, which provided robust tools for graph creation and manipulation. Additionally, the course materials and resources from the EECS 4414 course at York University were instrumental in guiding our methodology and analysis.

References:

- https://networkx.org/
- https://www.eecs.yorku.ca/~papaggel/courses/eecs4414/