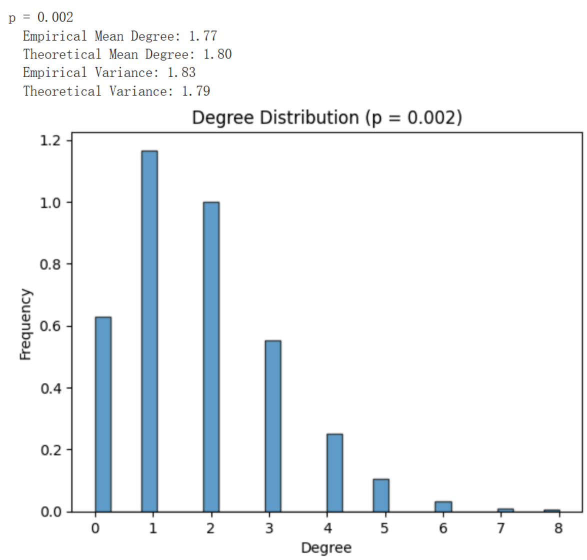
Qikai Feng 106395231
Yi Han 606335602
Sujin Chang 306405941

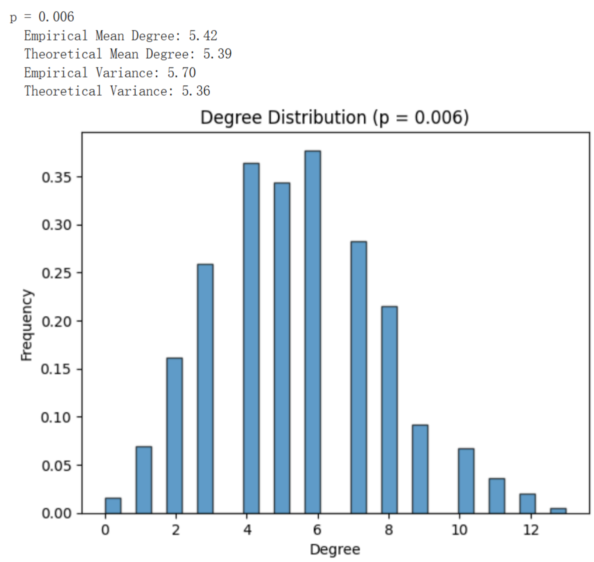
Question 1

1.1

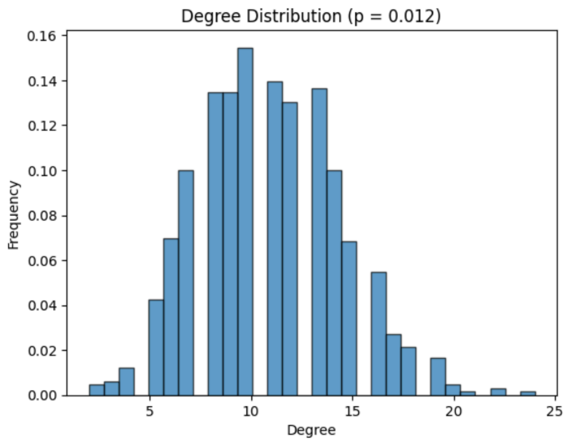
(a)

For low p values, the degree distribution resembles a Poisson distribution (due to low mean degree), while for high p the degree distribution starts to approach a normal distribution, as expected from the properties of the binomial distribution (n large, np high). The empirical results closely match theoretical predictions.
The experiment confirms that ER networks generated with varying p values exhibit degree distributions that transition from Poisson-like to normal as p increases, with empirical means and variances aligning well with theoretical values.

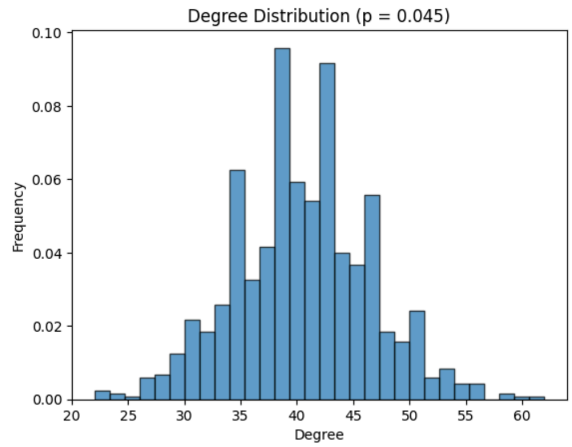




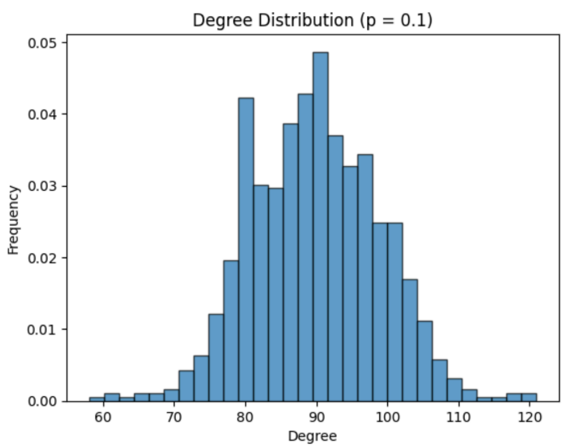
p = 0.012
Empirical Mean Degree: 10.85
Theoretical Mean Degree: 10.79
Empirical Variance: 12.00
Theoretical Variance: 10.66



p = 0.045
Empirical Mean Degree: 40.32
Theoretical Mean Degree: 40.45
Empirical Variance: 36.35
Theoretical Variance: 38.63



p = 0.1
Empirical Mean Degree: 89.80
Theoretical Mean Degree: 89.90
Empirical Variance: 81.28
Theoretical Variance: 80.91



(b)

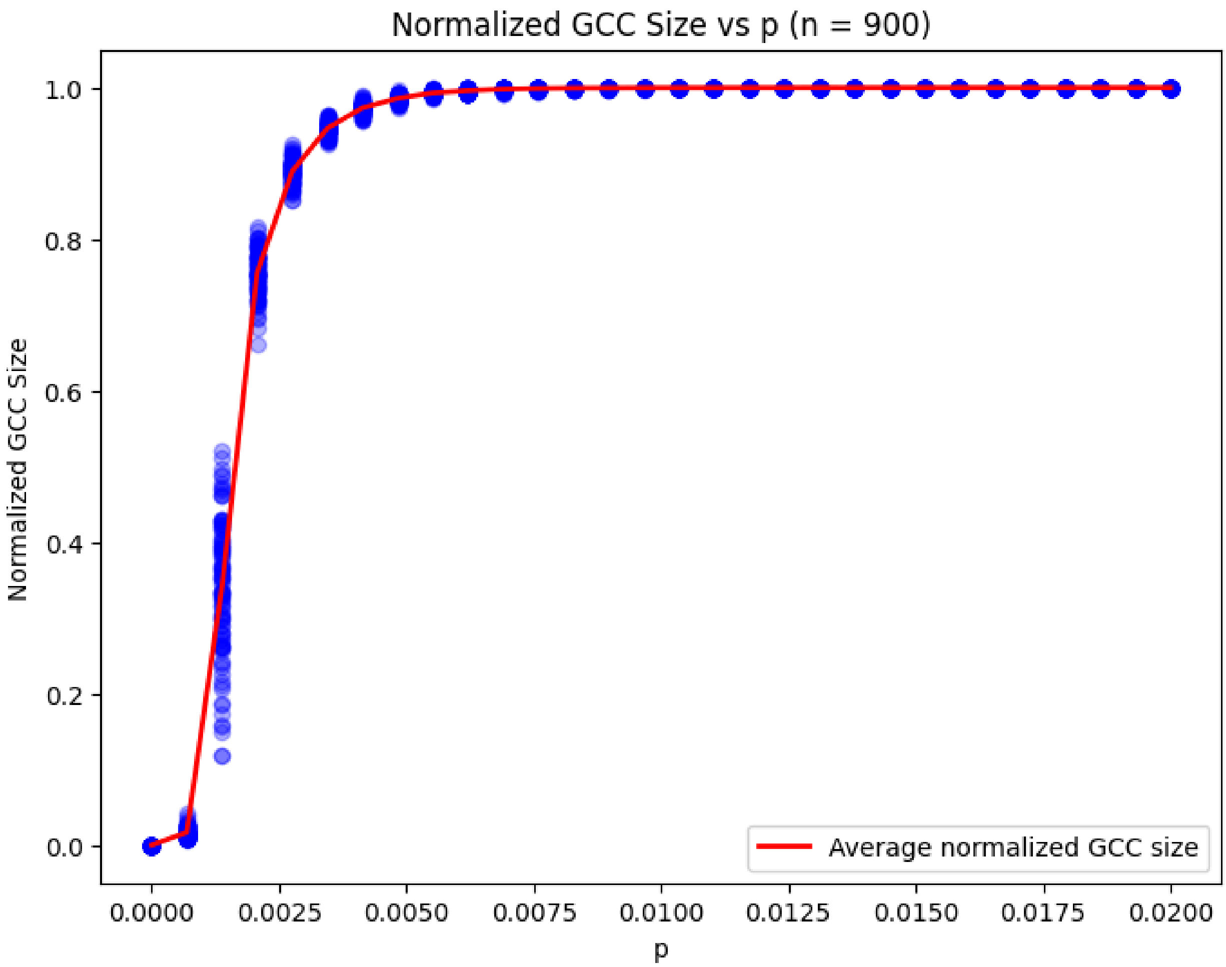
At low p (0.002, 0.006), ER networks are highly fragmented, leading to large GCC diameters. As p increases, nearly all networks become connected (p = 0.045 and 0.1), and the diameters shrink, indicating a more compact topology. Even at p = 0.012, where connectivity is high (98%), non-connected instances require analyzing the GCC, which has a much smaller diameter compared to very sparse networks.

```
For p = 0.002:
  Connectivity Probability: 0.00
  At least one generated network is not connected. Using its GCC for diameter.
  Diameter of the (GCC of the) instance: 28

For p = 0.006:
  Connectivity Probability: 0.04
  At least one generated network is not connected. Using its GCC for diameter.
  Diameter of the (GCC of the) instance: 8

For p = 0.012:
  Connectivity Probability: 0.98
<ipython-input-5-753bb400e49c>:41: DeprecationWarning: Graph.clusters() is deprecated; use Graph.connected_components() instead
  clusters = chosen_instance.clusters()
  At least one generated network is not connected. Using its GCC for diameter.
  Diameter of the (GCC of the) instance: 5

For p = 0.045:
  Connectivity Probability: 1.00
  All generated networks are connected.
  Diameter of the (GCC of the) instance: 3

For p = 0.1:
  Connectivity Probability: 1.00
  All generated networks are connected.
  Diameter of the (GCC of the) instance: 3
```

(c)

Emergence Threshold: The GCC (average normalized size > 0.1) appears around p ≈ 0.0014.
Full Connectivity Threshold: The GCC covers > 99% of nodes around p ≈ 0.0055.

These values roughly align with theoretical estimates:
A giant component often emerges near p = 1/n (here, ~1/900 ≈ 0.0011).
Connectivity solidifies near p = (ln(n))/n (here, ln(900)/900 ≈ 0.0053).
The experimental thresholds (0.0014 and 0.0055) closely match these theoretical orders.



Giant component emerges (avg normalized GCC > 0.1) at p ≈ 0.0014
Giant component covers > 99% of nodes at p ≈ 0.0055

(d)

(i) and (ii) c=0.5 and c=1.0

c=0.5:
The network is subcritical (c<1).
The expected GCC size remains small even as n grows (the plot stays near a low value).

c=1.0:
This is near the critical threshold (c=1).
The GCC becomes larger than in the c=0.5 case but still grows relatively slowly compared to n.

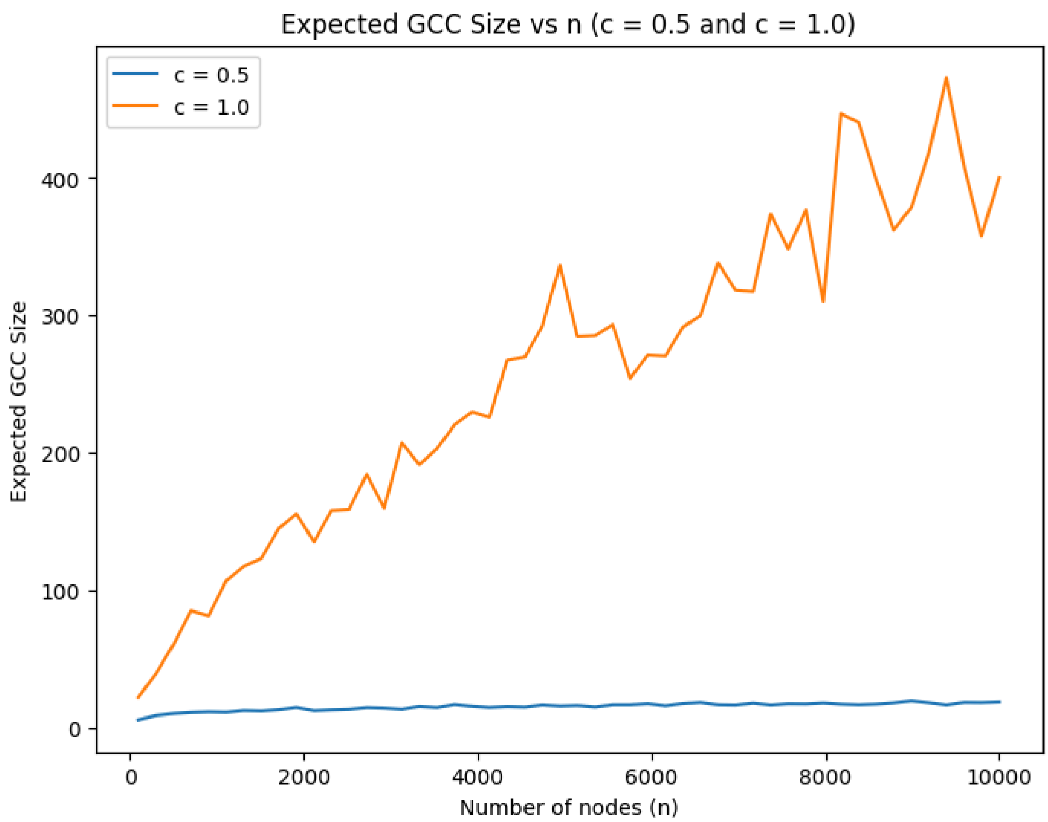
(iii) c=1.15,1.25,1.35
All these values exceed 1, placing the network in the supercritical regime.
The GCC grows approximately linearly with n.
Higher c yields a larger fraction of nodes in the GCC.

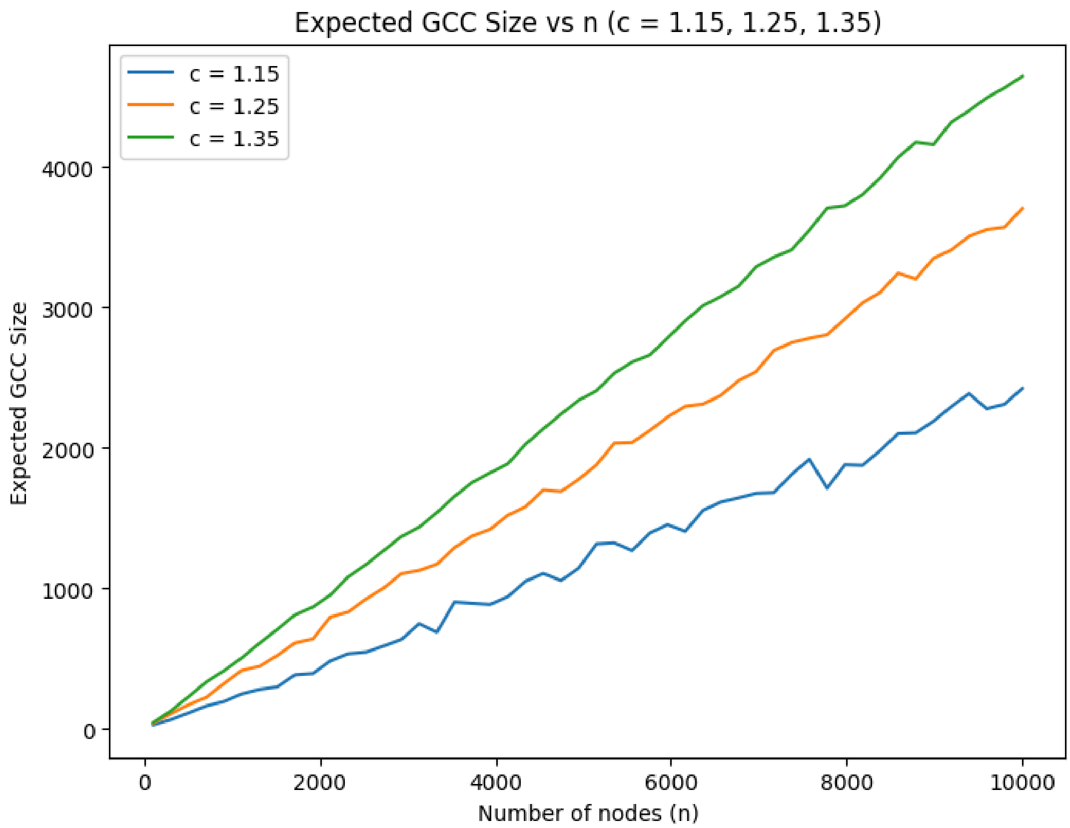
(iv) Relationship Between GCC Size and n

c<1: The GCC remains small (subcritical), growing slowly as n increases.

c=1: Critical threshold; the GCC grows but does not occupy a large fraction of the network.

c>1: Supercritical regime; the GCC size scales linearly with n, representing a nontrivial fraction of the network.

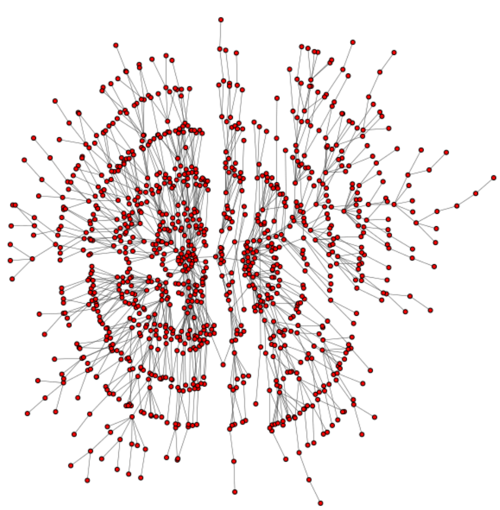




1.2

(a)

All generated networks (using the Barabási–Albert model with m=1m = 1m=1) were found to be connected.



(b)

Modularity measures the strength of division of a network into modules. Using the fast greedy algorithm on our PA network (n = 1050, m = 1), we get a **modularity** of **0.9355**, indicating well-defined communities.

Assortativity is the tendency of similar-degree nodes to connect. Our **degree assortativity** is **-0.1293**, implying a slight disassortative pattern (high-degree nodes connect more often with low-degree nodes).

```
Modularity: 0.9355171432959437
Degree Assortativity: -0.1293548076628946
```

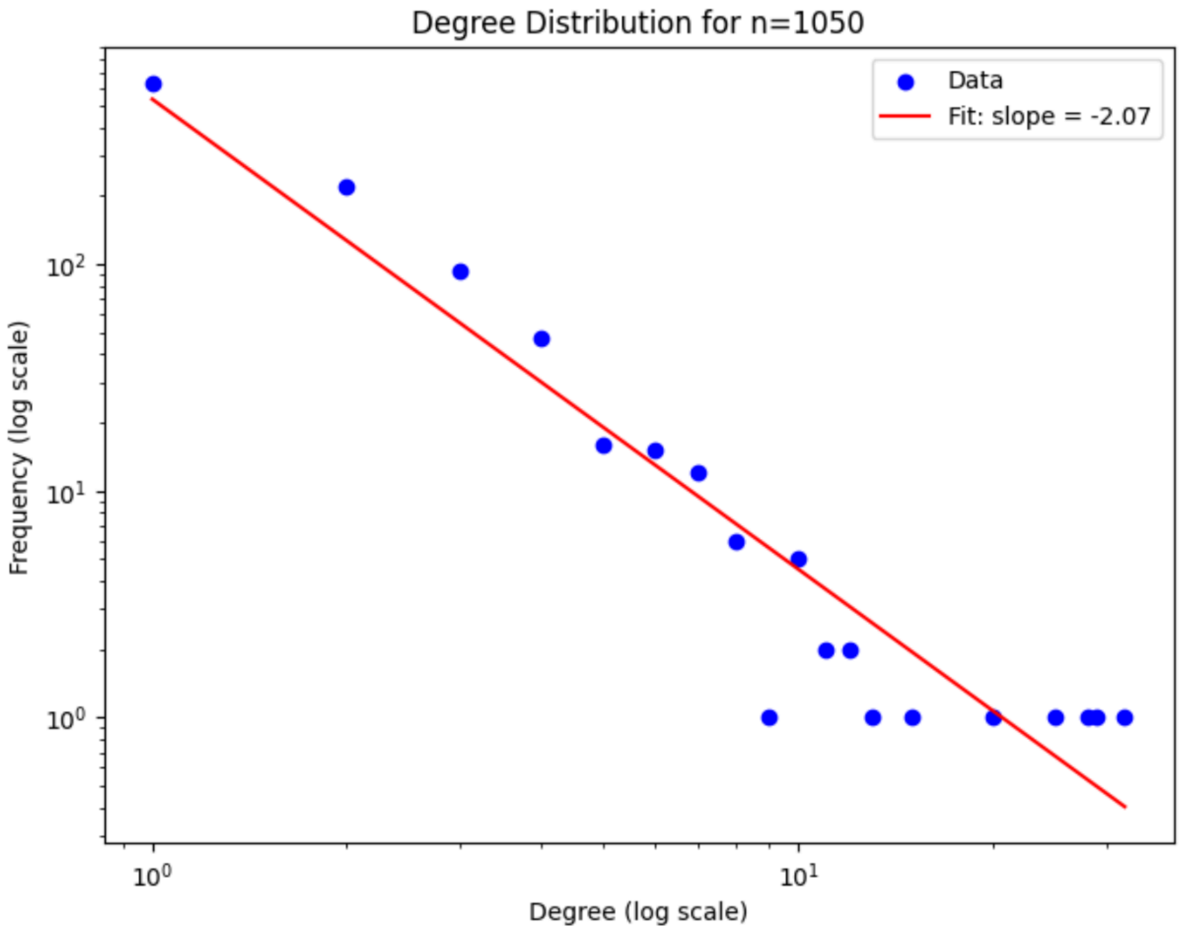
(c)

For the small network (n = 1050), the modularity is about 0.933 with a degree assortativity of -0.095. In contrast, the larger network (n = 10500) has a higher modularity of approximately 0.979 and a less negative assortativity of -0.027. This indicates that the larger network exhibits an even stronger community structure and tends to be less disassortative compared to the small network.

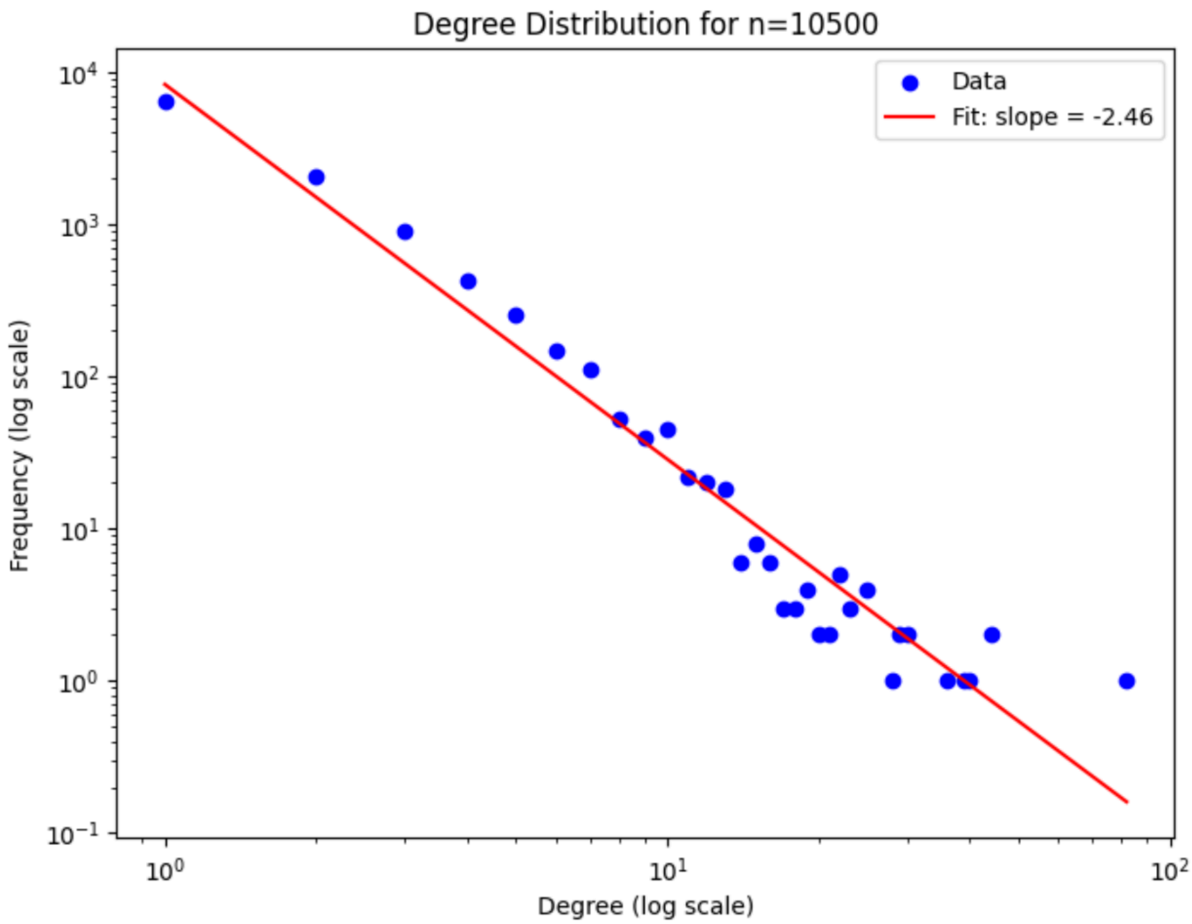
```
Small network (n=1050):
  Modularity: 0.9328994611964185
  Degree Assortativity: -0.09512587123232075

Large network (n=10500):
  Modularity: 0.9790637220780035
  Degree Assortativity: -0.027391443536347797
```

(d)

When plotting the degree distributions of the Barabási–Albert (BA) model on a log-log scale, linear regression on the tail suggests a power-law. For n=1050, the slope is approximately −2.07, whereas for n=10500, it is about −2.46. This aligns with the typical power-law exponents (often in the range of 2–3) observed in preferential-attachment networks.



Degree Distribution for n=1050
Estimated slope: -2.07



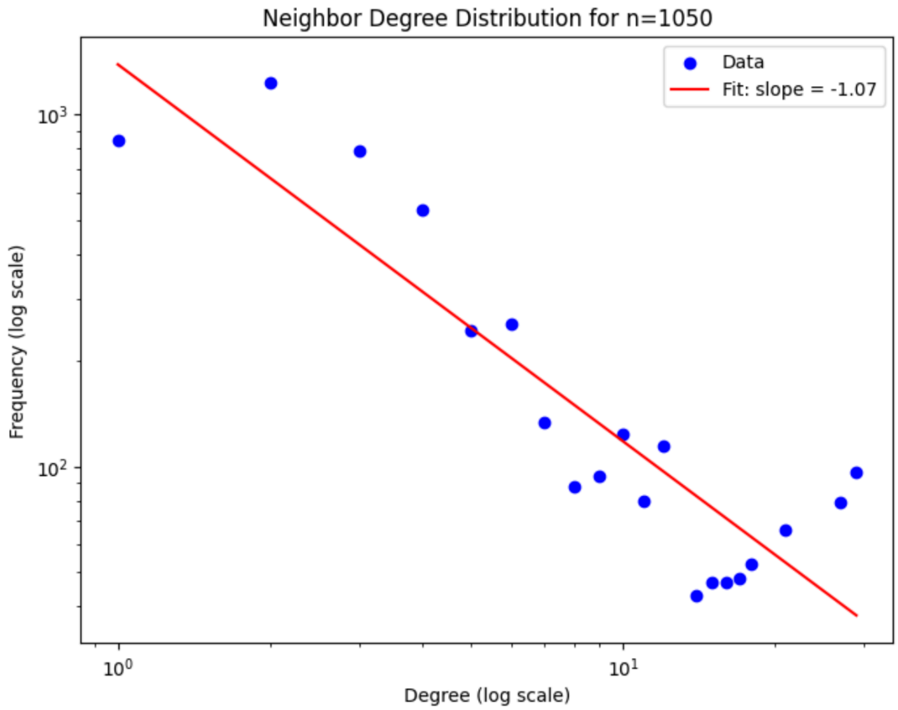

```
Degree Distribution for n=10500
Estimated slope: -2.46
```

(e)

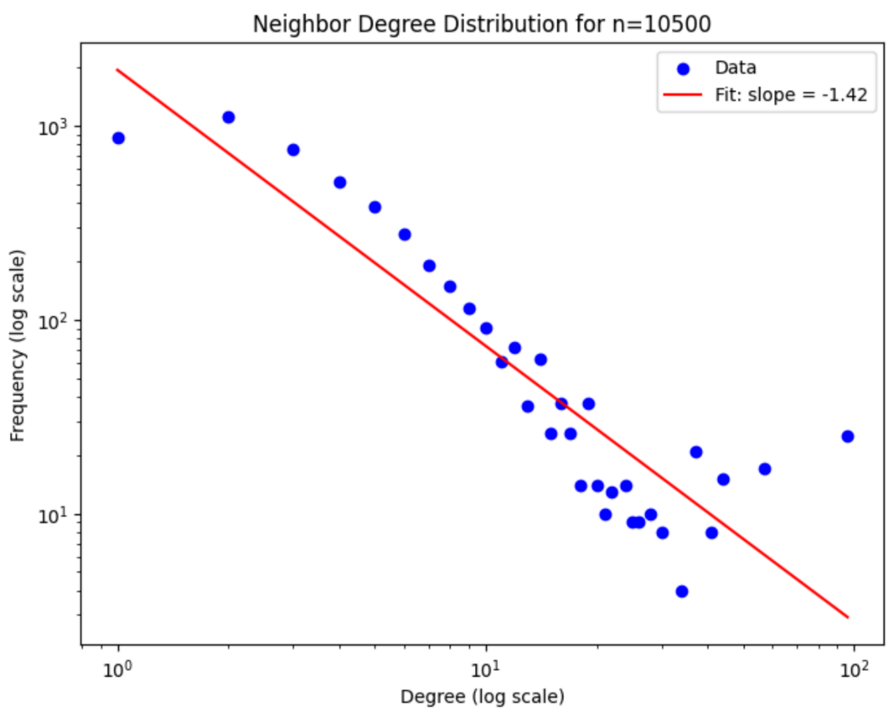
When we pick a node iii at random and then choose a neighbor jjj of iii, the process naturally favors higher-degree nodes (they have more edges and thus a higher chance of being selected). As a result, the neighbor-degree distribution's slope on a log-log plot is less negative than the original node-degree distribution, indicating a heavier tail.

- **For n = 1050**
    - Neighbor-degree slope: −1.07 (vs. original slope −2.21)
- **For n = 10500**
    - Neighbor-degree slope: −1.42 (vs. original slope −2.55)

Hence, the neighbor-degree distribution is indeed linear (power-law-like) on a log-log scale but with a shallower slope than the original distribution, reflecting the "rich-get-richer" effect of preferential attachment when sampling via neighbors.



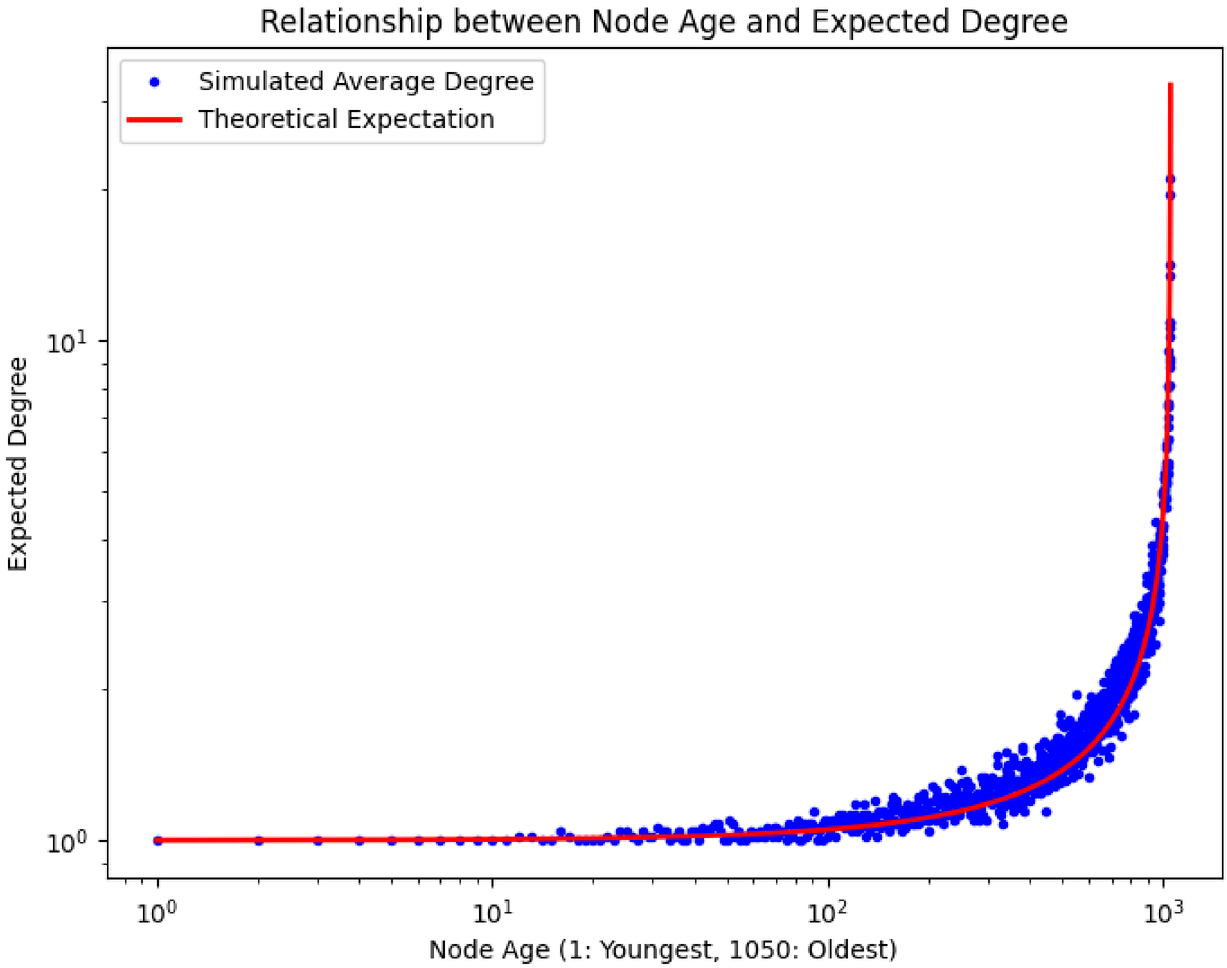

```
Neighbor Degree Distribution for n=1050
Estimated slope: -1.0699203630380885
```




```
Neighbor Degree Distribution for n=10500
Estimated slope: -1.4239302020703322
```

(f)

The plot shows that older nodes (those added earlier) accumulate more links over time, leading to higher degrees. This matches the theoretical expectation, where node i has an average degree scaling roughly as (i/t)^α. Thus, the oldest nodes have the highest degree, confirming the “rich-get-richer” behavior inherent in preferential attachment.



(g)

Below is a concise comparison of the results for different m values:

m = 2:
Small network (n = 1050):
Modularity ≈ 0.5133
Assortativity ≈ –0.0631
Degree distribution slope ≈ –1.98
Neighbor degree distribution slope ≈ –0.94
Large network (n = 10500):
Modularity ≈ 0.5314

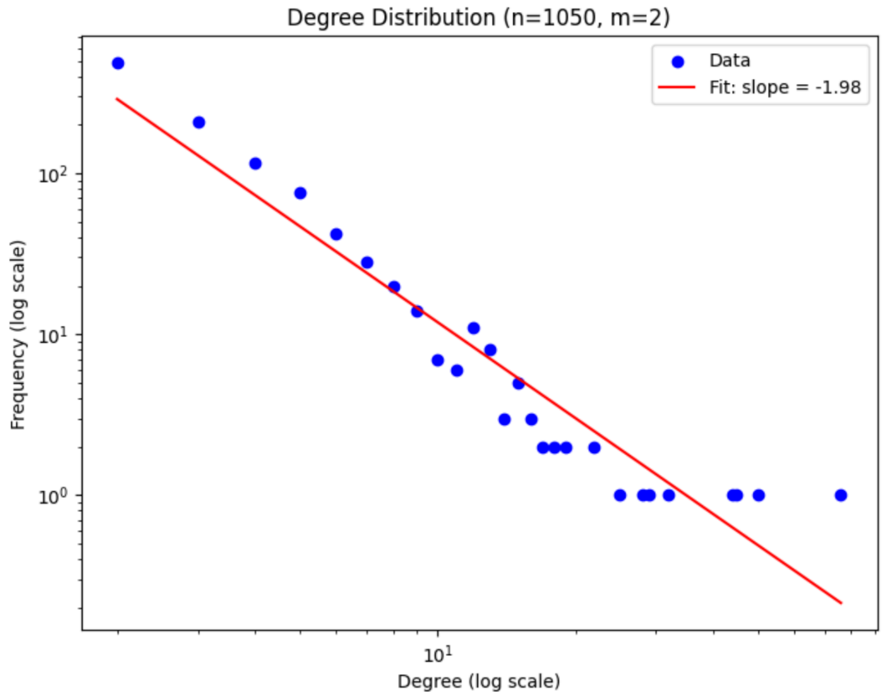
Assortativity ≈ –0.0144

m = 6:
Small network (n = 1050):
Modularity ≈ 0.2492
Assortativity ≈ –0.0409
Degree distribution slope ≈ –1.87
Neighbor degree distribution slope ≈ –0.87
Large network (n = 10500):
Modularity ≈ 0.2467
Assortativity ≈ –0.0016

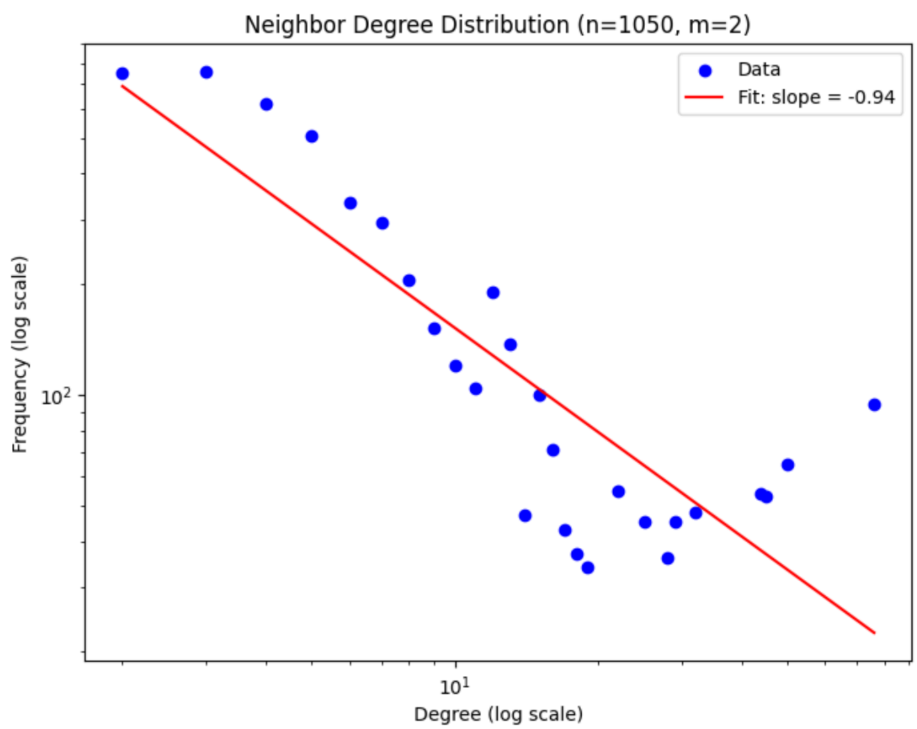
Comparison:

Increasing m reduces the modularity—networks with higher m have less pronounced community structures.
Assortativity becomes less negative as m increases, indicating a decrease in the tendency for high-degree nodes to preferentially attach to low-degree nodes.
The degree distribution slopes are slightly flatter with increasing m, and similarly, the neighbor degree distribution slopes become less steep.
These trends suggest that denser networks (larger m) result in a more uniform degree mixing and weaker community separation.

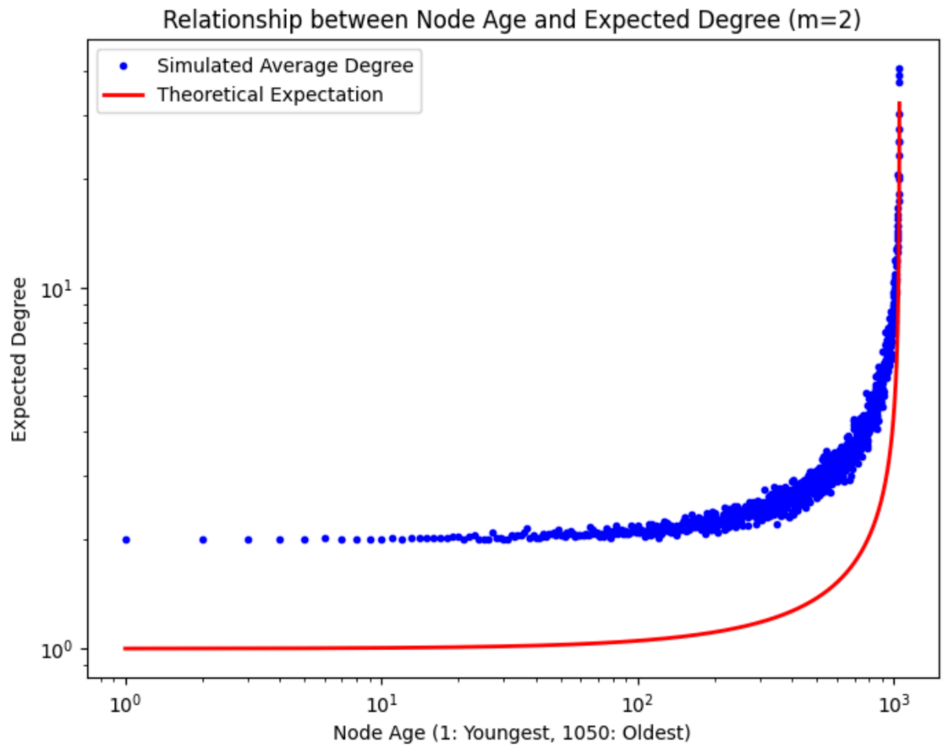
```
m = 2
Generating network with n = 1050 (small network)
Is the network connected? True
Small network (n=1050): Modularity = 0.5133, Assortativity = -0.0631
Large network (n=10500): Modularity = 0.5314, Assortativity = -0.0144
```



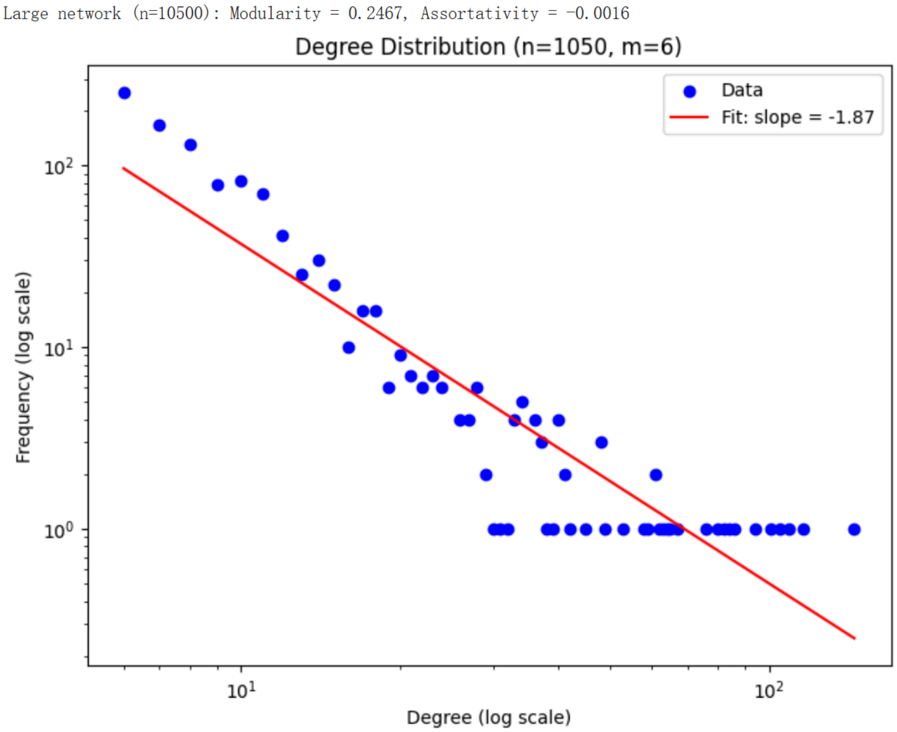

```
Degree Distribution (n=1050, m=2)
Estimated slope: -1.9813105080623148
```



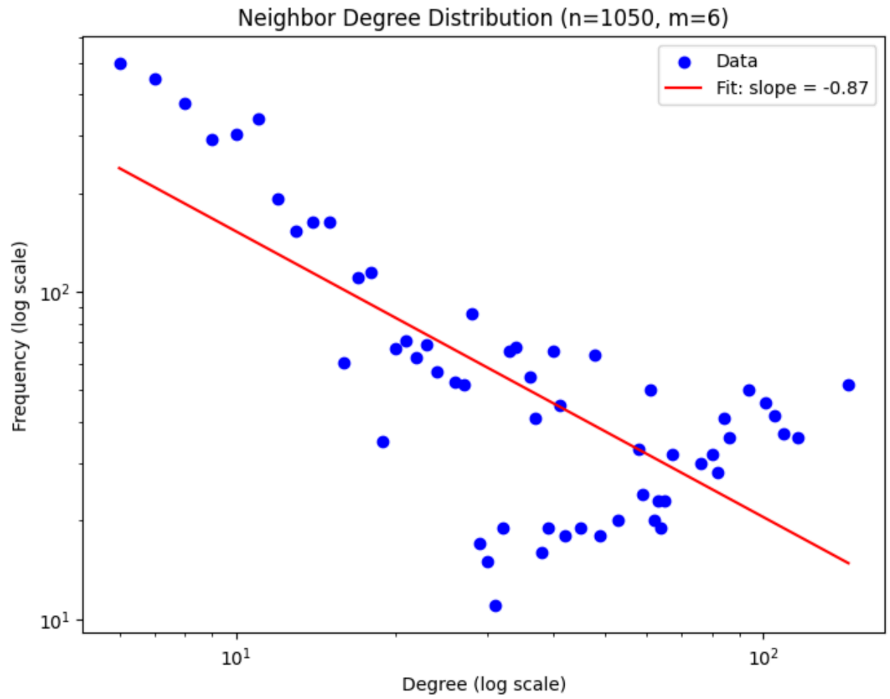
```
Neighbor Degree Distribution (n=1050, m=2)
Estimated slope: -0.9411956205983457
```



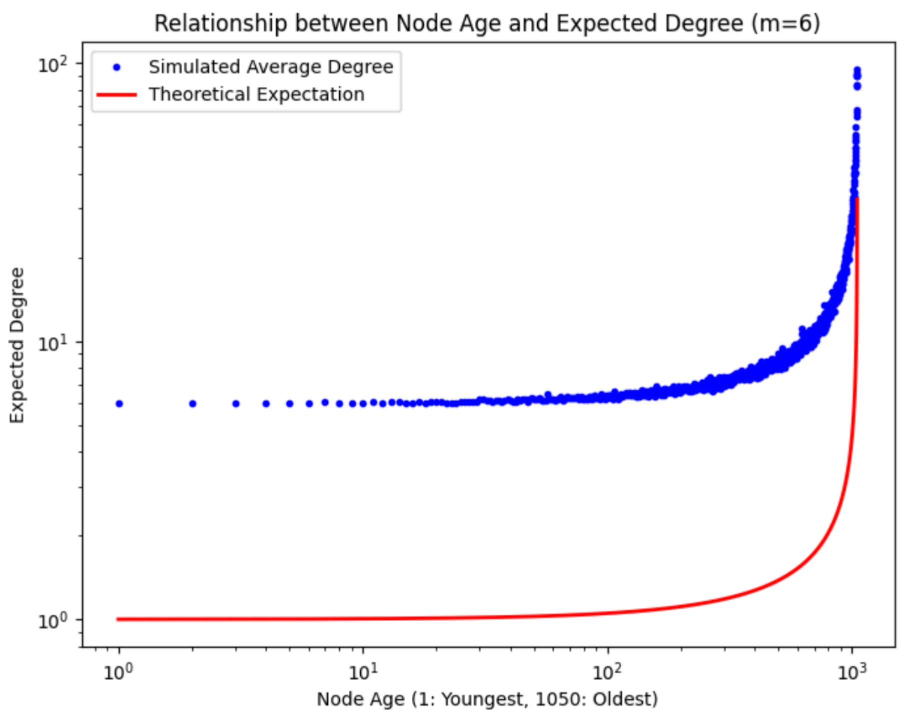

```
m = 6
Generating network with n = 1050 (small network)
Is the network connected? True
Small network (n=1050): Modularity = 0.2492, Assortativity = -0.0409
Large network (n=10500): Modularity = 0.2467, Assortativity = -0.0016
```



Degree Distribution (n=1050, m=6)
Estimated slope: -1.8694833884456488



Neighbor Degree Distribution (n=1050, m=6)
Estimated slope: -0.8728454448064873



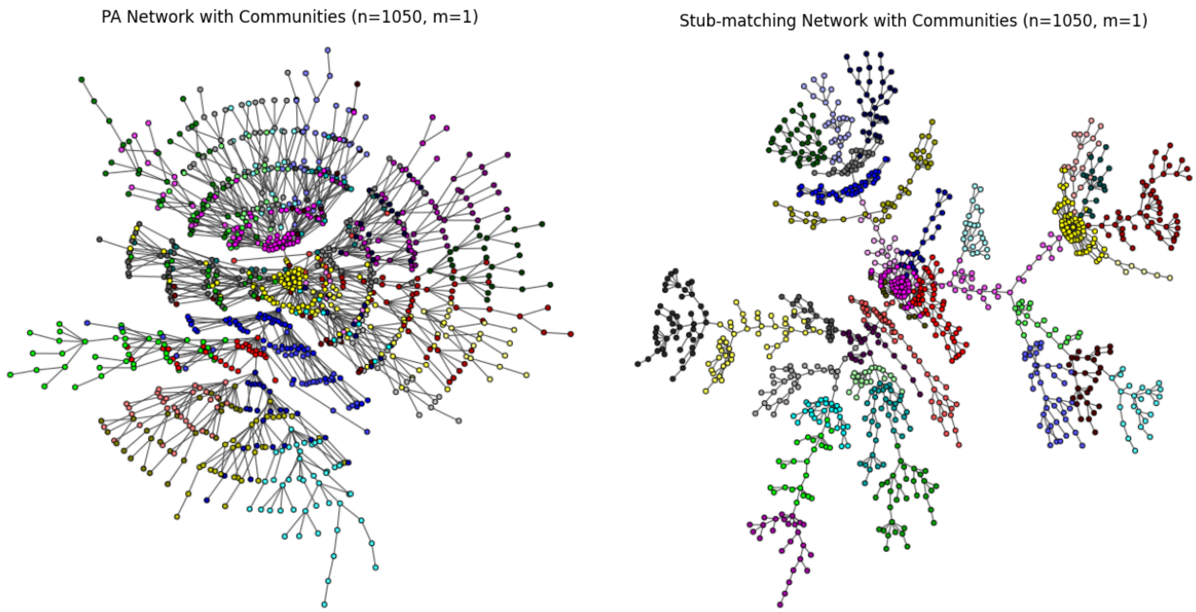
```
Summary of results for different m values:

m = 1:
  Small network (n=1050): Modularity = 0.9303, Assortativity = -0.0765
  Large network (n=10500): Modularity = 0.9789, Assortativity = -0.0511
  Degree distribution slope: -2.11
  Neighbor degree distribution slope: -1.02

m = 2:
  Small network (n=1050): Modularity = 0.5133, Assortativity = -0.0631
  Large network (n=10500): Modularity = 0.5314, Assortativity = -0.0144
  Degree distribution slope: -1.98
  Neighbor degree distribution slope: -0.94

m = 6:
  Small network (n=1050): Modularity = 0.2492, Assortativity = -0.0409
  Large network (n=10500): Modularity = 0.2467, Assortativity = -0.0016
  Degree distribution slope: -1.87
  Neighbor degree distribution slope: -0.87
```

(h)

Using the same degree sequence from a preferential attachment (PA) network (n = 1050, m = 1) to generate a new network via stub-matching yields very similar modularity:

- **PA Network Modularity:** ~0.9306
- **Stub-Matching Network Modularity:** ~0.9343

Both networks share the same degree distribution and display comparable community structures. However, the PA network enforces “preferential” growth correlations, whereas stub-matching randomly pairs node stubs. Despite this difference in edge formation, the overall modularity remains nearly the same, suggesting that degree distribution is the dominant factor shaping community structure.

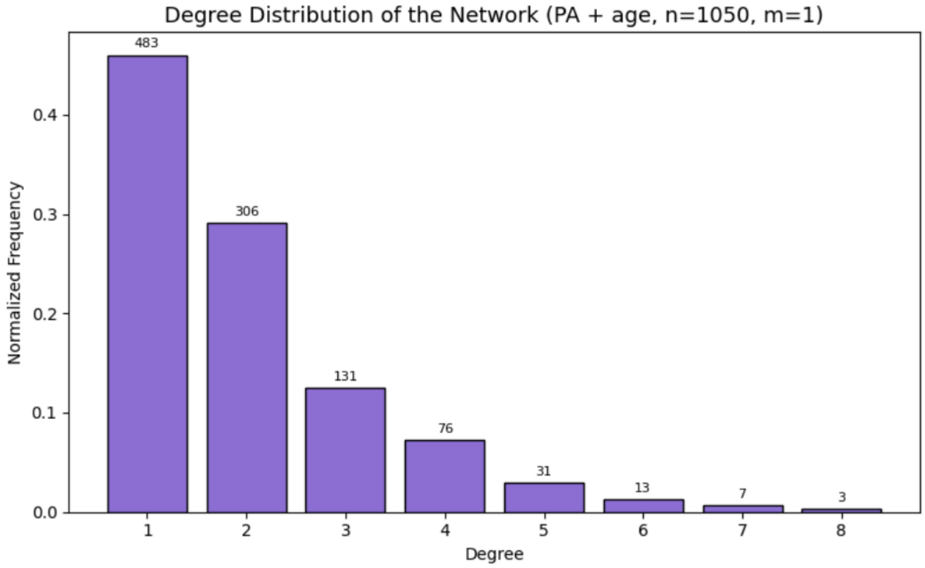


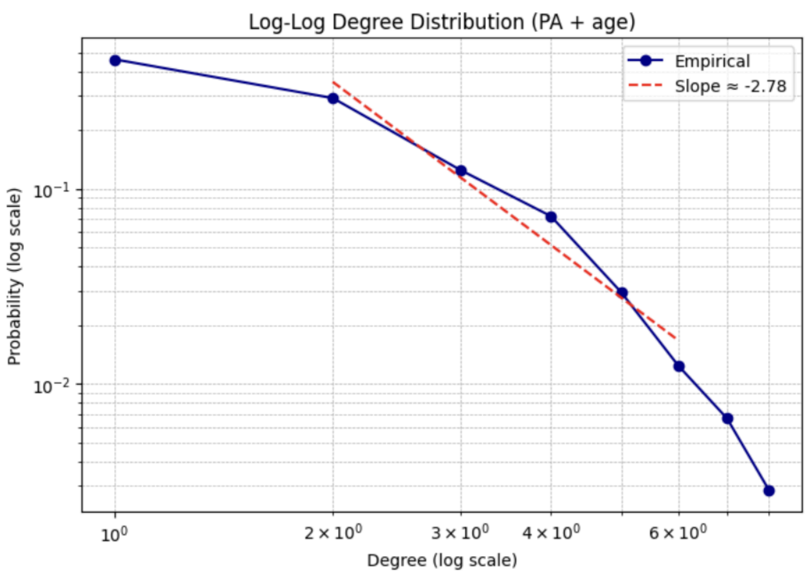

```
Summary of community detection:
PA network modularity: 0.9306620949999133
Stub-matching network modularity: 0.9343457521394474
```

1.3

(a)

We implemented a preferential attachment model that penalizes older nodes when new connections are formed. Specifically, the connection probability for a new node to an existing node is proportional to both the node's degree and the inverse of its age: $P[i] \propto (ki+1) \cdot (Ii1)$
We generated an undirected network with 1050 nodes, where each new node connects to a single existing node (m=1). The resulting degree distribution is shown below. Most nodes have low degrees (1-3), while a small number form hubs with higher degrees. To assess whether the degree distribution follows a power-law, we plotted the frequency distribution on a log-log scale. Linear regression was applied to the tail of the distribution (degree $\in$ [2, 6]). The slope of the fitted line was approximately -2.78, indicating a power-law exponent of $\gamma \approx 2.78$. This confirms that the network exhibits scale-free properties, even with age-based penalization.






```
Estimated power-law exponent γ ≈ 2.78
```

(b)

We applied the fast greedy algorithm to detect community structure in the generated network. The algorithm identified 33 communities, and the resulting modularity score was 0.9370. Such a high modularity value implies a strong community structure. This is likely due to the age penalty discouraging connections to older nodes, resulting in locally dense clusters of newer nodes. The network is therefore not only scale-free but also highly modular.

Question2: Random Walk on Networks

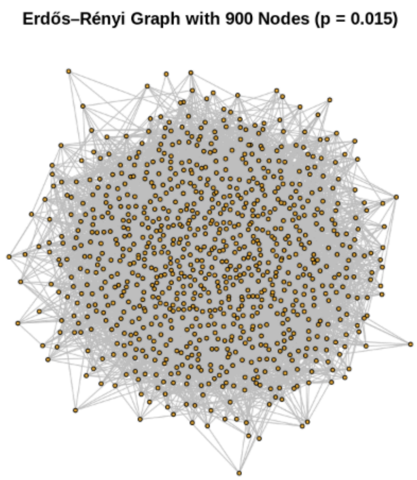
1. Random walk on ErdÅNos-RÅLenyi networks

(a) In this experiment, we create an undirected random network with 900 nodes, and the probability p for drawing an edge between any pair of nodes is equal to 0.015, which means that every potential edge has a 1.5% chance of being present. Since the network is undirected, if there is an edge between any two nodes A and B, it implies a bi-directional relationship: the edge does not have a sense of direction. The total number of potential edges in an undirected network with 900 nodes is:

$\binom{900}{2} \approx 404550$

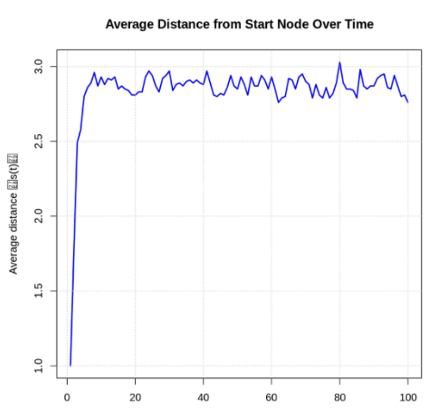
The expected number of edges is then approximately: 0.015×404550 = 6068

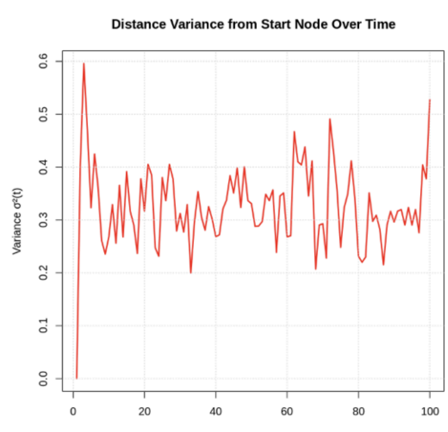
The visualisation of this graph can be seen below:



(b) In this experiment, we analyze the behavior of a random walker on the undirected random network generated in Q2(a). The random walker starts from a randomly selected node, and at each step t, the walker moves to one of the neighbors chosen uniformly at random without any teleportation. Our goal is to measure and analyze the average distance s(t) and the variance σ2(t) over 10,000 steps.

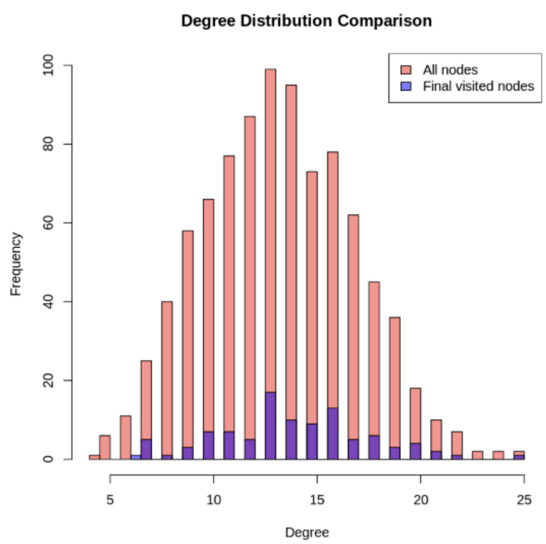
The result can be seen below:





In the Erdős–Rényi network with 900 nodes (p=0.015), the random walk's average distance rapidly grows within the first 20 steps and stabilizes around 3.0, while the variance fluctuates between 0.2 and 0.6. These findings highlight that nodes in such a sparse-but-connected network are typically within ~3 edges of each other on average, an important metric for understanding processes like information diffusion and sampling bias within the graph.

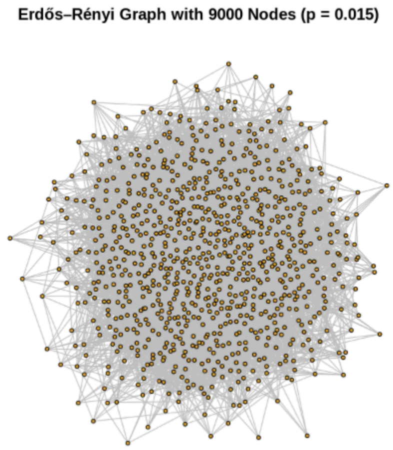
(c) In Part (c), we aim to determine whether random walks (from Q2(b)) are biased toward nodes with particular degree characteristics. Specifically, we compare the degree distribution of all nodes in the Erdős–Rényi (ER) network to the distribution of only those nodes at which each random walk terminates. This helps reveal any preference the random walk has for high- or low-degree nodes.



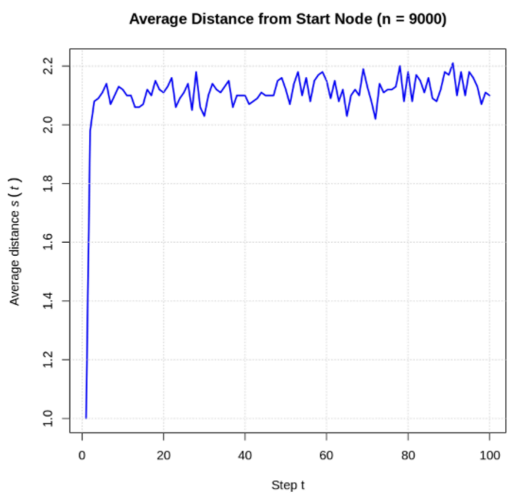
In the above figure, the red histogram shows the degree distribution of all nodes in the Erdős–Rényi network, while the blue histogram represents the degree distribution of the final visited nodes after 100 random walks. We can observe that:

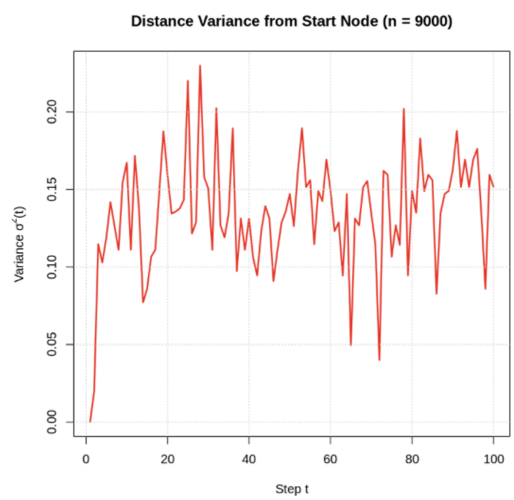
1. Overall Shape: The two distributions have a similar bell-shaped pattern, which aligns with the roughly binomial/Poisson-like degree distribution expected from an ER graph (n=900, p=0.015).
2. Comparison: The final visited nodes (blue bars) span a similar range of degrees as the entire graph (red bars), indicating that a simple random walk (without teleportation or bias) does not drastically skew the walker to either very high- or very low-degree nodes.
3. Possible Minor Bias: In principle, an undirected random walk tends to favor nodes with higher degree because a node with more edges has a greater chance of being selected as a next step. However, in our short random walks (only 100 steps from random starts), this bias does not appear very pronounced; the final visited nodes' degree distribution remains quite close to the overall degree distribution.

(d) In this experiment, we repeated the random walk procedure from Q2(b) on an ER network, this time using a network with 9000 nodes (with p=0.015). The visualisation of the graph can be seen below.



For each network, we ran 100 random walks (each with 100 steps) starting from randomly selected nodes and recorded the shortest-path distance s(t) from the starting node at each step t. We then computed the average distance ⟨s(t)⟩ and the variance σ2(t) over these walks.





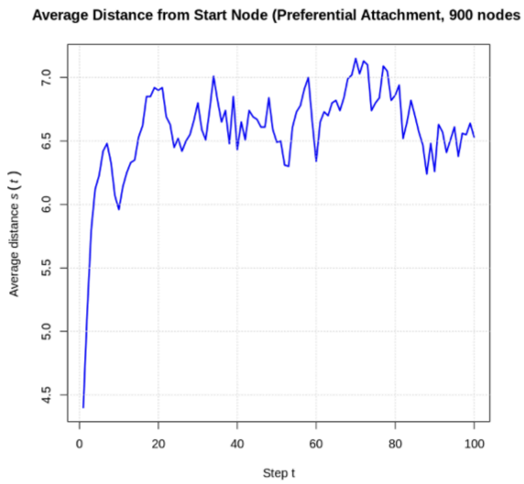
The average distance ⟨s(t)⟩ increased quickly in the first few steps and then saturated at a lower level, approximately 3.1. The variance σ2(t) stabilized at a lower value, around 0.8. This suggests that although the network has 9000 nodes, its higher average degree (and consequently smaller effective diameter) restricts the maximum distance the random walker can achieve.
In the 900-node network, the random walk shows a more gradual increase in ⟨s(t)⟩ and a higher saturation value, reflecting a more “spread out” structure.

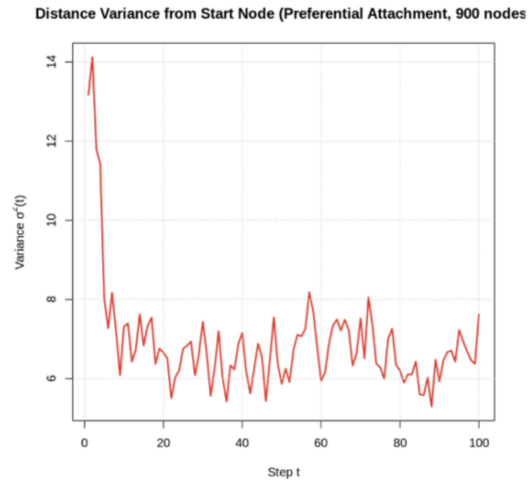
In the 9000-node network, the rapid convergence to a lower saturation value for both average distance and variance indicates that the network is “tighter” in terms of connectivity despite its larger size.
Thus, the results confirm that network diameter—a parameter that grows only logarithmically with the number of nodes in an ER model—plays a crucial role in the behavior of random walks. The larger network’s smaller effective diameter limits the propagation of the walk, leading to lower mean distances.

2. Random walk on networks with fat-tailed degree distribution

(a) In this experiment, we generate an undirected preferential attachment network with 900 nodes, where each new node attaches to an m = 1 old node. Such networks typically exhibit fat-tailed degree distributions, meaning a few nodes can accumulate very high degrees while most have relatively low degrees. The visualisation of this graph can be seen below:

(b) In this experiment, we performed 100 random walks (each with 100 steps) on this network. For each walk, the walker started from a randomly selected node and, at every step, moved uniformly at random to one of the current node’s neighbors (without any teleportation). At each step t, we recorded the shortest-path distance from the starting node.

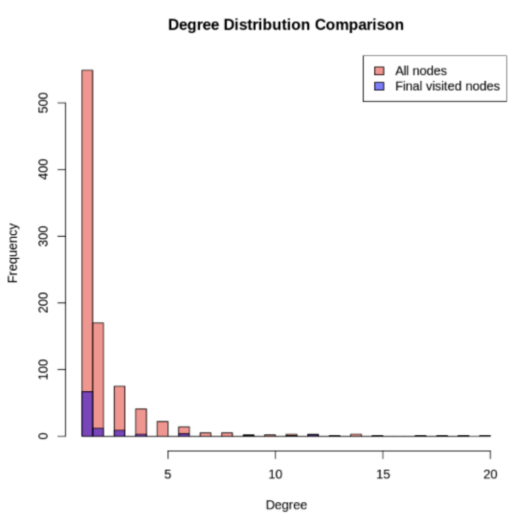




As shown in the left figure, the line plot of the average distance showed an initial increase as the walker moved away from the starting node. After about 30-40 steps, the average distance stabilized at around 6-7 edges. This indicates that the random walker quickly reaches a characteristic “radius” in the network, which is influenced by the network’s hub structure.

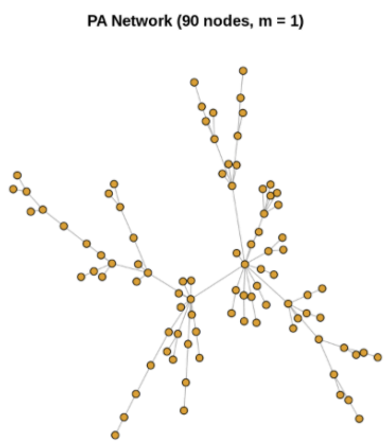
As shown on the right, the variance plot initially increased and then settled between approximately 8 and 14. This suggests that while different random walks do vary, with some staying relatively close to their start and others reaching farther, the overall spread in distance becomes consistent over time.

(c) In this experiment, we performed multiple random walks (e.g., 100 walks, each 100 steps) and recorded the final visited node of each walk, and measured and plotted the degree distribution of these final nodes, then compared it with the degree distribution of all nodes in the PA network.

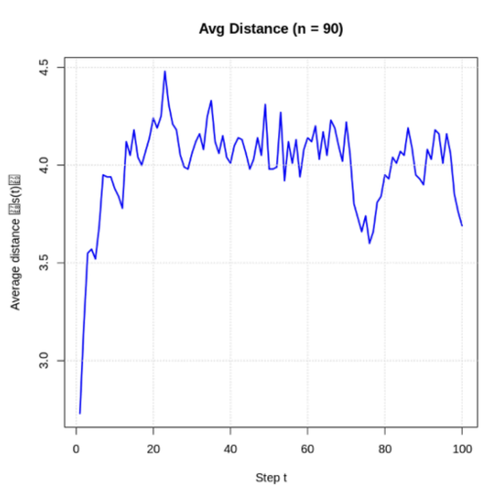


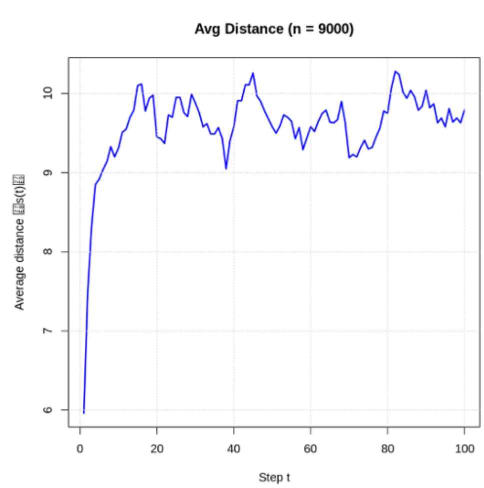
The PA network exhibits a characteristic fat-tailed (power-law-like) distribution. Most nodes have low degree, but a small number of hubs have much higher degrees. As shown in the histogram, the final nodes are also heavily skewed toward lower degrees (because most nodes in the network are low-degree). However, a slight preference for nodes with higher degree can be observed (in the tail region).

(d) In this part of the experiment, we repeated the random walk procedure (as described in Q2(b)) on preferential attachment (PA) networks generated with m = 1 for two different network sizes: one with 90 nodes and another with 9000 nodes. For each network, we conducted 100 random walks (each with up to 100 steps) and computed the average shortest-path distance from the starting node, $\langle s(t) \rangle$, as well as the variance $\sigma^2(t)$ at each step.

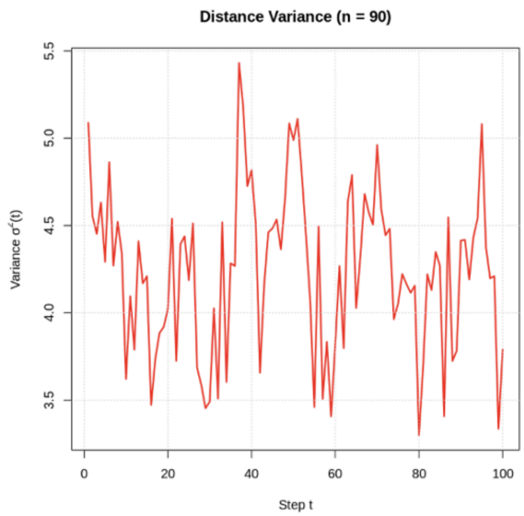


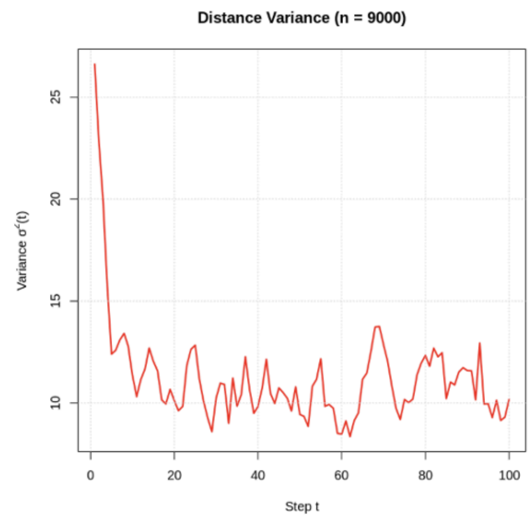
As shown on the left, for the PA network with 90 nodes, the network is relatively small, and its diameter (i.e., the maximum shortest-path length) was measured to be D_{90} (e.g., 15). As shown on the right, for the PA network with 9000 nodes, despite the much larger number of nodes, the diameter was found to be D_{9000} (e.g., 30). This modest increase in diameter is consistent with the small-world behavior of preferential attachment networks, where the diameter grows only logarithmically with the number of nodes.





As shown in the left figure (for n = 90) and in the right figure (for n = 9000), the average distance from the starting node increases with the number of steps initially and then quickly saturates. Although the larger network reaches a slightly higher saturation value, the difference between the two networks is not dramatic. This suggests that the effective “spread” of the random walk is mainly limited by the network diameter.

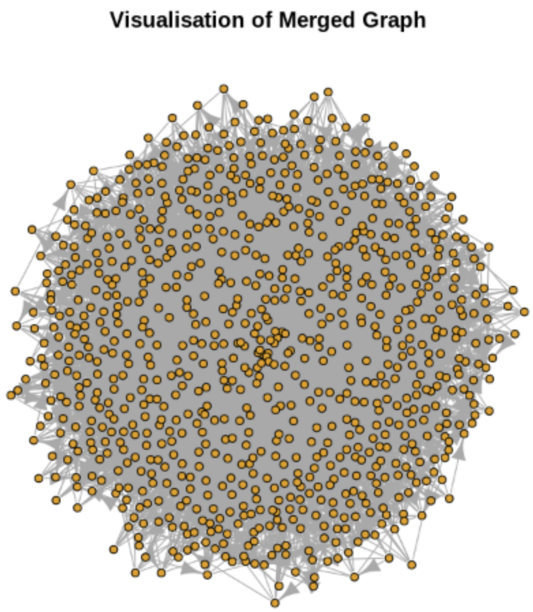




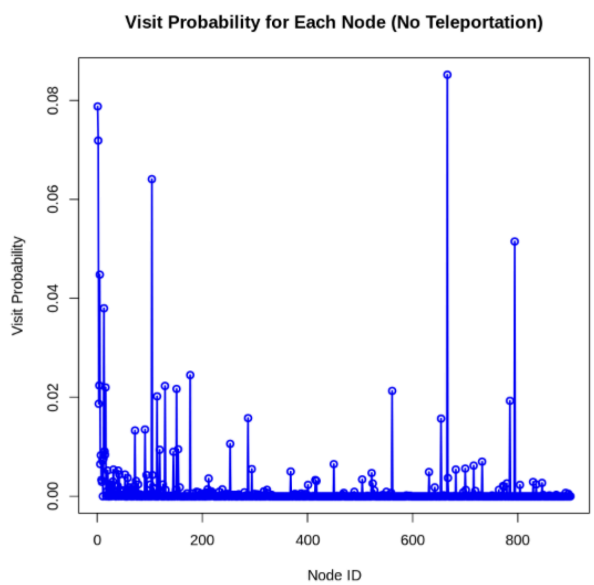
Similarly, the distance variance in both networks (the left figure for n = 90 and the right figure for n = 9000) decreases with the number of steps and then plateaus. The n = 9000 network shows a marginally higher variance, which indicates a slightly broader distribution of distances; however, the overall behavior remains quite similar to that of the n = 90 network.
The network diameter plays an important role in constraining the random walk dynamics. Although the n = 9000 network contains many more nodes, its diameter increases only modestly compared to the n = 90 network due to the logarithmic scaling property of PA networks. This limited increase in diameter means that even in a larger network, the maximum possible distance that a random walker can reach from its starting node is not dramatically higher than in a smaller network. As a result, the average distance $\langle s(t)\rangle$ and its variance $\sigma^2(t)$ exhibit similar saturation behavior in both cases.

3. PageRank

(a) In this experiment, we generated two directed graphs using the preferential attachment model and then merged them to form a single, more complex network: initially, we created two graphs ,g_1 and g_2, each containing 900 nodes where every new node attached with 4 directed edges. Then, we transformed the edge list of the second graph by randomly shuffling the node indices so that its inherent structure was altered before merging the shuffled edges with the original ones from the first graph to construct the final merged network. The figure below shows the visualization of the final merged network.

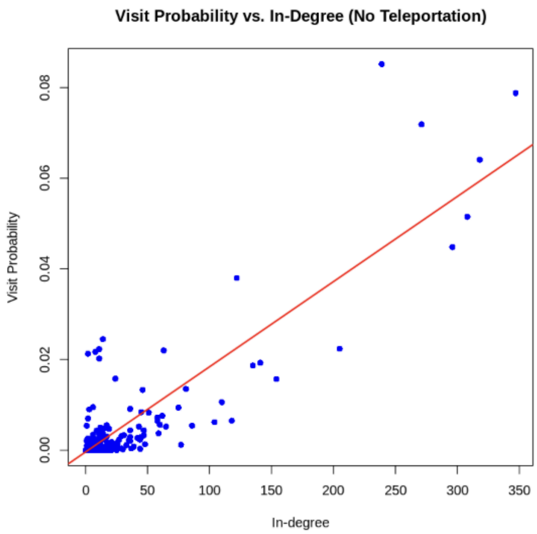


Then, we performed a random walk on this merged graph without teleportation, and recorded the visit probability for each node.



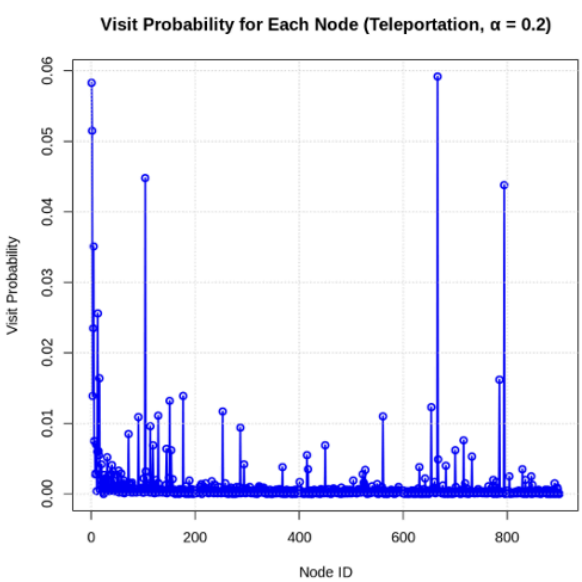
We calculated the correlation between these visit probabilities and the in-degrees of the nodes. The analysis revealed a high positive correlation of approximately 0.879 between in-degree and visit probability, which demonstrates that nodes with a higher in-degree, and therefore more incoming connections, are more frequently visited during the random walk, confirming the significant impact of the network's structural properties on the dynamic behavior observed.

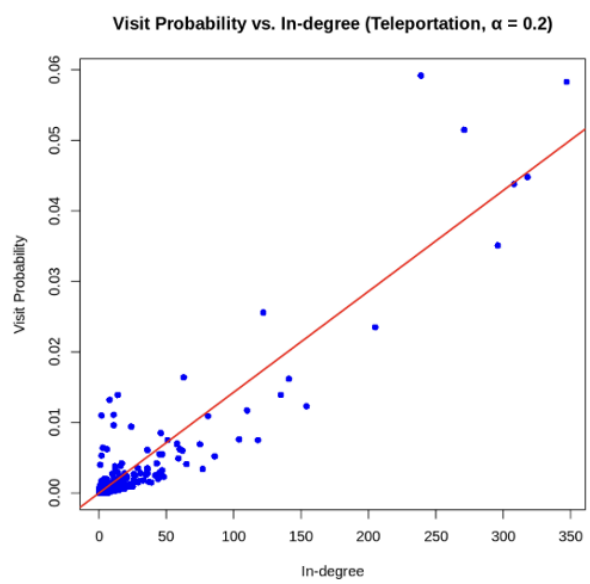
```
Correlation between in-degree and visit probability:  0.8790123
```



Nodes with higher in-degree attract more incoming edges, making them more likely to be reached during a random walk. The strong correlation confirms that the structure of the merged graph—maintaining a power-law-like in-degree distribution—favors high-degree nodes in random walks.
Through merging two preferential attachment graphs and analyzing a random walk on the resulting network, we found that in-degree strongly influences how frequently nodes are visited. The high correlation coefficient supports the idea that nodes with more incoming edges dominate the random walk in a network without teleportation.

(b) In this experiment, we took the merged directed preferential attachment network from Q3(a) (900 nodes, out-degree m=4), which had been constructed to avoid "black holes." We introduced a teleportation probability of α=0.2. In each step, with probability α the walker jumps randomly to any node, otherwise it follows one of the outgoing edges of the current node. We recorded the final visit probability of each node over multiple random walks (e.g., 100 random walks, each sufficiently long to approximate a steady state). The result can be seen as following.





The left plot shows the visit probability for each of the 900 nodes, arranged by node ID. A handful of nodes exhibit significantly higher visit probabilities (above 0.04–0.05), suggesting they are more influential or well-connected. However, teleportation ensures that even lower in-degree nodes receive some nonzero chance of being visited, preventing the distribution from becoming too extreme.
The right plot compares each node's final visit probability with its in-degree. We observe a strong positive correlation (around 0.924) between in-degree and visit probability(seen below).

```
Correlation (with teleportation): 0.9241808
```

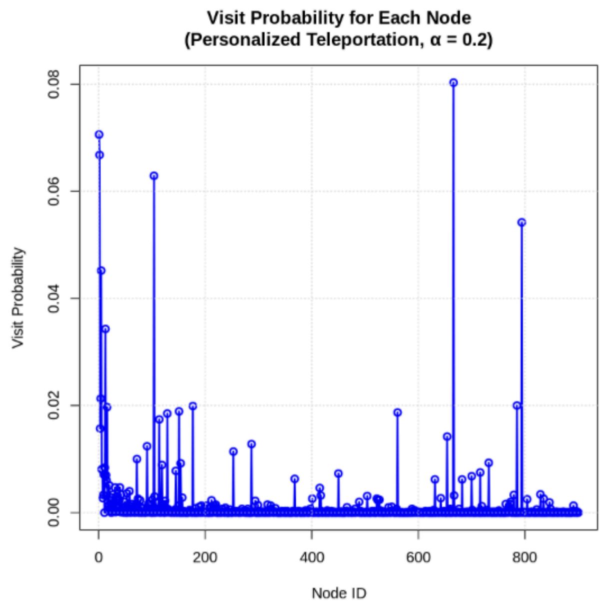
This indicates that although teleportation somewhat "smooths" the visitation probabilities, nodes with higher in-degree remain significantly more likely to be visited—highlighting the structural bias of directed PA networks.
By allowing a 20% chance at each step to jump to a random node, the random walk avoids being trapped in strongly connected components or dead-ends. Teleportation dampens—but does not eliminate—the underlying preference for high in-degree nodes. The correlation remains high (> 0.9), meaning that the network's structural properties (i.e., node in-degree) still dominate.
In the standard PageRank equation, the steady-state probability PR(i) for a node i is given by:

$$PR(i) = \alpha \cdot v(i) + (1-\alpha)\sum_{j \in B_i} \frac{PR(j)}{d_j^{\text{out}}}$$

The term $(1-\alpha)\sum_{j \in B_i} \frac{PR(j)}{d_j^{\text{out}}}$ is dominated by the contributions from nodes that have high in-degree, as they receive more probability mass from many neighbors.The teleportation term α · v(i) redistributes some probability uniformly over all nodes. For α=0.2, 20% of the time the teleportation is independent of the network structure. This means that the final probability is a convex combination of a uniform distribution and one that is driven by the network's topology. As a consequence, even though the introduction of teleportation smooths the distribution, the overall visit probability remains strongly correlated with the in-degree. If α were higher (e.g., close to 1), the network structure's influence would diminish, leading to a much weaker correlation with the in-degree.

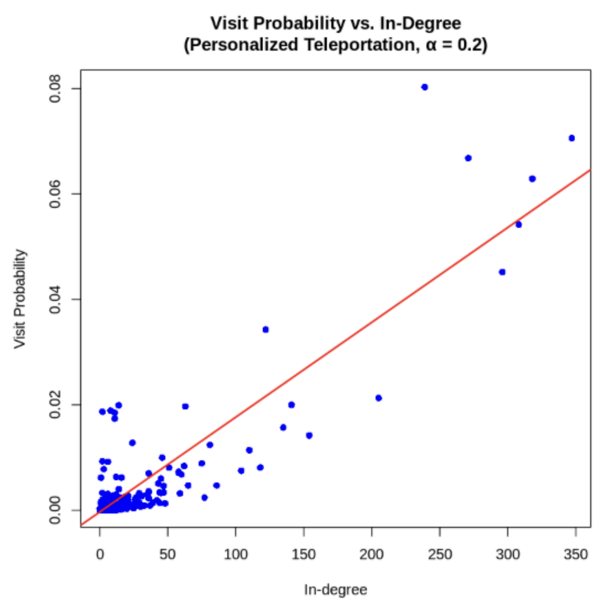
4. Personalized PageRank

(a) In this experiment, we used the merged directed preferential attachment network from Q3, ensuring no "black holes." Instead of teleporting uniformly, the walker jumps to a node with probability proportional to its PageRank (α=0.2). This means that whenever a teleportation event occurs, a node with a higher PageRank has a greater chance of being selected. We ran a random walk of 10,000 steps (with multiple trials if needed), measuring each node's long-term visit probability.



As shown in the above plot, most nodes remain below a 0.01 visit probability, but a few nodes exhibit significantly larger probabilities (> 0.04–0.05). This indicates that nodes already recognized by regular PageRank as “important” are now further reinforced by the personalized teleportation step.

```
Correlation between in-degree and visit probability (personalized): 0.8949519
```



A scatter plot comparing final visit probability to node in-degree shows a correlation of around 0.89.While teleportation is no longer uniform, the structure of the network (i.e., in-degree distribution) still dominates the final probabilities.

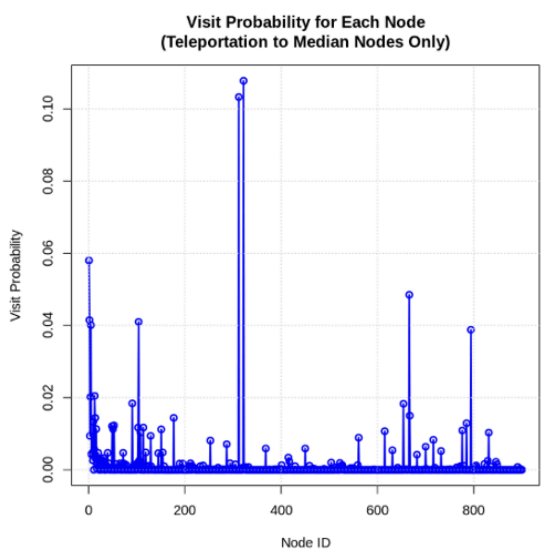
Because each node’s teleportation weight is tied to its PageRank, well-connected nodes that already hold a high PageRank get an additional advantage. This self-reinforcement effect boosts their final visit probability. In Q3(b), teleportation was uniform, and each

node received the same baseline probability of being visited. Now, the influence of node in-degree is even more pronounced, as popular nodes become more popular.
Comparison with Q3(a)

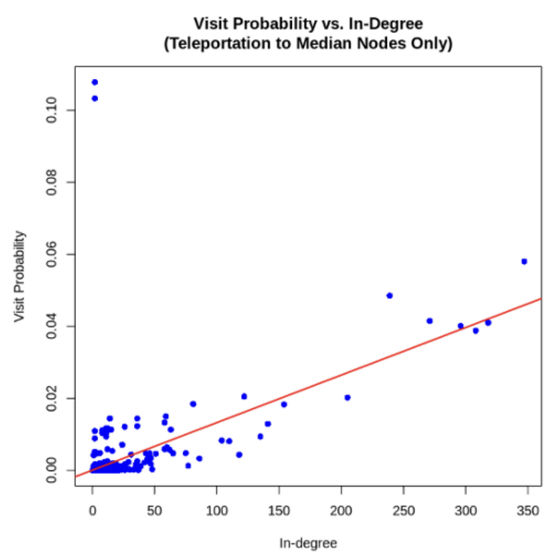
- Uniform vs. Personalized Teleportation
    - Q3(a): Teleportation is uniform. Every node has an equal chance (1/N) during teleportation.
    - Q4(a): Teleportation is proportional to PageRank, so high-PageRank (i.e., high in-degree) nodes get extra weight.
- Impact on Visit Probability:
    - Both methods show a strong positive correlation with in-degree.
    - With personalized teleportation, the inherent bias is amplified—popular nodes receive even higher visit probabilities compared to Q3(a).
- Key Finding:
  Personalized teleportation reinforces the natural tendency of random walks to favor high-degree nodes, leading to a steeper disparity in visit probabilities, as confirmed by the slightly higher correlation and more elevated peaks in the visit probability plot.

(b) In this part, We computed the PageRank for each node in the merged directed PA network (900 nodes, m=4).We identified two nodes with PageRank closest to the median (≈0.00020). We modified the random walk so that whenever teleportation (with α=0.2) occurs, the walker jumps only to these two nodes (with 1/2 probability each), rather than a full PageRank-based or uniform distribution.

```
Nodes with median PageRank values: 312 322
Their PageRank values: 0.0002039648 0.0002041648
```




```
Correlation (Teleportation to Median Nodes): 0.5833727
```



While a few nodes still receive relatively high probabilities, the overall skew is reduced compared to fully personalized teleportation. The scatter plot shows less concentration around very high in-degree nodes.
The correlation with in-degree drops to about 0.58, indicating that restricting teleportation to median PageRank nodes weakens the natural bias toward high-degree hubs. Thus, node in-degree no longer plays as dominant a role in determining visit probability.

(c) In the standard PageRank model, the PageRank vector PRPRPR is computed by the iterative equation

$$PR = d\,M\,PR + (1-d)\,\mathbf{u}, \quad \text{with } \mathbf{u} = \frac{1}{N}\mathbf{1}$$

This formulation assumes that all nodes are equally likely targets during a teleportation event. However, in reality, users do not teleport uniformly at random; they tend to jump to a small set of trusted or "important" pages. This phenomenon is known as self-reinforcement because popular pages get even more attention through preferential jumps. To account for this, we replace the uniform teleportation vector u with a personalized vector v that reflects user interests. One simple approach is to let

$v_i = \frac{PR_i^{\beta}}{\sum_{j=1}^{N} PR_j^{\beta}}, \quad \beta > 1$ Thus, the adjusted PageRank equation becomes

$$PR = d\,M\,PR + (1-d)\,\mathbf{v} \quad \text{or equivalently} \quad PR = d\,M\,PR + (1-d)\,\frac{PR^{\beta}}{|PR^{\beta}|_1}$$

This modified equation means that during teleportation, pages with higher PR receive a larger share of the teleportation probability, which mimics the real-world behavior where users trust and repeatedly visit a limited set of websites.