COMP0123: Complex Networks and Webs
Professor Shi Zhou

COURSEWORK 1 REPORT

November 15, 2023

Student ID: 20019613
University College London
Department of Computer Science

Task 1 - (15 marks)

- Calculate the average node degree and the maximum node degree of the 3 networks.
- Plot their degree distribution $P(k)$ on linear-linear scale and log-log scale, respectively.
- Estimate the power-law exponent of the degree distribution $P(k)$ of the author network only.
  - You can fit a curve by using the function polyfit from the numpy library.
  - Ideally, you can do the fitting on CCDF (the complementary cumulative distribution function) on log-log scale.
- Briefly discuss your results, e.g. difference of the networks.

Table 1

	Author Network	**Random Network**	**BA Network**
Average Degree $\bar{k}$	4.99	4.99	4.99
Maximum Degree k_{max}	51	13	131
Power-Law Exponent α	2.43	-	-

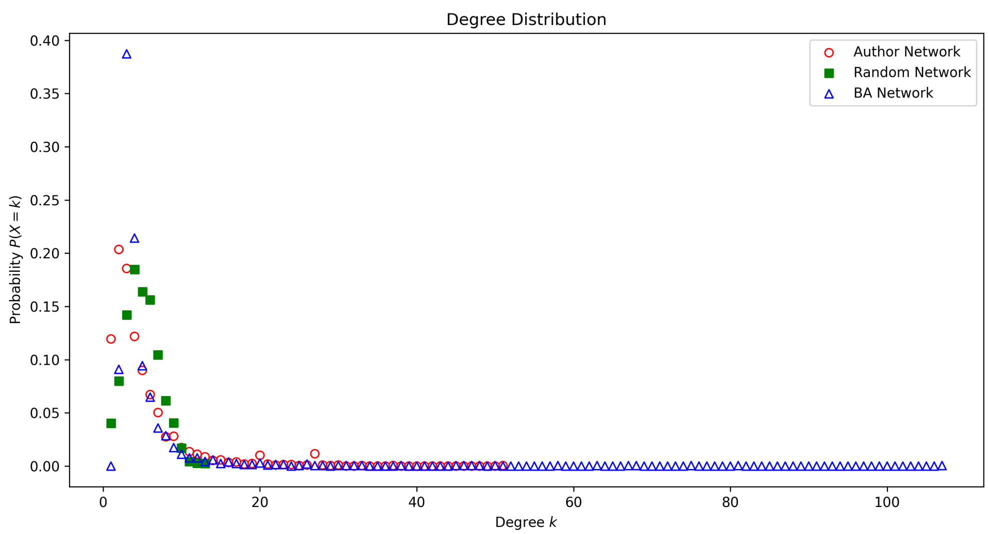


Figure 1: Degree Distribution $P(k)$ on linear-linear scale

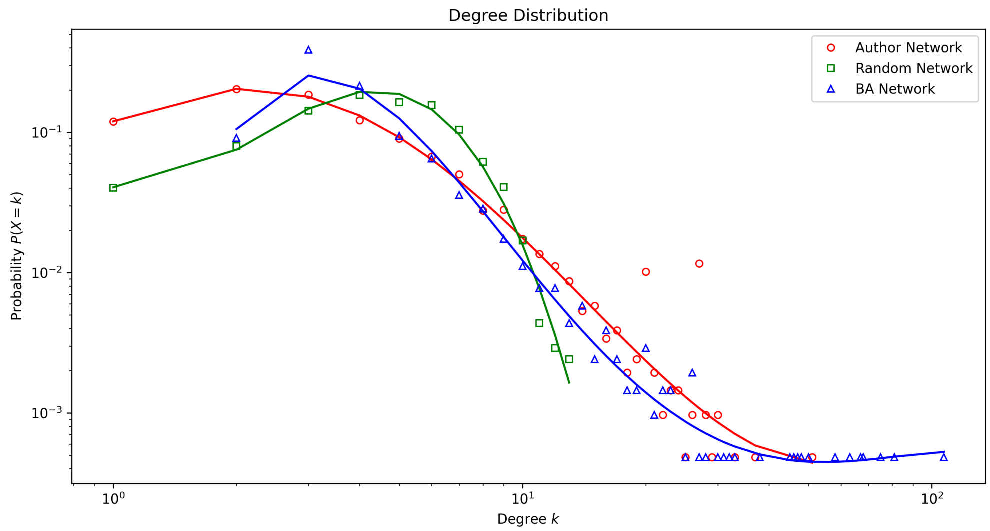


Figure 2: Degree Distribution $P(k)$ on log-log scale

The random network has a comparatively low maximum node degree of $k_{max} = 13$, suggesting a relatively even degree distribution, as evident in Figure 1. Each node in a random network has the same probability of being connected to any other node, indicating that high-degree nodes are extremely rare. We can confirm this as our results (Fig. 2) show a rapid decline in probability as the degree increases.

The BA model is renowned for its preferential attachment mechanism [1], which explains why our network has a maximum node degree of $k_{max} = 131$, the highest among the three networks. This characteristic is evident in Figure 2, where the BA network exhibits a more linear relationship over a broader range of degrees, indicating a scale-free network with a power-law degree distribution.

With a maximum node degree of $k_{max} = 51$, our author network lies between the other networks. Our degree distribution on the log-log scale (Fig 2) shows a distribution that seems to follow a curve rather than a straight line, suggesting it doesn't follow a strict power-law distribution, which is typical for real-world networks.

In both the author network and BA Network, we see that the degree distribution deviates from a strict power law and has short tails, with the BA network being longer.

We estimate the power-law exponent α of the author network to be 2.43 on the CCDF plot with an exponential cutoff, suggesting that the network has some scale-free properties. It also explains why the author network's curve in Figure 2 appears less steep than our BA Network, which usually has a higher power-law exponent.

Task 2 - (15 marks)

- Calculate and plot the nearest neighbour's average degree k_{nn} as a function of degree k, on log-log scale.
- Calculate the assortative coefficient of the networks.
- Briefly discuss your results

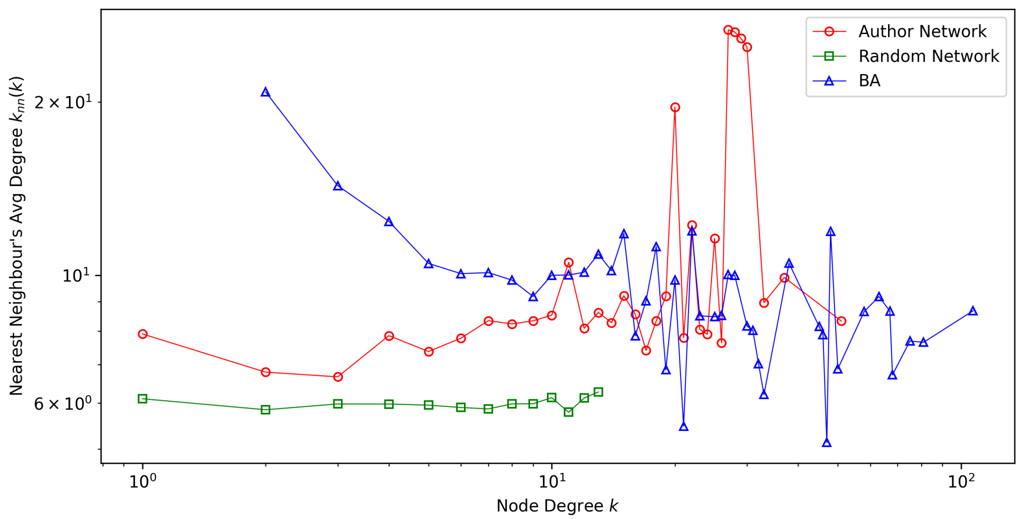


Figure 3: Nearest neighbour's average degree k_{nn} as a function of degree k

Table 2

	Author Network	**Random Network**	**BA Network**
Assortative coefficient r	0.47	0.00	-0.08

The data indicates distinct mixing patterns across the three networks (Fig. 3), where we quantify this result in Table 2 using Newman's proposed assortative coefficient [2]. The author network has a positive assortative coefficient of $r = 0.47$; thus, it exhibits assortative mixing, where nodes with similar degrees tend to be connected. This is expected in social networks as very active authors tend to create a collaborative environment between themselves.

The random network has an assortative coefficient of $r = 0$; thus, it is a non-assortative network. It has no preference for connections, aligning with the expected randomness of such networks. Similarly, the BA network has an assortative coefficient of $r = -0.08$, which is weakly disassortative. Intuitively, a BA network would be disassortative given its preferential attachment mechanism. However, we can see that the preferential attachment interferes with the connectivity of newly added nodes [3]. As the degree grows, the network becomes almost uncorrelated (Fig. 3); Nikolski [4] shows that a BA network does not generate degree correlation for large N.

Task 3 - (15 marks)

- Calculate the diameter and the average shortest path length of the network
- Calculate and plot the average node betweenness of k-degree nodes as a function of node degree k, where node betweenness is normalised, on a log-log scale.
- Briefly discuss your results.

Table 3

	Author Network	**Random Network**	**BA Network**
Diameter d	19	10	7
Average shortest path length ℓ	7.30	4.97	4.18

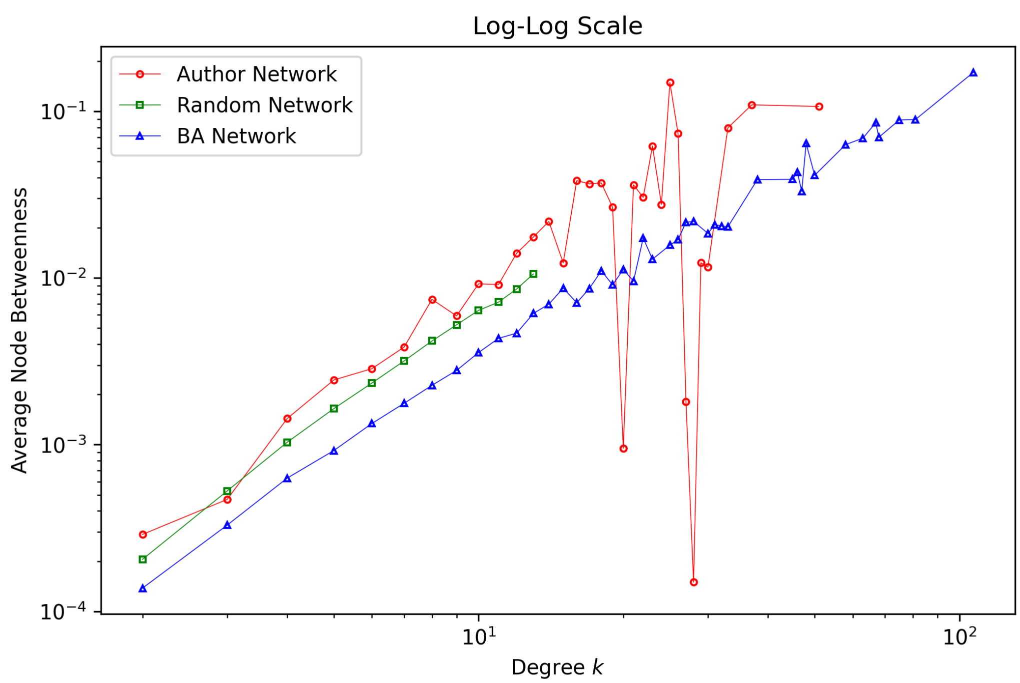


Figure 4: Average node betweenness (normalised) on a log-log scale

Table 3 shows that the BA Network has the smallest diameter of $d = 7$ and an average shortest path length of $\ell = 4.18$, indicating a highly interconnected network. This is typical for scale-free networks due to their hub-and-spoke structure. We can see how this is reflected in Figure 4, which demonstrates high betweenness centrality for the nodes with the most connections (hubs).

Random networks generally exhibit the small-world property, leading to relatively short paths between nodes due to their randomness in connections. Table 2 supports this with the random network having a relatively small diameter of $d = 10$ and an average shortest path length of $\ell = 4.97$. Given the uniform distribution of betweenness across nodes, we see less variation than the other networks (Fig. 4).

The author network has the largest diameter and the longest average path length, suggesting a more fragmented structure. We see a few nodes with high centrality (Fig. 4), indicative of critical roles in connecting different network parts. We can see how this reflects the real-world social systems where communities form based on shared interests or collaborations with relatively low links between these communities.

Task 4 - (15 marks)

- Calculate and plot the rich-club coefficient as a function of node rank on a log-log scale
- Calculate and plot the rich-club coefficient as a function of node degree on log-log scale
- Briefly discuss your result

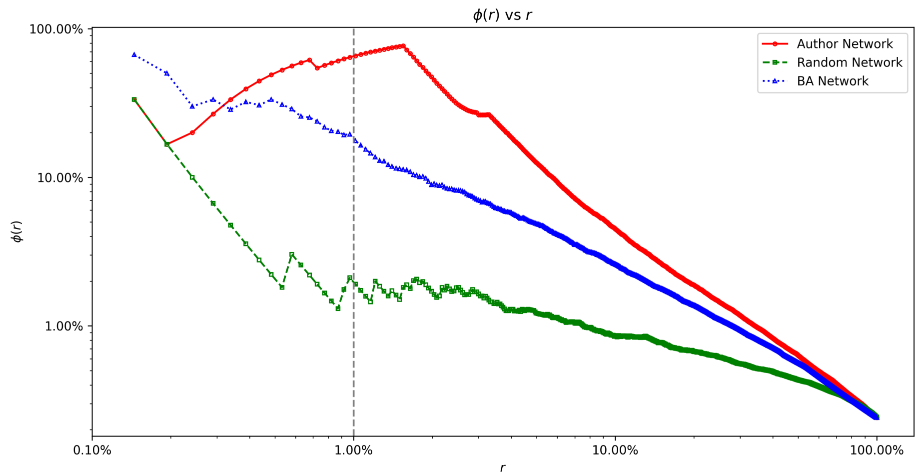


Figure 5: Rich-club connectivity $\phi(r)$ against node rank r

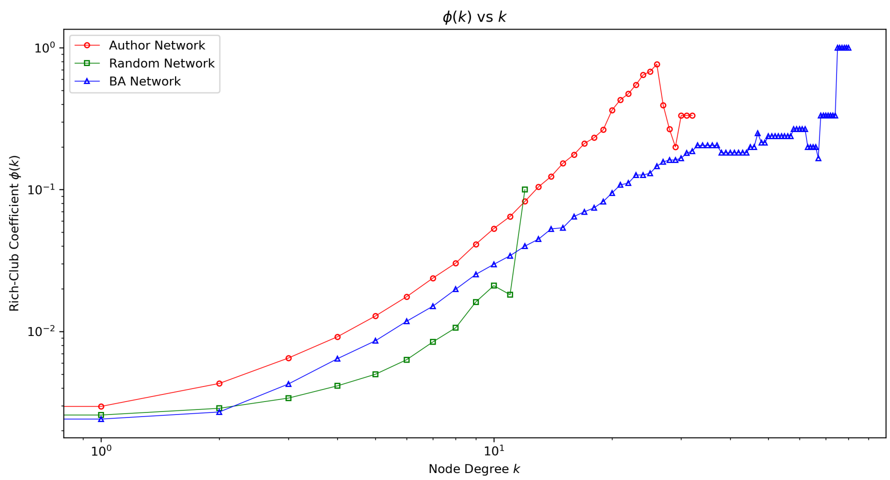


Figure 6: Rich-club connectivity $\phi(k)$ against node degree k

In Figure 5, the author network exhibits a pronounced rich-club phenomenon [5]. The top 1% rich nodes have 65% of the maximum possible number of links, compared with $\phi(1\%) = 18\%$ of the BA network and only $\phi(1\%) = 2\%$ of the random network.

The BA network lacks a prominent rich-club structure because of its growth mechanism; all new links connect with new nodes. Owing to its preferential attachment model, the likelihood of a newly added node ascending to a 'rich' status reduces as the network grows. As a result, rich nodes become less interconnected among themselves.

Likewise, the random network aligns with the expectation that no significant rich-club structuring exists in such networks due to its evenly distributed random connections.

As the degree size increases, the author network's tendency to form exclusive groups of rich nodes increases (Fig. 6). The BA network also shows this trend but to a lesser extent. Meanwhile, the random network does not show this pattern; all its connections are equal.

Task 5 - (15 marks)

- Obtain the community structure (with the largest modularity value) of the three networks
- Give the number of communities and the size (i.e. number of nodes) of the top 3 largest communities in each network.
- Visualise the 3 networks. In each network, show every community with a different colour.
- Briefly discuss your result

Table 4

	Author Network	**Random Network**	**BA Network**
Number of communities $\|C\|$	43	22	22
Modularity value	0.9	0.45	0.45
Largest community size $\|C_1\|$	132	152	191
2nd largest community size $\|C_2\|$	120	144	185
3rd largest community size $\|C_3\|$	119	139	169

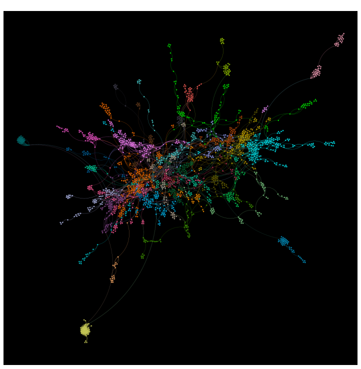


Figure 7: Author Network

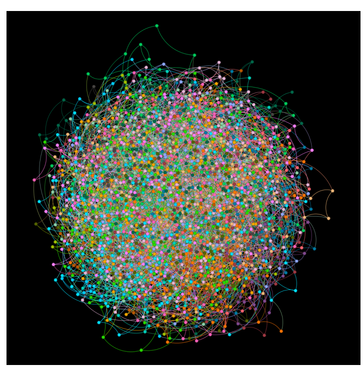


Figure 8: BA Network

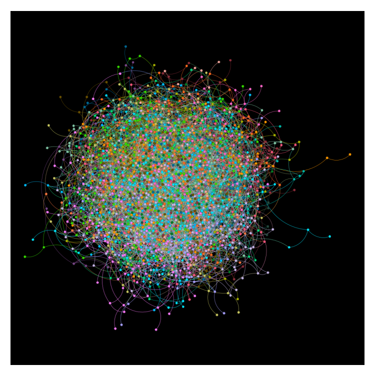


Figure 9: Random Network

The high number of communities and high modularity value show that author collaborations reflect the real-world nature of author collaborations. This is evident in our visual representation (Fig. 7), where separate communities exist.

Figure 8 lacks distinct clusters. The "rich get richer" mechanism of the BA model allows for a few highly connected hubs. These hubs are connected to many nodes across the entire network, which tends to prevent the formation of tightly-knit, isolated communities.

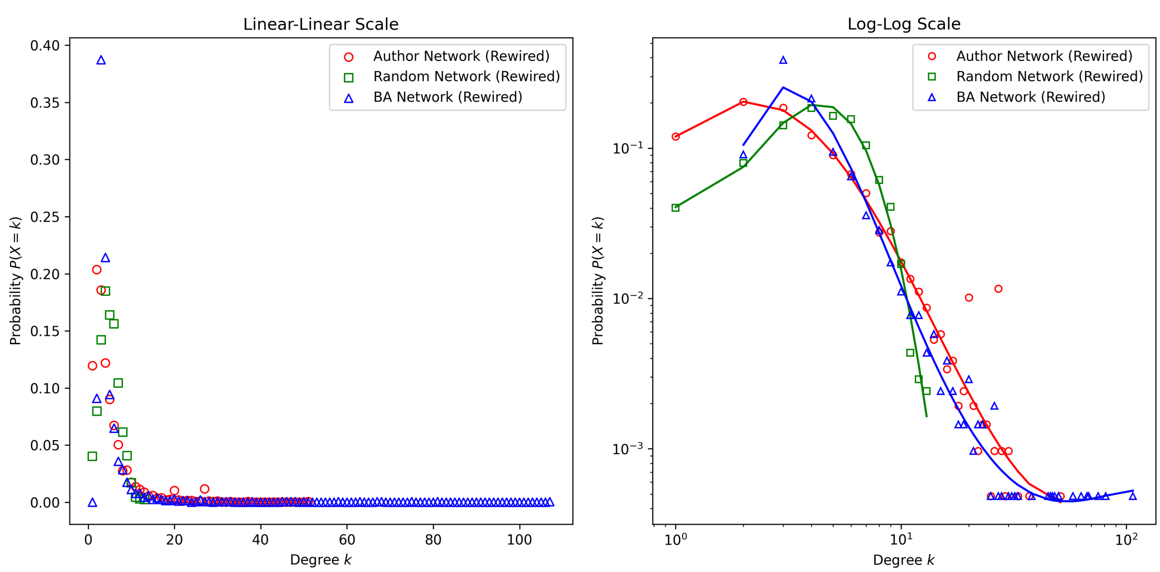
Applying the greedy modularity algorithm [6] gives us 22 communities. With comparatively fewer communities than the author network, we can take note of the more considerable size disparity among them. However, a modularity value of 0.45 does indicate that there are some community-like structures.

Even though the random network has the same number of communities as the BA network, we see less deviation in terms of community size. The equal likelihood of connecting any two nodes does not support the formation of strong clusters as it doesn't promote a higher density of links within certain node groups; random networks typically have a lower clustering coefficient.

Task 6 - (25 marks)

- Randomly rewire the 3 networks while preserving the degree distribution; and obtain the maximal random case of each network
- For the 3 randomised networks, plot their degree distribution
- For each of the randomised networks, calculate the average clustering coefficient, the assortative coefficient, the size of the giant component, and the average shortest path length in the giant component. Show these results and compare them with those of the 3 original networks in a table,
- Briefly discuss your result.

Figure 10: Degree distribution $P(k)$ of our rewired networks



Obtaining the maximal random case of each network should preserve the degree distribution [7]. Figure 10 supports this belief as we have identical degree distributions as our original networks (Fig. 1 & 2).

Table 5

Original	**Author Network**	**Random Network**	**BA Network**
Giant Component Size N	2068	2068	2068
Average Clustering Coefficient C	0.62	0.00	0.02
Assortative coefficient r	0.47	0.01	-0.07
Average shortest path length ℓ	7.30	4.94	4.02
Rewired			
Giant Component Size N	2064	2068	2068
Average Clustering Coefficient C	0.01	0.00	0.02
Assortative coefficient r	0.03	-0.01	-0.02
Average shortest path length ℓ	4.43	4.94	4.1

The maximal random rewiring of the author network decorrelates the network. The significant decrease in our average clustering coefficient is due to breaking up locally clustered nodes and spreading connections more uniformly, which diminishes our community structure. Random rewiring causes the total degree correlation to decrease [8]. We see a large drop in assortativity, which, as a result, typically reduces our average shortest path length [9].

This is in contrast to the random network and the BA network, where the maximal random case shows minimal difference; **(i)** average clustering coefficient ($C\pm$ 0.00), **(ii)** assortative coefficient ($r \pm 0.05$), and **(iii)** average shortest path length ($\ell \pm 0.08$).

Tools, Packages and Libraries

- Python
  - import random
  - import numpy
  - import matplotlib.pyplot
  - import matplotlib.ticker
  - import networkx
  - import sympy
  - import scipy.stats
- Gephi

References

[1] R. Barabasi, A. & Albert, "Emergence of scaling in random networks," *Science*, vol. 286, pp. 509–512, oct 1999.

[2] M. E. J. Newman, "Assortative mixing in networks," *Physical Review Letters*, vol. 89, oct 2002.

[3] G. G. Piva, F. L. Ribeiro, and A. S. Mata, "Networks with growth and preferential attachment: modelling and applications," *Journal of Complex Networks*, vol. 9, p. cnab008, 04 2021.

[4] C. Stegehuis, "Degree correlations in scale-free random graph models," *Journal of Applied Probability*, vol. 56, no. 3, pp. 672–700, 2019.

[5] S. Zhou and R. Mondragon, "The rich-club phenomenon in the internet topology," *IEEE Communications Letters*, vol. 8, no. 3, pp. 180–182, 2004.

[6] A. Clauset, M. E. J. Newman, and C. Moore, "Finding community structure in very large networks.," *Physical review. E, Statistical, nonlinear, and soft matter physics*, vol. 70 6 Pt 2, p. 066111, 2004.

[7] M. E. J. Newman, S. H. Strogatz, and D. J. Watts, "Random graphs with arbitrary degree distributions and their applications," *Phys. Rev. E*, vol. 64, p. 026118, Jul 2001.

[8] J. Qu, S.-J. Wang, M. Jusup, and Z. Wang, "Effects of random rewiring on the degree correlation of scale-free networks," *Scientific Reports*, vol. 5, p. 15450, 10 2015.

[9] R. Xulvi-Brunet and I. M. Sokolov, "Reshuffling scale-free networks: From random to assortative," *Phys. Rev. E*, vol. 70, p. 066102, Dec 2004.