University College London

COMP0123 Complex Networks and Web (2023/24)

Coursework 1 Report

Task 1

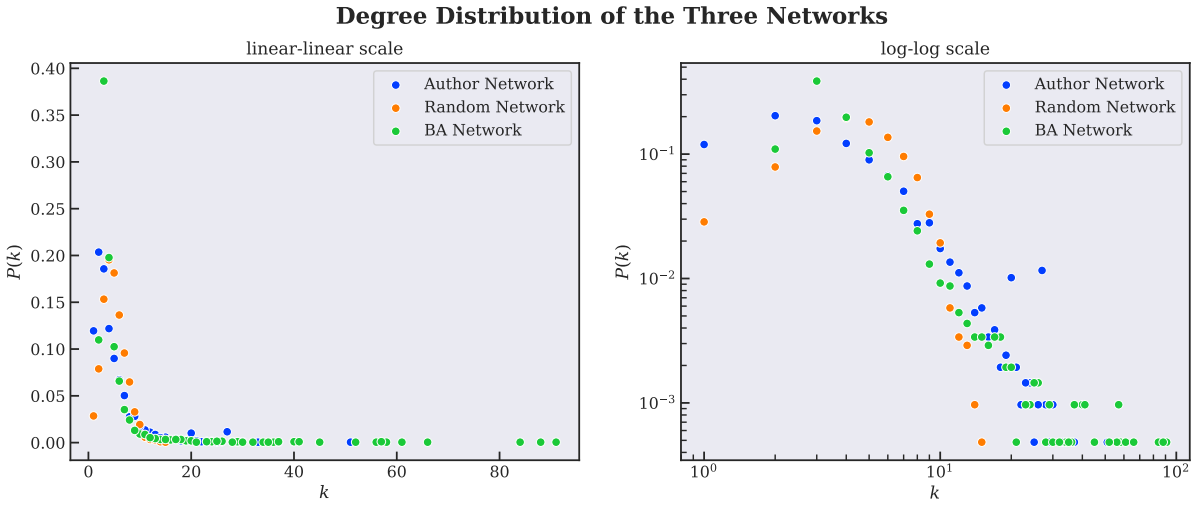
- Calculate the average node degree and the maximum node degree of the 3 networks.
- Plot their degree distribution $P(k)$ on linear-linear scale and log-log scale, respectively.
- Estimate the power-law exponent of the degree distribution $P(k)$ of the author network only.
  - You can fit a curve by 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.

Answer

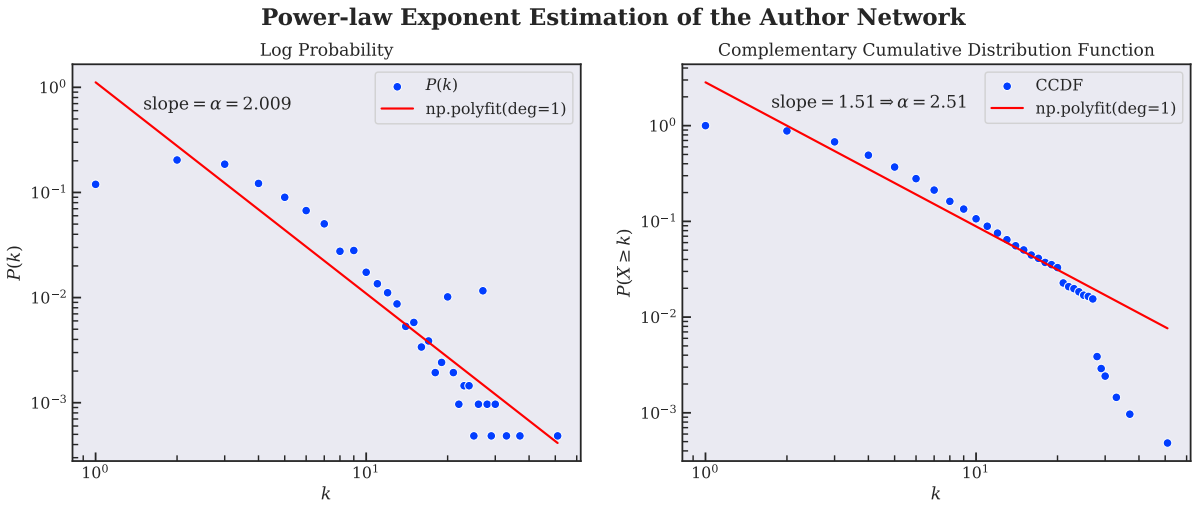
- Average node degree and the maximum node degree of the 3 networks:

Network	Average Node Degree	Maximum Node Degree
Author Network	5.00	51
Random Network	5.00	15
BA Network	5.00	101

- Plots of degree distribution $P(k)$



- Power-law exponent estimation of the Author Network



The resulted exponents of two different fittings are shown as follows:

Fitted on the log probability	$\alpha = 2.01$
Fitted on the CCDF	$\alpha = 2.51$

- Discussion
  - Degree Distribution

    The BA network is designed to be a scale-free (power-law) networks which has a heavy tail, meaning that the majority of nodes have only a few edges but there are a small amount of nodes having a large number of edges. The Author network show similar feature like a few hubs having large number of links as the BA network. Hence, confirms that social network exist as a power-law network. However, the random network does not mimic a power-law network. It does not have a long, heavy tail and most nodes possess similar degree with the average node degree. Its degree distribution bears a strong resemblance to a binomial distribution.

  - Exponential Estimation of the Author Network

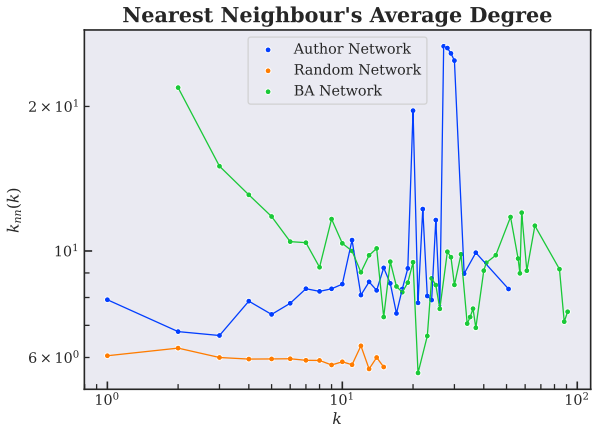
    When fitting on the log probability distribution, the fitted curve tends to fit closer to high degree nodes since the least square fitting treats every data point with the same weight. Hence, it does not reflect the power law exponent accurately as those high degree nodes indeed represent a minority of nodes in the network. Some noise of the distribution also adversely affect the exponent estimation. Another fitting on CCDF is done to provide a more accurately estimation. The CCDF curve is smoother and less noisy. The least square fitting is done without the last 6 data points. These data point are dropped since they reflect the exponential part of the CCDF curve.

Task 2

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

Answer

- Nearest neighbour's average degree



- Assortative coefficients

Network	Average Node Degree
Author Network	0.47
Random Network	-0.02
BA Network	-0.08

- Discussion

  The Author network shows assortative mixing as shown by the plot and also the assortative coefficient. Its nodes tend to connect with nodes of similar degree, which is a usual phenomenon in scientific collaboration network. The random network is neutral as expected since new edges of the network are attached to nodes randomly during the network creation. On the other hand, the BA network is expected to be disassortative since new nodes are preferentially attached to existing nodes with high degree. However, the assortative coefficient of the BA network shows that the network is only slighly disassortative. Also, the KNN plot of the BA network only shows a disassortative mixing pattern for the first half of the plot, while the remaining part shows a neutral mixing pattern.

Task 3

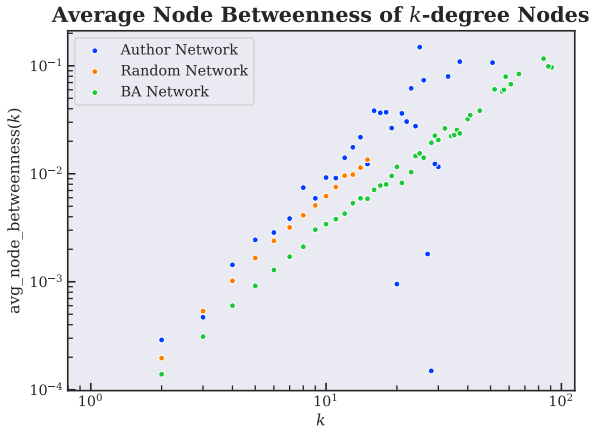
- Calculate the diameter and the average shortest path length of the network.
- Calculate and plot the average node betweenness of k-degree nodes as a function of node degree k, where node betweenness is normalised, on log-log scale.
- Briefly discuss your results.

Answer

- Diameter and the average shortest path

Network	Diameter	Average Shortest Path
Author Network	19	7.30
Random Network	10	4.94
BA Network	7	4.10

- Average node betweenness of k-degree nodes



- Discussion

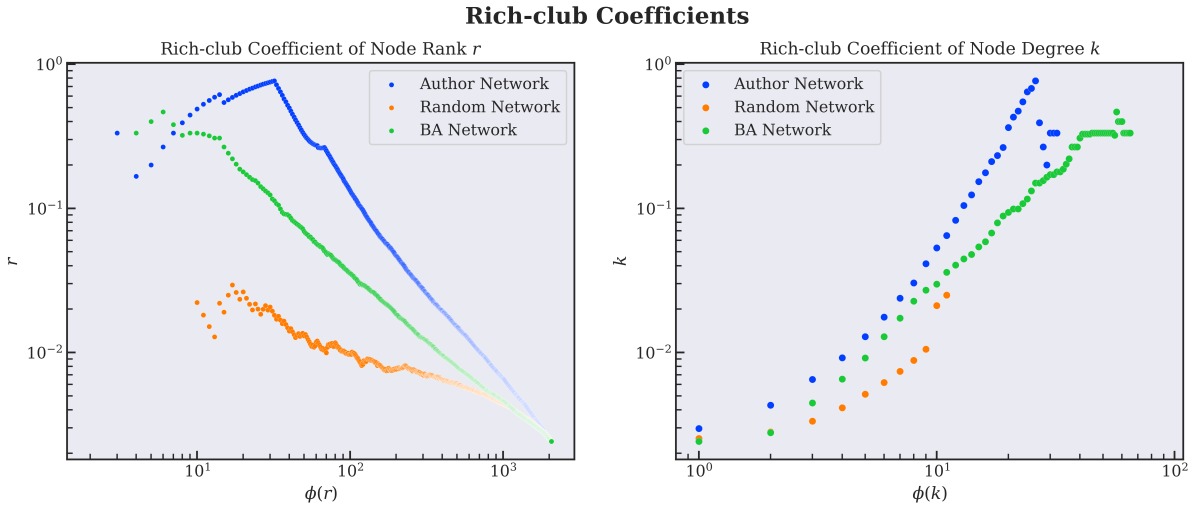
  The random network and the BA network have similar diameter and average shortest path. Both networks are more connected than the Author network, which has the highest diameter and average shortest path. These statistics are consistent with the plot of average node betweenness of k-degree nodes. Overall, nodes with higher degree have higher betweenness since they have higher chance to be contained in shortest paths between other nodes. For the Author network, there are some high degree nodes having significantly low average betweenness. This explains that the network having the highest diameter and average shortest path.

Task 4

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

Answer

- Rich-club coefficient



- Discussion

  Rich clubs are observed in the Author network and BA network while that in the Author netowrk is more pronounced. However, it is interesting to note that the nodes group with lowest rank (or highest degree) does not have the highest rich club coefficient, which quantify the density of interconnectivity between nodes group. During BA network creation, preferential attachment guarantee rich nodes gain more edges while not essentially increasing the interconnectivity between rich nodes. In contrary, the random network does not exhibit a rich club phenomenon since edges are added randomly, disregarding the nodes degree.

Task 5

- Obtain the community structure (with the largest modularity value) of the 3 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.

Answer

Community structure are obtained using Louvain community detection algorithm with largest modularity value.

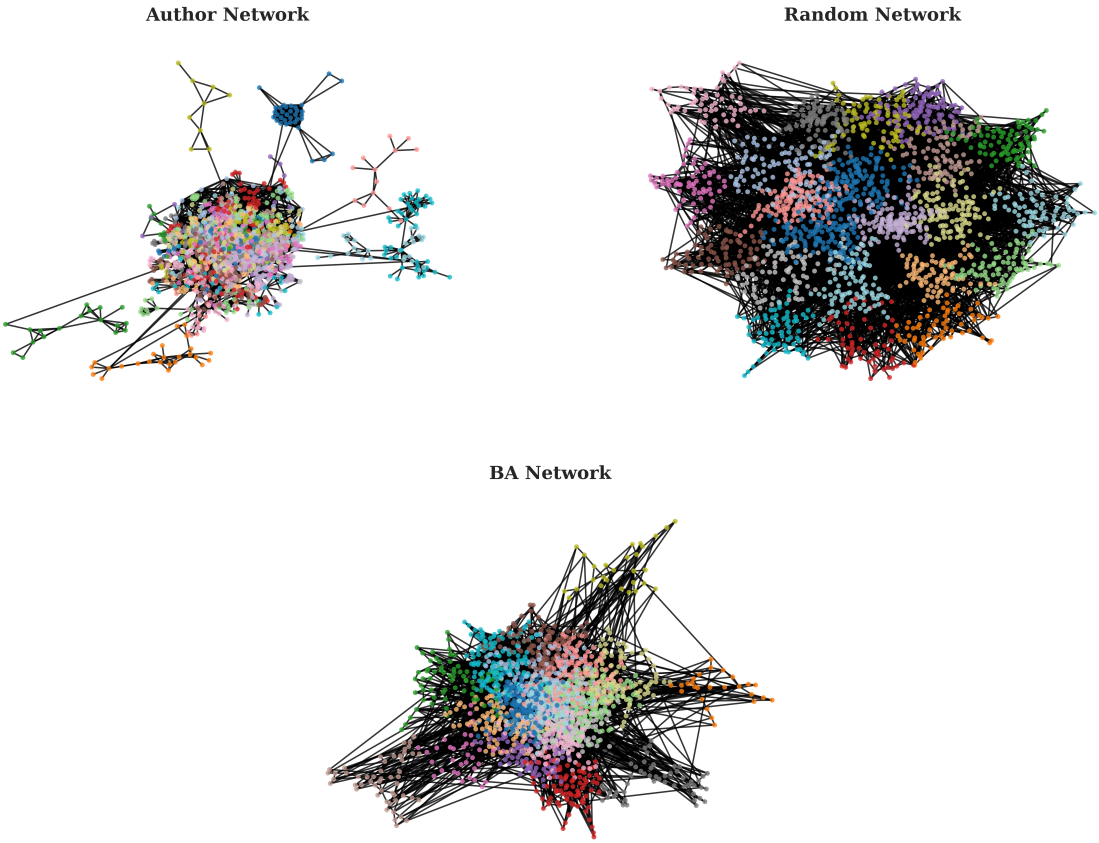
- Modularity

Network	Modularity value
Author Network	0.90
Random Network	0.45
BA Network	0.45

- Number of communities and the size of the top 3 largest communities

Network	Number of Community	Size of Top 3 Largest Communities
Author Network	43	135, 124, 119
Random Network	23	141, 137, 136
BA Network	20	194, 193, 185

- Visualising the communities



- Discussion

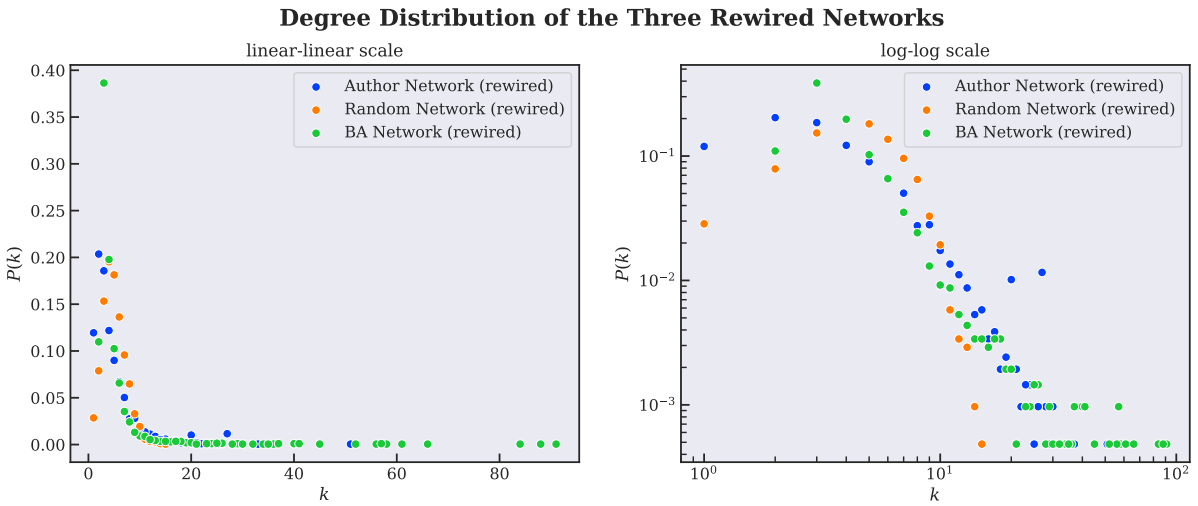
  Comparing with the random network and BA network, the Author network has more explicit and pronounced communities as observed from the plot and its much higher modularity. This can be understood since the two other networks have connections established randomly, while in real world collaboration, authors often work with people of the same field or organisation. Although communities are detected in the random network and BA network, they are not as separable as those in the Author network. Communities in the random network and BA network are still highly connected. Another feature observed is that the BA network's communities are much larger then others, which is a consequence of preferential attachment where highly connected nodes are created and form large communities.

Task 6

- 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 with those of the 3 original networks in a table.
- Briefly discuss your result.

Answer

- Degree distribution of the rewired network



- Comparing statistics of original networks and rewired networks

Network	Avg. Clustering Coefficient	Assortative Coefficient	Size of Giant Component	Avg. Shortest Path in Giant Component
Author Network	0.62	0.47	2068	7.30
Author Network (rewired)	0.01	-0.01	2068	4.41
Random Network	0.003	-0.02	2068	4.94
Random Network (rewired)	0.001	-0.01	2068	4.94
BA Network	0.01	-0.09	2068	4.08
BA Network (rewired)	0.01	-0.02	2068	4.16

- Discussion

  After maximal random link rewiring, all three networks become neutral in node degree correlation while the networks remain connected. The community structure and node correlation in the Author network are completely removed after the rewiring since the rewiring process is done randomly. The average shortest path in the network is also significantly reduced, meaning that the Author network becomes more connected. For the other two networks, changes are not significantly since the edges of the network are also generated randomly during the network creation.

Tools

All data processing, calculation and visulisation in this coursework are done using Python programming. All Python libraries used and their corresponding usages are listed as follows:

- matplotlib - data visualisation
- seaborn - data visualisation
- numpy - numerical calculation
- networkx - network manupilation