

Network Science

Rubric Assignment 2

- All results must be submitted through a well-documented Jupyter Notebook.
 - Submit a zip folder containing the data, Jupyter Notebooks, and the PDF of Jupyter Notebook for each question separately. Use the **filename convention: GroupNo_Assignment2**
-

1. Implement Watts and Strogatz's small-world network model. **(20 marks)**
 - Submit a jupyter notebook named as Q1 and write the code for the algorithm implementation. **(10 marks)**
 - Compute and plot the 'scaled clustering coefficient' and 'scaled characteristic path length' of Watts and Strogatz's network models with increasing value of rewiring probability. Create a plot with proper axis labeling. **(10 marks)**
2. Write the code to implement the Barabasi-Albert (BA) algorithm for generation of scale free networks. **(10 marks)**
 - Submit a jupyter notebook named as Q2.
 - Code to implement BA algorithm from the scratch without using any inbuilt library. **(5 marks)**
 - Create a final network (with a minimum 100 instances).
 - Find and Plot the (a) average clustering coefficient, (b) characteristic path length, and (c) degree distribution. **(2+2+1 marks)**
3. Modify the Barabasi-Albert algorithm to accentuate/strengthen the bias of rich getting richer phenomenon such that the probability of a newly added node getting connected to an existing node is now "proportional to the square of its degree." Compute and compare topological features a comparable size of networks created using BA algorithm. Create variants of higher order. **(10 marks)**
 - Submit a jupyter notebook named as Q3.
 - Code to implement modified BA algorithm as discussed in question. **(5 marks)**
 - Compute and compare topological features of networks created using BA algorithm. **(2 marks)**
 - Create the network of higher order. **(3 marks)**
4. (a) Implement the 'random' node deletion strategy in a network. Measure the response in terms of 'characteristic path length' of the network when increasing fraction of nodes (f) are removed. Also measure the size of the giant cluster (S). You may use existing graph analysis libraries (such as igraph/NetworkX). **(5 marks)**
 - Submit a jupyter notebook named as Q4 and write code of a-d.
 - Implement random node deletion strategy in a network as per the question. **(4 marks)**

- Calculate the size of giant cluster. **(1 mark)**
- (b) Extend the study to compare the response of random and scale free networks to node deletions according to (i) random, and (ii) targeted strategies. **(5 marks)**
- Compare the response of node deletions according to random and targeted strategies. **(2.5 +2.5 marks)**
- (c) Perform the above-mentioned analysis on any real-world (scale-free) network of your choice ($500 < n < 5000$). **(5 marks)**
- Perform the analysis as mentioned. **(5 marks)**
- (d) Comment on your observations, knowing that the observed pattern of response in (i) characteristic path length and (ii) size of giant cluster/average size of fragmented component is expected to be as follows, as reported by Albert et al., Nature, 406, 378 (2000). **(5 marks)**
- Comment and compare your observations as reported by Albert et al. **(5 marks)**
5. (a) Construct the Yeast Protein Interactome (YPI) using data available from Yeast Interactome Database (http://interactome.dfci.harvard.edu/S_cerevisiae/). Plot its degree distribution as well as $C(k) \times k$ plot. Visualize the network using Cytoscape and highlight importance of a protein in terms of its degree/connectivity. **(5 marks)**
- Submit a jupyter notebook named as Q5 and write the code of parts a-c.
 - Construct YPI network. **(1 mark)**
 - Plot degree distribution with proper axes labeling. **(1 mark)**
 - Plot $C(k) \times k$ with proper axes labeling. **(1 mark)**
 - Visualize the network using Cytoscape. **(1 mark)**
 - Highlight the importance of protein in terms of its degree. **(1 mark)**
- (b) Identify key proteins in YPI based on their number of interactions. Compute and plot the fraction of proteins that are essential for a given degree (k). (Find the information of essential genes from existing data sources from the internet.) **(5 marks)**
- Identify key proteins in YPI. **(3 marks)**
 - Plot fraction of proteins that are essential for a given degree (k) with proper axes labeling. **(2 marks)**
- (c) Implement the random and targeted node deletion strategies (Refer to Question 3 above) for YPI. Compute and plot the robustness (S and $< s >$) of YPI for these strategies. **(5 marks)**
- Implement random and targeted node deletion strategies for YPI. **(2.5 marks)**
 - Plot robustness (S and $< s >$) of YPI for random and targeted node deletion with proper axes labeling. **(2.5 marks)**