Network Biology

Coding Assignment 1

Note: You are responsible for the backup of data as well as results, which will be used for evaluation.

Code for the following tasks. Do not use existing libraries (such as NetworkX/igraph) for network analysis. You need to submit the Jupyter Notebook, its PDF as well we the data files for evaluation.

1. Draw a simple network (of ~10 nodes and ~20 edges) of each of the following classes: undirected & unweighted (UD-UW); undirected and weighted (UD-W); directed and unweighted (D-UW); directed and weighted (D-W). Store these networks using (i) adjacency matrix and (ii) edge list representation. **[2]**

2. Write a program to compute the degree (for undirected graphs) and in/out-degree (for directed graphs) of each node. **[3]**

3. Starting from the adjacency matrix (A) of the UD-UW graph, compute A^2. What information, if any, does it hold in the context of the graph you have drawn in response to exercise 1? What about higher-order matrix multiplications of A? **[3]**

4. Download data of networks available from Mark Newman's page (http://www-personal.umich.edu/~mejn/netdata/). Read the data, and measure the degree of nodes (using the program written in exercise 2). Compute and plot the 'degree distribution'. **[2]**

5. Write a program to create a random graph/network (UD-UW) of n nodes and e edges, and to compute its 'degree distribution'. **[5]**

6. For real-world networks used in exercise 4, create 100 instances of random controls (by preserving the number of nodes and edges). Compute and plot its average degree distribution. Comment on differences in degree distributions observed in real-world networks and their random counterparts. **[5]**

7. Read the data of real-world networks into Cytoscape. Use Cytoscape to compute their degree distributions. Visualize the networks and export their images in JPG. **[2]**

8. Write a program to compute the **clustering coefficient** of node (C_i) and average clustering coefficient of the network (C) starting from the adjacency matrix of an undirected graph. **[5]**

9. Write a program to compute the **shortest path length** between a pair of nodes (L_{ij}) and the characteristic path length of the network (L) starting from the adjacency matrix. (Hint: Look at well-known algorithms such as that by Dijkstra.) **[5]**

10. Download the HPRD data. Compute the number of nodes and edges. Plot the degree distribution of the protein-protein interaction network on a log-log scale. **[3]**