Frontiers of Network Science 6250/4250, Fall 2024
Assignment 1, due at 11:59 am on Tuesday, October 22nd

For each of the two real networks you selected, create two artificial networks with the same number of nodes and approximately the same number of edges. The first artificial network should be an ER random graph, while the second should be a BA scale-free network. For ER graphs, compute the average degree of the nodes in your network to get a similar number of edges in the generated random graph. For the BA network, choose the minimum degree to get the number of edges in the scale-free network closest to the actual network. As a result, you will work with two sets of three networks: real, its ER version, and its BA version, with all three networks having the same number of nodes and a similar number of edges but at least two different degree distributions.

For each of the six networks:

1. Provide a detailed description of the two networks that you chose. Do not forget to specify the number of nodes and edges, what they represent, how, when, and by whom the network was collected, its significance and meaning, whether the edges are directed and weighted, etc.

2. Compute the global properties of a network: (i) the diameter, radius, and average eccentricity; (ii) the number of connected components (for directed graphs, weakly and strongly connected) and the size of the largest one; (iii) the node average degree and degree variance.

3. Compute and plot: (iv) the degree distribution (for directed graphs, also in-degree and out-degree distributions), (v) the length distribution of each node's longest shortest paths, (vi) the clustering coefficient distribution, (vii) the connected components size distribution.

   Compute and list (viii) the average values and (ix) variances for the measurements (v)-(vii).

   Ensure that you use a log scale on the vertical axis or log on both axes (log-log plot) when needed. (*) For each distribution, also plot the best-fit line on the same plot.

Hint: for 2(i) and 3(v), repeatedly use breadth-first walk from the node to find its longest shortest path (one path per node)

Optional: Render a high-quality (vector) graphic representation of each of your two chosen networks and include it in your report. If you find that the network is too extensive to have all nodes and edges drawn directly without affecting the presentation quality of the figure, proceed as follows. Use filtering, collapsing expansion, hierarchical representation, or other techniques to reduce visual clutter) Make the visualization of your network visually appealing (i.e., legible, with proper layout and labels when necessary) and meaningful (provide visual cues that should help readers understand your analysis of the network, e.g., different colors and sizes of nodes reflect different values of metrics which you discuss in your analysis). Provide the legend explaining the meaning of different colors, sizes, types of lines, etc. Justify your choices.

You can use whatever tools (either your own or third party) you deem appropriate for the job. If you use a third-party tool (e.g., Gephi, Neo4j), it must be free of charge (or at least a fully functional evaluation version should be available). Please document which tools you use for which task, including the URL of the tool Web site.

If you are using your tools (e.g., writing your programs), please specify which ones and for which tasks and provide the source code and executable for ThinkPad running Microsoft Windows 10, along with the relevant instructions on how to run them.

If you are using third-party frameworks or applications (e.g., Matlab, Microsoft, Excel), indicate which tool you used and for which task and provide all user files (".m," ".xls," etc.).

Briefly justify choosing a particular tool for a particular task in your report.

Grading
The maximum numbers of points assigned to Tasks are as follows:
Task 1: 10 pts, **Task 2:** 15 pts, Task 3: 15 pts. **Optional task:** 10 points
Partial answers of partially correct answers will earn partial credit.

Each student should do the assignment individually; students should not collaborate on it.
Submit your solution to the email TA with the necessary files, answers, and plots as attachments.