DSC190/291: Network Science and Graph Theory **Winter 2025**

Homework Assignment 4

Topics: Power-law distributions and Barabási-Albert Networks

Instructor: **Prof. Gal Mishne**
Deadline: **Friday Feb. 14 (2/14/2025)**
Submission: **Gradescope**
Write full answers to the questions below. Show your calculations.

Note: GenAI tools are permitted solely for problems: 2, 3, 4.

1. *Power law*

   Consider an on-line news site, such as cnn.com or nytimes.com, which consists of a front page with links to many different articles. The people who operate such sites generally track the popularity of the various articles that get posted, asking questions like "As a function of k, what fraction of all articles have been viewed by k people?" Let's call this the popularity distribution of the articles. Now suppose that the operators of such a news site are considering changing the front page, so that next to each link is a counter showing how many people have clicked on the link (e.g., next to each link it will say something like, "30,480 people have viewed this story," with the number getting updated over time.)

   (a) What effect do you think this change will have on the behavior of people using the site?

   (b) Do you expect that adding this feature will cause the popularity distribution of the articles to follow a power-law distribution more closely or less closely, compared to the version of the site before these counters were added? Give an explanation for your answer.

2. *Graph visualization - Cora*

   The *Cora* dataset consists of 2708 scientific publications classified into one of seven classes. The citation network consists of 5429 links. The class labels of each paper is given in the last column of the file *cora.content*. Download the dataset and read the README file.

   - Construct the network using the citation information given in *cora.cites*.
   - Visualize the network (for example using *networkx*) and color the nodes by the node labels (paper classes). Use two different graph layouts graph layouts to visualize the graph such that one layout is informative (you can mostly tell the papers apart by topic) and another which is uninformative.

   (Note the purpose of this exercise is for you to consider how different graph layouts affect what you can understand about a network's structure from its plot. Useful visualizations will be important in your project.)

3. *Generating Barabási-Albert Networks*

   In Python implement the Barabási-Albert model model and generate a network with $N = 10^4$ nodes using $m = 4$. As initial condition use a fully connected network with 4 nodes.

(a) Calculate the degree distribution at intermediate steps, namely when the network has 10^2, 10^3 and 10^4 nodes.

(b) Compare the distributions at these intermediate steps by plotting them together and fitting each to a power-law with degree exponent γ. Do the distributions "converge"? Explain.

(c) Plot together the cumulative degree distributions at intermediate steps.

(d) Measure the average clustering coefficient as a function of N.

(e) Measure the degree dynamics of one of the initial nodes and of the nodes added to the network at time $t = 100, t = 1,000$ and $t = 5,000$ (i.e. plot the degree of these nodes over time). Explain the difference you observe (if any).

4. *Modifying the Barabási-Albert Model*

Based on your BA model implementation, implement in Python the following models with $N = 10^4$ nodes and $m = 4$. As initial condition use a fully connected network with 4 nodes.

(a) Modify the BA model as follows: every 10th time step instead of adding a node, delete an existing node from the network at random. (Note you need to run the simulation here for longer than $t = 10^4$ time steps, why?)

(b) Modify the BA model as follows: every 10th time step instead of adding a node, delete $2m$ links at random.

(c) Modify the BA model to model A we learned in class (growth without preferential attachment).

(d) Invent your own simple modification to the BA model, describe it and implement it.

Submit the following:

- For each of the 4 modified networks and the BA model from the previous question, visualize the network when the network has 10^2 and 10^3 nodes (10 plots). Compare the modified networks to the original model, explain visual differences you see, if any.
- For each of the 4 modified networks, plot the degree distributions when the network has 10^4 nodes. Do the networks follow a power law distribution? Explain.