CptS 591: Elements of Network Science
Spring 2021
Assignment 1. Due by Fri Feb 19, 2021, 11:59pm

In this assignment you will do basic network analysis on a set of real-world networks and synthetically generated random networks. For the analysis, you will use the software package igraph.

Datasets: the assignment involves *three* networks from Mark Newman's collection of Network Data (`http://www-personal.umich.edu/~mejn/netdata/`). The networks in the collection are given certain descriptive names. The three networks chosen for this assignment are called: (1) Political blogs, (2) Neural network, and (3) Internet. Read the brief descriptions there so you get an idea for what the networks are modeling.

Your task: Download the three networks. These data sets are in GML format, which is a format the package igraph understands. "Read" the three networks into igraph. In addition to these three networks, generate three Erdos-Renyi random networks $G(n, p)$ using igraph, the first with the parameters $n = 2000$ and $p = 0.01$, the second with the parameters $n = 2000$ and $p = 0.005$, and the third with the parameters $n = 2000$ and $p = 0.0025$. Perform and report on the following tasks for all six networks.

Your submission will be one PDF file consisting of your solutions and an appendix describing the procedure you followed (simple listing of your igraph-python/R code).

1. (40%) Obtain/calculate the information/quantities for each of the six networks needed to complete the following table:

Network	Type	n	m	c	d	l	L	cc_l	cc_g

   where: Type indicates whether the network is directed or undirected; n is the number of nodes; m is the number of links; c is the number of connected components (in the case of a directed network, separately report on the number of strong and weak connected components); d is the maximum degree; l is the average path length; L is the diameter; cc_l is the average local clustering coefficient; and cc_g is the global clustering coefficient (3 times number of triangles/ number of connected triplets). Note that igraph calculates global clustering coefficient using its *transitivity_undirected* function and the average local clustering coefficient using its *transitivity_avglocal_undirected* function.

2. (25%) Plot the *degree* distribution of each network. This is a p_k vs k plot, where p_k corresponds to the probability that a randomly chosen node has degree k (in other words, p_k shows the fraction of nodes having degree k). What observations can you make from the plots?

3. (25%) Plot the *pathlength* distribution of each network. This is a p_l vs l plot, where p_l corresponds to the probability that two randomly chosen nodes have a shortest path of length l between them (in other words, p_l shows the fraction of node-pairs connected by a shortest path of length l.) Hint: see igraph's function called *path_length_hist*. What observations can you make from the plots?

4. (10%)) Do the analysis described in (1)–(3) on your favorite real-world network (of course outside the three described in the dataset for this assignment). This could be a network that your research is related to, or just a network that you are interested in learning something about. Remember to include a brief description of the network you worked with – at a minimum the description should state what the nodes and links are.