**Problem Set 1**

**Pioneer Course: Networks: from Searching the World Wide Web to Slowing the Spread of Infectious Diseases**

Warning this may be challenging and different than high school problem sets. If you have problems please email me for help and hints! Hint: use the networkx website (https://networkx.readthedocs.io/en/stable/) to look up functions and web search to find sample python code. (You should not need more than 1 or 2 python commands for any part of this problem set!)

Hand in all your results in a single pdf document (where you copied and pasted key results from your jupyter sessions). Include other files, typically the Jupyter file, separately.

Read chapter 1 in the textbook

1. Watch the video “How Kevin Bacon Cured Cancer. https://www.youtube.com/watch?v=2rzxAyY7D7k

Write two questions that you want to know after watching it?

1. Write a paragraph or two describing something that you are interested in and how it could be modeled/studied using networks. Make sure to be precise: what are the nodes? When are the edges?
2. Find 2 interesting networks on the WWW and download them. One should be in Pajek (.net) format and the other in a different network format and should be read into networkx using their correct reader: https://networkx.readthedocs.io/en/stable/reference/readwrite.html
   1. Using networkx, read them, plot them, compute some of their properties and write a paragraph about them and your analysis of them. Hint: Do a web search for “networkx read graph” to see what formats networkx can read.
   2. Look at each of the files using a text editor and try to understand the format. Describe it in your own words. (Don’t look this up using google, try to do it on your own.)
3. Generate some networks using networkx. For each one, plot them for n=5,10,25 (where n is the number of nodes) and compute the number of nodes and edges, average clustering coefficient and average path length. Summarize your results and try to explain them using the networkx documentation. (Note that n is the number of nodes in the graph for each of these.)
   * 1. cycle graph
     2. complete graph
     3. gnp graph for p=0.2
     4. gnm graph for m=n/5
4. Look at the networkx documentation for the “gnp\_random\_graph” function and click on the “source” link to see the underlying code. Try to explain how it constructs the random network. (Only consider the undirected version.)
5. Consider a gnp random graph (not the “fast gnp”) as a function of n and p. Find the largest values of n, for p =0.01 and 0.5, which don’t crash and run in under 2 minutes. Compute and plot the time it takes to construct these graphs as a function of n for both p’s.
6. Estimate the amount of time you spent on this problem set and add any comments you might have on it. (This will help me with future problem sets.)