Network Science
PHYS 5116, Fall 2016

Prof. Albert-László Barabási, Dr. Sean P. Cornelius, Dr. Emma K. Towlson

Assignment 2, due by Friday, November 4th, 6 PM

Write your name at the top of your assignment before handing it in. Staple all pages together. If you hand in a digital/scanned copy, please be nice to your instructors and hand in a ***single****, combined file. And please include your name in the file name (e.g.* Cornelius.pdf*)!*

1. **Barabási-Albert networks, plotting things in log-log scale**
   Problem 5.12.1 in the book. Various network libraries including NetworkX already have implementations of the BA model, but you will need to implement it yourself in order to do parts (a) and (e). Include your source code with your solution. When you plot things for parts (a), (b), and (c), remember all the rules we learned in class about plotting things in log-log scale!

2. **Network visualization**
   Download the dataset *netscience.zip* from the website. As described in the text file, this is a coauthorship network of scientists working on network theory and experiment compiled by Mark Newman in 2006. Visualize the largest component of the network (Gephi may be the easiest choice of software, but you are free to use your favorite), considering the following:

   - Degree
   - Other measures of centrality
   - Community structure

   Make appropriate use of color, size and layout to create a clear and informative visualization. Describe your approach and comment on your observations (about a paragraph for each).

3. **Degree Correlations**

   Problem 7.9.1 from the book.

4. **Communities and modularity resolution limit**
   Consider a network consisting of C subnetworks of m nodes each. Every subnetwork is a one-dimensional lattice, in which the nodes lie on a circle and each connects to its k nearest neighbors on either side. This is the starting point for the Watts-Strogatz model before any rewiring occurs. The subnetworks are in turn arranged in a big ring, and each is joined to its two neighboring subnetworks by one link each.

   (a) Calculate the expected modularity of the natural partition of this network, *i.e.* the one in which the communities are the C subnetworks.

   (b) Calculate the expected modularity of the partition in which there are $C/2$ communities, consisting of neighboring *pairs* of the original C subnetworks.

   (c) Comparing your answers to parts (a) and (b), show that for given values of m and k, there is an upper limit to the number of subnetworks C, beyond which the "natural" partition violates the maximum modularity hypothesis. Your answer should be in terms of m and k.

   (d) When C is at the upper limit you derived above, what is the *expected* number of links between two adjacent communities (according to a null model)? Use this answer to interpret your finding in (b). Why does the limit exist? Specifically, why do other partition than the "natural" one look like more plausible descriptions of the community structure (from a modularity standpoint) once this limit is exceeded? How does varying m and k up or down change the limit and why?

5. **Boot camp: Centrality, modularity. (Compulsory for Network Science PhD students only)**
   Quite often in real datasets, you will find that the various network centrality measures are highly correlated with one another. To understand the differences between these measures, for each of the following 6 cases, come up with an example network and a particular node in that network that has:

a) High closeness centraliy but low degree centrality, and vice versa.

b) High betweenness centrality but low closeness centrality, and vice versa.

c) High degree centrality but low betweenness centrality, and vice versa.

For each of these, you should hand-draw the network you have in mind, highlight the node, and explain why it satisfies the given criteria. You should conceive of examples simple enough that you can *manually* calculate the two centrality measures (at least approximately or asymptotically), directly showing the disparity between the two centrality measures. So do that.

Now, manually calculate the modularity of the following networks and explain why numbers come out the way they do:

d) The “star” network on N nodes. This is a single hub connected to each of $N-1$ “leaf” nodes of degree 1.

e) A wheel network on N nodes. I.e., the star network above but with additional links between pairs of adjacent leaves (so that they each have degree 3).

f) The “lollipop” network: a complete graph on m nodes connected to a chain of k nodes.

As always, answers should be functions only of the provided parameters.