# Math 168, Networks, UCLA, Spring 2018
## *Problem Sheet 2*

(submit to CCLE by Monday 16 April 2018 at 5:00 pm)

1. *Reading.* Read Chapter 6 of Newman's book.
2. Do Problem 6.1 of Newman's book.
3. Do Problem 6.3 of Newman's book.
4. Do Problem 6.5 of Newman's book.
5. *Empirical Data and Visualization.*
   (a) Download data for at least two small unweighted, undirected networks of somewhat different sizes (e.g., one of them should have at least (roughly) 5–10 times as many nodes as the other). Cite your sources for these data sets.

   As you can see, numerous data sets are available online. [Note: The Colorado Index of Complex Networks is one good place to look.]

   (b) Using some software, draw visualizations of these networks. One possibility is the following: `http://netwiki.amath.unc.edu/VisComms/VisComms`. There are numerous other software packages.

   (c) Plot the degree distribution for each of the two networks that you downloaded. What are you able to conclude from these degree distributions?

6. *Random graphs.* In later lectures, we will discuss ensembles of random graphs. The simplest and most famous type of random graph is an Erdős–Rényi (ER) graph (which was first studied by Solomonoff and Rapoport).

   In particular, consider the random-graph ensemble $\mathcal{G}(N, p)$, which is defined as follows: Suppose that there are $N$ nodes. Between each pair of distinct nodes, a single edge exists with uniform and independent probability $p$. There are no self-edges. A single graph $G \in \mathcal{G}(N, p)$ is generated using this process, and it is interesting to study the properties of collections ("ensembles") of graphs that are generated in this way. The probability with which each simple graph $G$ with $m$ edges appears in a graph $G \in \mathcal{G}(N, p)$ is

$$P(G) = p^m (1-p)^{{}_N C_2 - m} \,. \tag{1}$$

   (a) Write down the total probability of drawing a graph $G$ with exactly $m$ edges from the ensemble $\mathcal{G}(N, p)$, and use it to find the mean number $\langle m \rangle$ of edges.

   (b) Calculate the expected mean degree of an ER graph.

   (c) Show, under an appropriate assumption (which you should state), that the degree distribution for an ER graph (in expectation over the ensemble) satisfies

$$p_k \sim e^{-c} \frac{c^k}{k!} \,, \quad N \to \infty \,, \tag{2}$$

   where $c = (N-1)p$.

   (d) By doing calculations in MATLAB, or in some other program, compare the expectations that you calculated above to sample means from a set of ER networks. Indicate explicitly what sizes and parameter values you consider. Also indicate explicitly the number of elements in your ensemble. What happens as you take a sample mean over a larger number of ER graphs?

**Note 1:** Writing "in expectation over the ensemble" gets repetitive rather quickly. In the applied literature, when one writes something like "the degree distribution of an ER graph", it is normally understood to be a calculation that is in expectation over the ensemble. To be more precise, one can use notation like $G$ for a single realization and $\mathcal{G}$ for an ensemble. Expectations over graph ensembles are often compared to "sample means", in which one calculates a quantity of interest for each of some number of draws from a graph ensemble and then averages over those results. In Newman's book, he often writes about "averaging over" things, and you should compare that language to what I have stated in this paragraph.

**Note 2:** The term "ensemble" comes from statistical mechanics. Using more mathematical language, a random-graph ensemble is a probability distribution on graphs.