1 Introduction

This problem set has 3 theoretical problems and 1 experimental problem. To do the experimental problem, you must be enrolled in the course through Classes V2. See the rest of the instructions below.

The problem set is due in class on Thursday, Sept. 12. If you cannot come to class, you can put it in Dan Spielman’s mailbox on the first floor of AKW before then.

2 Homework Policy

You may discuss the problems with other students. But, you must write your solutions independently, drawing on your own understanding. You should cite any sources that you use on the problem sets other than the textbook, TFs and instructor. This means that you should list your collaborators.

It goes without saying that you should write your own code for the experimental part.

You may not search the web for solutions to similar problems given out in other classes. If you think this policy needs any clarification, please let me know.

Problem 1: Isolated Vertices

In Lecture 2, we proved that in graphs chosen from the distribution $\mathcal{G}(n,p)$ with $p = (1 - \epsilon) \ln n / n$, the expected number of isolated vertices is at least $n \epsilon / 2$ (I remark that you could improve this bound to make it much closer to $n \epsilon$). However, this does not necessarily imply that there is probably an isolated vertex. We will show that the probability of there being an isolated vertex is at least some fixed constant, independent of $n$.

a. Prove that it is very unlikely that there are more than $2en \epsilon$ isolated vertices.

b. Prove that there is some absolute constant $c > 0$, possibly depending on $\epsilon$, so that with probability at least $c$ a graph chosen from $\mathcal{G}(n,p)$ has at least one isolated vertex.

**Hint:** Combine part a with the bound on the expected number of isolated vertices.
Remark: Using more sophisticated techniques, one can prove that this holds for $c$ close to 1. I’ll tell you how later.

Problem 2: Choose-2 Graphs

Here’s a model of a random graph that I call a “choose-$k$” graph. In this model, each vertex chooses $k$ vertices uniformly at random. We make these choices independently for each vertex, and then add edges between each vertex and the $k$ vertices that it chooses. I will allow these choices to be made with replacement, and I will allow a vertex to choose itself. When a vertex chooses itself, we will just skip that edge. So, a vertex could get fewer than $k$ neighbors if it chooses some neighbor more than once or if it chooses itself.

To be concrete, the following procedure could be used to create the graph:

1. Set $E = \emptyset$
2. For $a = 1$ to $n$
   a. For $i = 1$ to $k$
      Choose $b$ uniformly in $\{1, \ldots, n\}$ and, if $b \neq a$, add $(a, b)$ to $E$.

The resulting graph will have close to $kn$ edges. Some vertices could have degree much larger than $2k$ if they are chosen many times. While these vertices have natural directed analogs, we will consider the undirected versions.

Prove that for sufficiently large $n$, a random choose-2 graph is probably connected. I suggest you begin by proving that it is unlikely that it has any isolated vertices.

Problem 3: Choose-$k$ Graphs

Prove that there is some constant $k$ so that for sufficiently large $n$, a random choose-$k$ graph with $n$ vertices probably has diameter less than $2 \ln n$.

You probably cannot solve this problem until after the third lecture.

Problem 4: Experiments in Percolation

We will study Erdös-Rényi random subgraphs of a given graph. The data gathered in these experiments will inform our discussion of percolation in later lectures.

In this problem, you will be assigned a graph. Your job will be to run experiments on the graph, and to report the results. In each experiment, there will be some number $p \in (0, 1)$. You will
choose to keep each edge in the graph with probability \( p \), independently. You will then measure the size of the largest connected component in the resulting graph.

To get your graph, follow the links on the course homepage under “PS1”. It will lead you to a web server that will ask for your Yale NetID. You must be registered for the class on Classes V2 to get a graph.

Once you have your graph, you should run the experiment for two different ranges of \( p \). The first is \( i/10 \) for \( i \) an integer in 1 to 9. The second is \( i/(4d_{ave}) \) for \( i \) an integer in 2 to 10, where \( d_{ave} \) is the average degree of your graph. If your graph has fewer than 100,000 nodes, you should run each experiment 100 times. If it has more nodes, you can run each experiment 10 times. For each range, you should create a text file containing your results as a comma separated table, with the probabilities on the top line. Each column should list the largest components from each run of the experiment, again separated by commas (this format is called CSV). You will later upload these files to the same server. You may include extra whitespace in your files, but you should not have other text: you want to make it easy for the server to parse your file.

If you go to the web page for PS1, you will find example files that I generated for the graph called grid2. I recommend that you check if you get similar results.

You should submit an explanation of how you ran this experiment, such as a printout of code along with an explanation of how you used it, with the problem set.

Notes:

1. We want to do this study on undirected graphs. Check if your graph is undirected. If it is not, make it so.

2. Your graph might not be connected or have isolated vertices. That is OK.

3. This list might be updated later.