Homework 2 due Feb 6

Select five problems from the list below that interest you and solve them. Create one additional problem that interests you and solve it. Also carry out the mini project.

Mini project: Look on the web for a data base that can be converted to an undirected graph. For example in science there is a data base of proteins and interactions. Each protein can be represented by a vertex and two proteins that interact are connected with an edge. Find a data set that will yield a graph with at least 1000 vertices and a number of edges that is roughly in the range or one to two times the number of vertices. Thus the graph will be quite sparse and have many connected components. Find all connected components and create a table with the number of components of each size. So for example there may be 312 isolated vertices. 124 components consisting of two vertices connected by an edge, 29 components with three vertices, ect.

Exercise 1: In $G\left(n,\frac{1}{n}\right)$ what is the probability of a vertex of degree $\log n$?

Exercise 2: How do you create a graph with statistical properties of an airline route map?

Exercise 3: Simplify $\left(1-\frac{1}{n}\right)^{n\log n}$

Exercise 4: Let f(n) be a function that is asymptotically less than n. Some such functions are 1/n, a constant d, logn or $n^{\frac{1}{3}}$. Show that

$$\lim_{n\to\infty}\left(1+\frac{f(n)}{n}\right)^n = e^{f(n)}.$$

Exercise 5: Evaluate $\left(1-\frac{1}{2^k}\right)^{2^k}$ for k=3, 5, and 7. How close is it to 1/e?

Exercise 6: Generate a graph $G\left(n,\frac{d}{n}\right)$ with n=1000 and *d*=2, 3, and 6. Count the number of triangles in each graph. Try the experiment with *n*=100.

Exercise 7: The global clustering coefficient of a graph is defined as follows. Let d_v be the degree of vertex v and let e_v be the number of edges connecting vertices adjacent to vertex v. The global clustering coefficient c is given by

$$c = \sum_v \frac{2e_v}{d_v(d_v-1)}$$

In a social network, for example, it measures how many pairs of friends of each person are themselves friends. If many are, the clustering coefficient is high. What is c for a random graph? Compare this value to that for some social network.

Exercise 8: Let G(n,p) be a random graph and let x be the random variable which is the number of unordered pairs of non adjacent vertices(u,v) such that no other vertex of G is adjacent to both u and v. Prove that if $\lim_{n\to\infty} E(x) = 0$, then for large n there are almost no disconnected graphs i.e. $\text{Prob}(x=0)\to 1$ and hence $\text{Prob}\left(\text{G is connected}\right)\to 1$.

Exercise 9: Prove that the expected shortest distance between two nodes in a connected component of a random graph G(n,1/n) is $\log n$. If this is not correct what is the correct answer.

Exercise 10: Let $S(n,p)$ be a random subset of $\{1,2,\cdots,n\}$ of size k. Consider the property P that $S(n,p)$ contains a 3-term arithmetic progression. Show that $n^{\frac{1}{3}}$ is a threshold for property P.

Exercise 11: Try to find either a necessary or a sufficient condition for a monotone property to have a sharp threshold. **Research problem.**

Exercise 12: Is there a good example of a Boolean function where we can exhibit the above phenomena?

Exercise 13: Consider a random process for generating a Boolean function *f* in conjunctive normal form where each of c clauses is generated by placing each of *n* variables in the clause with probability *p* and complementing the variable with probability ½. What is the distribution of clause sizes for various *p* such as $p=\frac{3}{n}$, ½, other values? Experimentally determine the threshold value of *p* for *f* to cease to be satisfied.

Exercise 14: Consider graph 3-colorability. Randomly generate the edges of a graph and compute the number of solutions and the number of connected components of the solution set as a function of the number of edges generated. What happens?

Exercise 15: For a random CNF formula with n variables how many satisfying assignments are there as the number of variables increases?

Exercise 16: Let $M(n,p)$ be the multi set formed by drawing *pn* integer from the set $\{1,2,\cdots,n\}$ with repetition.

a) How large must p be in order to have some integer appear twice?
b) Does a sharp transition occur?
c) How large must p be in order for every integer to occur in $M(n,p)$?
d) How does the frequency of occurrences of integers evolve with increasing p?

Exercise 17: In $N(n,p)$ what is the threshold for $N(n,p)$ to contain a) a perfect square, b) a perfect cube, c) an even number, d) three numbers such that x+y=z

Exercise 18: Modify the proof that every monotone property has a threshold for sets of integers to work for a) G(n,p) b) 3-CNF SAT.

Exercise 19: The threshold property seems to be related to uniform distributions. What if we considered other distributions. Consider a model where *i* is selected with probability $\frac{p(n)}{i}$. Is there a threshold for perfect squares? Is there a threshold for arithmetic progressions?

Exercise 20: Explain why the property that $N(n, p)$ contains the integer 1 has a threshold. What is the threshold?

Exercise 21: Consider G(n,p).
a) Where is phase transition for 3-colorability?
b) For vertex cover?

Exercise 22: Is there a condition for a sharp threshold?

Exercise 23: List five increasing properties – connectedness, Hamilton circuit. List five non increasing properties – even number of edges, odd number of connected components.

Exercise 24: Consider generating a random graph by flipping two coins, one with probability p_1 of heads and the other with probability p_2 of heads. Create an edge if either coin comes down heads. The graph generated will be in G(n,p) for what value of p.

Exercise 25: Prove that if Q is an increasing property and $p_2 > p_1$, then

$$\text{Prob}\left[G(n, p_1) \text{ has Q}\right] \leq \text{Prob}\left[G(n, p_2) \text{ has Q}\right]$$

Exercise 26: Is there a threshold for Sudoku?

Exercise 27: Show that when the threshold for connectivity of g(n,p) is reached the graph has diameter $O(\log n)$.

Exercise 28: Select n points in R^d. For small n all points lie in some half space bounded by a hyper plane through the origin. For large n they do not. What is the threshold?

Exercise 29: Birthday problem: What is the number of integers that must be drawn with replacement from a set of n integers so that some integer will be selected twice? Show that a phase transition occurs at $\sqrt{n}$.