Problem Set 3

Out: 29 Nov.; Due: 14 Dec.

Notes: *These problems are designed to test your understanding of the last part of the class (Lectures 21 through 26). Put your solutions in the box outside my office—677 Soda Hall—by* **5pm on Wednesday Dec. 14**. *Late solutions will not be accepted. Depending on grading resources, we reserve the right to grade only a subset of the problems (but since you don't know which subset, you're strongly advised to do all of them!).*

Take time to write **clear** *and* **concise** *answers. None of the problem parts requires a long solution; if you find yourself writing a lot, you are either on the wrong track or giving an inappropriate level of detail. You are encouraged to form small groups (two or three people) to work through the problems, but you* **must** *write up your solutions on your own.*

1. A contact process on a graph

Let $G = (V, E)$ be an arbitrary connected undirected graph with n vertices and m edges. Consider the following random process on G, which is a very simple model for the spread of rumors or infections. Initially, each vertex of G is colored either black or white (with no constraints). Then, at each step, all vertices simultaneously update their colors, independently of all other vertices, as follows:

- with probability $\frac{1}{2}$ do nothing
- else pick a neighboring vertex u.a.r. and adopt the color of that vertex

(Note that *all* vertices make these decisions before any vertex changes color.) It should be clear that this process will eventually terminate with all vertices black or all white, after which no further change is possible.

(a) Let the random variable X_t denote the sum of the degrees of all the white vertices at time t. Show that X_t is a martingale with respect to the obvious filter.

(b) Use the optional stopping theorem to compute the probability that the process terminates in the all-white configuration, as a function of the initial configuration.

(c) Use the optional stopping theorem again to show that the expected duration of the process is at most $O(m^2)$ steps. [NOTE: Justify carefully any claims you make about the conditional variance of X_t.]

2. Vertex-disjoint cycles

Let G be an arbitrary k-regular directed graph (i.e., every vertex has in- and out-degree k). In this problem, we will show, using the Lovász Local Lemma (LLL), that G contains at least $\lfloor \frac{k}{3 \ln k} \rfloor$ vertex-disjoint cycles.

(a) Suppose the vertices of G are partitioned into $c = \lfloor \frac{k}{3 \ln k} \rfloor$ components by assigning each vertex to a component chosen independently and u.a.r. For each vertex v, let A_v be the event that v has no edge to another vertex in its component. Show that $\Pr[A_v] \leq k^{-3}$.

(b) Let D_v denote the "dependency set" of event A_v (i.e., A_v is independent of all events A_u except for those in D_v). Show that $|D_v| \leq (k+1)^2$.

(c) Deduce from parts (a) and (b) and the LLL that G contains at least c vertex-disjoint cycles.

3. A hands-on approach to making the Lovász Local Lemma algorithmic

Recall from class (Lecture 22) that we proved using the Lovász Local Lemma (LLL) that any k-SAT formula in which no variable appears in more than $2^k/k$ clauses must be satisfiable. However, this proof does not lead to a useful algorithm because the probability of the random assignment being satisfying is exponentially small. In this problem, we will see how to derive an algorithmic version of this result, with rather weaker constants. (Of course, we could always appeal to the general Moser-Tardos constructive version of the LLL to get an optimal result, but it is instructive to do this by hand.)

Specifically, we will prove that if ϕ is any k-SAT formula such that no variable appears in more than $2^{k/20}/k$ clauses, then there is a randomized algorithm that finds a satisfying assignment of ϕ with high probability in time polynomial in the number of clauses, n. Throughout we will assume that k is a sufficiently large constant (and in fact the running time will be exponential in k).

We will make heavy use of the notion of the "dependency graph" associated with applications of the LLL, as mentioned in class; in the context of satisfiability, this is the graph whose vertices are clauses, with an edge between two vertices iff the two clauses share a variable. Note that for ϕ as above, the degree of the dependency graph is at most $d = 2^{k/20}$.

(a) In Stage 1 of the algorithm, we assign truth values to the variables of ϕ one at a time by flipping a fair coin. If at any time a clause has more than $k/2$ of its variables set, but is not yet satisfied, we *freeze* all the remaining variables in the clause, i.e., we don't set their values in Stage 1. At the end of Stage 1, we will be left with (presumably many) satisfied clauses, plus some as yet unsatisfied clauses containing frozen variables. Call these latter clauses "surviving." Note that every surviving clause must contain at least $\lfloor k/2 \rfloor$ frozen variables.

Let ϕ' denote the reduced formula consisting only of surviving clauses and frozen variables. Use the (symmetric) LLL to show that ϕ' is satisfiable (and hence the partial assignment of Stage 1 can be extended to a satisfying assignment).

(b) Let G' denote the dependency graph of ϕ'. Stage 2 of the algorithm is based on the following:

Claim: *With probability $1 - o(1)$, no connected component of G' has size larger than $C \log n$, where C is a constant (that depends on k).*

Without proving the Claim, show how it (together with part (a)) leads to a polynomial time randomized algorithm that finds a satisfying assignment with high probability.

(c) In preparation for proving the Claim, let G denote the dependency graph of the original formula ϕ, and G_4 the graph obtained from G by connecting only those vertices whose distance in G is *exactly* 4. Show that, to prove the Claim, it is enough to show that G' contains no set S of vertices of size larger than $(C \log n)/d^3$ such that all elements of S are at pairwise distance at least four in G and S is connected in G_4.

(d) Now let S be a set of vertices in G of size r, such that all elements of S are at pairwise distance at least four in G. Argue that the probability that all clauses in S survive Stage 1 is at most $\left(2^{-k/2}(d+1)\right)^r$. [HINT: These events are independent—why?]

(e) Show that the number of connected components S in G_4 of size r is at most md^{8r}. [HINT: Each connected component can be described by a spanning tree, which in turn can be described by a closed walk around its exterior. G_4 has m vertices and degree $\leq d^4$.]

(f) Combine parts (c), (d) and (e) to deduce the Claim in part (b). This completes the analysis of the algorithm.

4. Random walks and electrical networks

This problem concerns the connection between random walks on undirected graphs and electrical networks discussed in Lecture 23. In this problem, you will probably need the classical "Rayleigh's Monotonicity Principle," which states that the effective resistance R_{uv} between any pair of vertices u, v cannot be increased by decreasing resistances in the graph (which includes "shorting" an edge by merging its endpoints, and adding new edges), and cannot be decreased by increasing resistances in the graph (which includes removing edges).

(a) Let G be a regular graph with n vertices. Show that the cover time of G is at most $O(n^2 \log n)$. [NOTE: This is much smaller than the cover time of general graphs, which as we saw in class can be as large as $\Theta(n^3)$. With a more careful analysis, the $\log n$ factor can actually be dropped.]

(b) Now let G be a regular graph in which every vertex has degree strictly larger than $\frac{n}{2}$. Prove that the cover time of G is $O(n \log n)$. [HINT: Show that, between every pair of vertices u, v, there are $\Omega(n)$ edge-disjoint paths of length at most four.]

(c) Give an example of a regular graph with all vertex degrees $\frac{n}{2} - O(1)$ whose cover time is $\Omega(n^2)$; you should justify the value of the cover time. [NOTE: It should be easy to generalize your example to smaller degrees. Together with part (b), this establishes a "phase transition" in the cover time for regular graphs that occurs at degree roughly $\frac{n}{2}$.]

5. Another slow way to shuffle cards

Here is yet another card-shuffling process (assuming as usual a deck of n cards). At each step, pick a random pair of *adjacent* cards and switch their positions. To eliminate periodicity, we also allow a probability $\frac{1}{n}$ of doing nothing at every step. Thus, more formally, at each step we pick $j \in \{0, 1, \dots, n-1\}$ uniformly at random; if $j = 0$ do nothing, else switch the cards in positions j and $j + 1$.

(a) Briefly justify why this process will converge to the uniform distribution on permutations of the deck.

(b) Consider the following coupling for this Markov chain. Let X, Y denote the permutations of the two decks of cards at some time. Define $S = \{j_1, \dots, j_k\}$ to be the set of positions j such that *neither* the cards in position j *nor* the cards in position $j + 1$ are matched. Also, add $j_0 = 0$ to S. We let X choose position $j \in \{0, 1, \dots, n-1\}$ u.a.r. Then Y chooses its position j' as follows:

$$j' = \begin{cases} j & \text{if } j \notin S; \\ j_{i+1} & \text{if } j = j_i \in S \text{ (where } j_{k+1} \text{ is understood as } j_0). \end{cases}$$

Finally, X and Y make their moves according to j and j' respectively. This is clearly a valid coupling as both j and j' are uniform. *Make sure you understand this coupling before proceeding.*

Using the above coupling, argue that the mixing time of this shuffle is at most $O(n^4)$. [HINT: First, show that the coupling never destroys matches, and also that no card in one deck can "jump over" the same card in the other deck. Then think about the expected time for any given card in one deck to reach the bottom.]

(c) Improve the above mixing time bound (using the same coupling) to $O(n^3 \log n)$. [NOTE: This is very close to the precise answer, which is $\Theta(n^3)$.]