College of Computer & Information Science — Spring 2016
Northeastern University — Handout 2
CS7880: Network Algorithms and Analysis — 14 February 2016

Sample Solution to Problem Set 2

Problem 1. (5 + 5 = 10 points) Condition for decoding in gossip protocol

In our study of gossiping using network coding, we considered the following scenario. We have a network of n nodes, each having an L-bit message that they want to share with all the other nodes. Each message $\vec{m_i}$ is an element of $\mathcal{F}_2^L$. We studied a network coding algorithm in which each node broadcast a packet consisting of two vectors, an n-bit vector $\vec{\mu}$ from $\mathcal{F}_2^n$, and an L-bit message vector $\vec{m}$ equaling the linear combination $\sum_i \mu(i)\vec{m_i}$, where $\mu(i)$ represents the ith component of the $\vec{\mu}$ and the sum is the component-wise addition mod 2. The particular message vector that a node transmits is drawn uniformly at random from the subspace spanned by the messages that it has received thus far.

We further showed the following statement: if S_v denotes the coefficient vectors received by v, and there is no vector in $\mathcal{F}_2^n$ that is orthogonal to the subspace spanned by S_v, then v can decode all of the n original messages.

Let B_n denote the set of n vectors in F_2^n:

$$\{(1, 0, 0, \ldots, 0), (0, 1, 0, \ldots, 0), \ldots, (0, 0, \ldots, 1)\}.$$

(a) Prove or disprove: B_n is a basis for F_2^n in the sense that every vector in F_2^n can be written as a linear combination of vectors in B_n.

Answer: True. Let n vectors in B_n be denoted e_i through e_n. Any vector $v = (v_1, \ldots, v_n)$ is simply $\sum_{1 \le i \le n} v_i e_i$.

(b) Show why the following condition is not sufficient for decoding: none of the n vectors in B_n is orthogonal to the subspace spanned by S_v.

Answer: Let n be even, and S_v be the subspace spanned by the singleton set $\{x = (1, 1, 1, \ldots, 1)\}$. Then $e_i \cdot x$ equals 1 for each i. Yet, none of the n messages can be decoded.

Answer:

Problem 2. (10 + 5 + 15 = 30 points) SIS model on the star network

Let S_n denote the star graph with n vertices: a root node 0 with edges to the remaining $n - 1$ nodes labeled 1 through $n - 1$.

(a) Show that the largest eigenvalue of the adjacency matrix of S_n is $\sqrt{n-1}$.

Answer: By Courant-Fischer, the largest eigenvalue is given by the following, where A is the adjacency matrix of S_n

$$\max_{x \neq 0} \frac{x^T A x}{x^T x}.$$

For $x = (x_1, \ldots, x_n)$, we have

$$x^T A x = 2x_1(x_2 + x_3 + \ldots + x_n).$$

We need to maximize this subject to the condition that $\sum_i x_i^2 = 1$. It is easy to see that all x_i, for $i > 1$, are equal. So we need to maximize $2(n-1)x_1 x_2$ subject to the condition that $x_1^2 + (n-1)x_2^2 = 1$. Setting $x_1^2 = 1-(n-1)x_2^2$, we obtain $2(n-1)x_1x_2 = 2(n-1)\sqrt{1-(n-1)x_2^2}x_2$, which is maximized when $(n-1)x_2^2 = 1-(n-1)x_2^2$, or when $x_2 = 1/\sqrt{2(n-1)}$. So we obtain $x_1 = 1/sqrt2$ and the largest eigenvalue equal to $\sqrt{n-1}$.

(b) Show that the edge expansion of S_n is 1.

Answer: Consider any cut $(X, \overline{X})$ in S_n. Without loss of generality, let X have size at most $n/2$. If X contains the root, then the number of edges from X to $\overline{X}$ is $|\overline{X}| \geq |X|$; otherwise, the number of edges from X is $|X|$. So the ratio of the number of edges to the size of the smaller set of the cut is at least 1. And there exists a cut (with one side being any leaf node) that has an expansion of exactly 1.

From the results that we have shown in class it follows that for S_n (i) if the infection rate $\beta < 1/\sqrt{n}$, then the epidemic die-out occurs in expected $O(\log n)$ time, assuming a recovery rate of 1 (per unit time); and (ii) if the infection rate $\beta > 1$, then the epidemic die-out occurs in expected exponential time.

We want to understand what happens when $1/\sqrt{n} \leq \beta \leq 1$.

(c) In this exercise you need to derive an upper bound on the expected epidemic die-out time when $\beta = C/\sqrt{n}$ for a constant $C > 1$.

We will consider a simplified discrete variant of the process. Assume that every node is infected at time 0. Let $S_j(t)$ be 1 if j is infected at t, and 0 otherwise. Let $H_j(t)$ denote a random variable that takes 1 with probability 1/2, and 0 otherwise (this represents healing of node j, if infected). Let $R_j(t)$ denote a random variable that is 1 with probability β, and 0 otherwise (this represents receipt of an infection from 0). Finally, let $T_j(t)$ denote a random variable that is 1 with probability β, and 0 otherwise (this represents transmitting an infection from 0). Define for $0 < j < n$ and $t > 0$:

$$S_j(t+1) = \max\{0, S_j(t) - H_j(t), S_0(t)R_j(t)\}.$$

Finally, define for $t > 0$:

$$S_0(t+1) = \max\{0, S_0(t) - H_0(t), \max_{j>0}(S_j(t)T_j(t))\}.$$

Show that in $\tau = O(\log n)$ steps, $S_j(\tau)$ is 0 for all j with probability $1 - 1/n^\alpha$ for some constant $\alpha > 0$.

Answer: We provide a sketch of the proof, leaving the details. We will establish that the expected time to die-out is $O(\log n)$, as opposed to establishing a high probability bound as mentioned above. A high probability bound can be established by a much more careful argument in the second phase of the analysis below.

The first thing to observe is that if the root is infected, the expected number of leaves it will infect is at most $C\sqrt{n}$. Using a standard Chernoff bound, this number is at most $C(1+\varepsilon)\sqrt{n}$ for any constant $\varepsilon > 0$, with very high probability $(1 - e^{-\Omega(\sqrt{n})})$.

Our second observation is that if x be the number of leaves that are infected at the start of a round, then the number of leaves that will recover during this round is $x/2$ (ignoring the possibility that they will be infected by the root). Furthermore, by a Chernoff bound, the probability that the number of recovered leaves is at most $x/4$ is at most $e^{-x/8}$. Thus, with probability at least $1 - e^{-x/8}$, the number of remaining infected leaves is at most $3x/4$.

Putting the above two observations together, we obtain that in a sequence of $O(\log n)$ rounds, the number of infected leaves is $O(\sqrt{n})$ with probability at least $1 - e^{-\Omega(\sqrt{n})}$. Let us call this the first phase, which succeeds with probability at least $1 - e^{-\Omega(\sqrt{n})}$.

We now reach an interesting second phase of the analysis. The key challenge is that if the root is infected in some round, $\Omega(\sqrt{n})$ leaves may be infected the next. Consequently, for the infection to die out, the root should be cured and not be infected for a non-trivial period of time, while the number of infected leaves gets smaller and smaller.

Let us suppose first that the root recovers and is remains infection-free. Starting with a state of L infected leaves, under the assumption that the root is continually not infected, the probability that the number of infected leaves decreases by a factor of at least $3/4$ every round until no leaves are infected, is at least the following product, where k is the largest integer such that $L3^k/4^k > 1$.

$$\prod_{0 \le i \le k} \left(1 - e^{-3^i L/(8 \cdot 4^i)}\right) \quad \ge \quad \delta \left(1 - e^{-x/8} - e^{-4x/(3 \cdot 8)} - \cdots - e^{-4^i x/(3^i \cdot 8)} + \cdots\right),$$

where x is a sufficiently large integer constant, and δ is a sufficiently small positive constant, dependent on x. For δ and L suitably chosen, the above bound is at least γ for some positive constant γ.

The above probability was under the assumption that the root remains infected. We now argue that there is a constant probability that the root is not infected during the above process.

Let L denote the number of infected leaves in a given round. The probability that the root is not infected by any of the leaves equals $(1 - C/\sqrt{n})^L \ge e^{-2LC/\sqrt{n}}$, where we use the fact that $1 - x \ge e^{-2x}$ for $x > 0$ sufficiently small. By the above argument,the probability that the number of infected leaves is at most $3L/4$ the following round, and continues to decrease by a factor of $3/4$ (assuming the root is not infected) is at least γ. Since the events of one round are independent of the events of subsequent rounds, the probability that root never becomes infected during the above process in which the number of infected leaves decreases by a factor of at least $3/4$ is at least

$$\prod_{0 \le i < k} e^{-2CL3^i/(4^i\sqrt{n})} = e^{-(CL/sqrtn)\sum_{0 \le i < k} 3^i/4^i} \ge e^{-8CL/\sqrt{n}},$$

which is at least a positive constant η for $L = \Theta(\sqrt{n})$.

We thus obtain that the probability that once we have a state of L infected leaves, the probability that the root recovers and then remains infection free, and all the infected leaves recover over $O(\log n)$ rounds is at least $\gamma\eta/2$ for positive constants γ and η. This is the probability of success of the second phase of the process.

Let ν equal the product of $\gamma\eta/2$ and $1 - e^{-\Omega(\sqrt{n})}$, the success probabilities of the first two phases, each of which run for, say $T = O(\log n)$ rounds. The expected number of rounds before which all nodes are infection free is at most

$$T(\nu + 2(1-\nu)\nu + 3(1-\nu)^2\nu + \cdots) \leq T/\nu = O(\log n).$$

Problem 3. (15 + 5 + 15 = 35 points) Myopic routing on a network with randomly chosen long-range links

Consider n nodes labeled 0 through $n-1$ organized in a ring network; so every node has exactly two neighbors. Now suppose each node has two additional *long-range links*, each of which is to a node in the ring drawn uniformly at random, with replacement. Let G denote the resulting random graph. The total number of edges is exactly $3n$. (Note that the graph is undirected; even though the long-range links are selected by each node in a directional manner, the edges formed are undirected.)

(a) Show that the diameter of G is $O(\log n)$ with high probability, i.e., with probability at least $1 - 1/n^c$ for some constant $c > 0$, where the hidden constant in the big-Oh notation bound for the diameter may depend on c.

Answer: Fix arbitrary nodes u and v. Let B_u (resp., B_v) denote the set of $2c\log n$ nodes nearest to u (resp., v) in the ring, for a constant c that will be chosen sufficiently large. For any set S of nodes let $L(S)$ denote the set of nodes that are one-hop away from S through the long-range links from the nodes in S. We first prove that there exists a constant $c' > 0$ such that for any set S of size in $[2c\log n, n/c']$, the size of $L(S)$ is at least $3|S|/2$ with probability $1 - 1/n^c$. The event that $|L(S)|$ is smaller than $3|S|/2$ implies that all of the $2|S|$ long-range links land in a region of size less than $3|S|/2$. We calculate this probability as follows:

$$\begin{aligned}
\binom{n}{3|S|/2}\left(\frac{3|S|}{2n}\right)^{2|S|} &\leq \left(\frac{2en}{3|S|}\right)^{3|S|/2} \cdot \left(\frac{3|S|}{2n}\right)^{2|S|} \\
&= \left(\frac{3e^3|S|}{2n}\right)^{|S|/2} \\
&\leq \frac{1}{n^c},
\end{aligned}$$

for $2c\log n \leq |S| \leq n/(3e^3)$; so our claim is true with $c' = 3e^3$. (In the above inequality, we use the inequality $\binom{a}{b} \leq (ea/b)^b$, for integers $a > b > 0$.)

Consider the sequence of sets defined as follows. Let $L_0 = B_u$, and $L_i = L(L_{i-1})$ for $i > 0$. By the above argument, there exists $i = O(\log n)$ such that with probability at least $1 - O(\log n)/n^c$, L_i exceeds $n/(3e^3)$.

Since the destination of long-range links are drawn uniformly at random from the ring, the set of nodes that constitute L_i is drawn at random from the set of all n nodes. What is the probability that $L_i \cap B_v$ is empty? This is at most

$$\frac{\binom{n-c\log n}{n/(3e^3)}}{\binom{n}{n/(3e^3)}} \leq \left(1 - \frac{c \log n}{n}\right)^{n/(3e^3)} \leq \frac{1}{n^c},$$

for c chosen to be a sufficiently large constant.

Consider the myopic search algorithm for routing a message from a source s to destination t. Every intermediate node u forwards the message to its neighbor v (on the ring or via a long-range edge) such that v is the node nearest to t along the ring among all neighbors of v, breaking ties arbitrarily; formally, $(v - t) \bmod n$ is minimum among all neighbors v of u.

(b) Show that the myopic search algorithm always terminates.

Answer: In every step, the distance to the destination decreases by at least 1; hence the routing eventually reaches the destination. Therefore, the algorithm terminates.

(c) Show that there exist s and t such that the expected time for myopic search to complete is $\Omega(\sqrt{n})$.

Answer: The main point about this exercise is to observe that even though any two vertices are within $O(\log n)$ distance in the graph, with high probability, the maximum, over all s and t, of the expected time myopic search takes to go from s to t is $\Omega(\sqrt{n})$.

Let s and t be two diametrically opposite nodes on the ring. Consider the following process, starting from $s = v_0$, $i = 0$. If v_i is t, the process stops; otherwise, we select two random nodes x and y chosen uniformly at random from the ring, and set v_{i+1} to be the vertex that is closest to t along the ring among four vertices: $\{x, y\}$ and the two neighbors of v_i on the ring. If we move to one of x or y, we refer to this as a *long-range step*; otherwise, we refer to this as a *ring step*.

We would like to determine how long does the process take before it stops. We claim that with probability at least $1/2$, the process takes $\Omega(\sqrt{n})$ rounds. The selection of the random nodes x and y in each step is independent of all other selections. So the probability that in $\sqrt{n}$ rounds, none of the long-range steps hit the interval $[t - \sqrt{n}, t + \sqrt{n}]$ is at least

$$\left(1 - \frac{2\sqrt{n}}{n}\right)^{\sqrt{n}} = \left(1 - \frac{2}{\sqrt{n}}\right)^{\sqrt{n}} \leq \frac{1}{e^2}.$$

The total amount of progress made by the ring steps is at most $\sqrt{n}$. So with probability at least $1 - 1/e^2$, the myopic search procedure does not reach t in $\sqrt{n}$ steps. Therefore, the expected time for myopic search to complete is $\Omega(\sqrt{n})$.