

Solution to Final Exam, Problem 1

COS 340 - Spring 2012

I pledge my honor that I have not violated the honor code on this exam.

You are given a collection of tasks each of which must be assigned a time slot in the range $\{1, \dots, n\}$. Each task j has an associated interval $I_j = [s_j, t_j]$ and the time slot assigned to j must fall in the interval I_j . Each time slot can be assigned to at most one task. Show that it is possible to assign time slots to tasks satisfying these constraints if and only if the following condition holds:

For any integers x, y with $1 \leq x \leq y \leq n$, the number of intervals I_j contained in $[x, y]$ is at most $y - x + 1$.

Consider a bipartite graph $G(X + Y, E)$. For a set W of vertices in G , let $N(W)$ denote the set of all vertices adjacent to some vertex in W . Hall's marriage theorem states that there exists a matching that saturates X (ie covers all of the vertices in X) if and only if for every subset W of X , $|W| \leq |N(W)|$. We now formulate the above problem as a bipartite matching problem.

Let X have a unique vertex for each job I_j . Let Y consist of n unique vertices corresponding to the range of times $1, \dots, n$. For each vertex v in X (that is, for each job) and each vertex u in Y (that is, each time), draw an edge from v to u if the job corresponding to v can be assigned the time corresponding to u . That is, draw an edge between u and v if the time slot corresponding to u falls within the interval of time for the job corresponding to vertex v . It should be clear that if a matching exists that saturates X , then we have a time slot for each job I_j . Namely, we can assign each job I_j corresponding to a vertex v in X to the time slot corresponding to its partner in the matching. It should also be clear, to prove the other side, that if a satisfying assignment of jobs to time slots exists, then we can construct a matching that saturates X . Namely, we can add an edge from vertex $v \in X$, corresponding to job I_j , to vertex $u \in Y$ if the assignment of jobs to times happened to assign job I_j to the time slot corresponding to u .

Now, because we have a matching saturating X in this graph if and only if there exists a satisfying assignment of jobs to time slots, we want to show that there exists a matching saturating X if and only if the condition stated in the problem holds.

If the condition holds, we have a matching that saturates X :

We want to show that if the condition holds for all subsets of jobs, then Hall's theorem is satisfied for all subsets of vertices in X in the graph $G(X + Y, E)$ constructed above since this would imply that a matching that saturates X exists. Namely, we want to show that for any subset S of jobs (corresponding to a subset S' of vertices in X), that $|S'| \leq |N(S')|$. First, consider any arbitrary subset of jobs S . Because S consists of jobs potentially spanning disjoint intervals, we break S up into pieces such that each piece corresponds to a subset of jobs S_i whose intervals overlap (so the intervals corresponding to the S_i are disjoint from one another). Next, we define the intervals T_i to be the intervals $[x_i, y_i]$ spanned by the corresponding subsets S_i of S . Now we can relate these quantities to the graph we constructed earlier. Since we created a unique vertex in X for each job, it should be clear that $|S_i| = |S'_i|$ where S'_i denotes a subset of vertices in X corresponding to the jobs in S_i . Further, since we created a vertex in Y for each time slot, it should also be clear that $|T'_i| = y_i - x_i + 1$ where T'_i denotes a subset of vertices in Y corresponding to the time slots spanned by the interval T_i . Note that because we only added edges to the graph between jobs (vertices in X) and the time slots they span (vertices in Y), and since the S_i were constructed such that all of the jobs within them have overlapping time intervals, that $|N(S'_i)| = |T'_i|$. Further, by the condition we have that the number of jobs in any of the S_i must be $\leq y_i - x_i + 1$ where $[x_i, y_i]$ is the interval spanned by the jobs in S_i . This makes $|S'_i| \leq |T'_i| = y_i - x_i + 1 = |N(S'_i)|$ for all of the disjoint groups of jobs, which makes the following true of their union (since we defined the S_i and T_i to be disjoint sets): $|\bigcup_{S'_i \in S'} S'_i| = |S'| \leq |\bigcup_{T'_i \in T'} T'_i| = |N(S')| \rightarrow |S'| \leq |N(S')|$, the inequality we wanted. This implies that Hall's theorem holds for the graph constructed if the condition holds and that, therefore, a matching

that saturates X exists if the condition holds.

If we have a matching that saturates X , then the condition holds: This is true trivially since if we found a matching that saturates X , then, looking at any interval $[x, y]$, the number of jobs completely contained within that interval can be at most $y - x + 1$ since if this number were higher, we would not have been able to find a matching for every vertex (job) in X , since the number of jobs that must be assigned to this interval would exceed the number of time slots available.

Having proven that there exists a satisfying assignment if and only if a matching that saturates X exists, and having proven that a matching that saturates X exists if and only if the condition holds (using Hall's theorem), we conclude that there exists a satisfying assignment if and only if the condition holds.

Solution to Final Exam, Problem 2

COS 340 - Spring 2012

I pledge my honor that I have not violated the honor code on this exam.

Consider an experiment where we toss a coin n times. Let X be the number of times when the number of heads so far equals the number of tails so far. e.g. if we toss a coin 6 times and the outcomes are $HTTTHHH$, then $X = 2$ in this case.

1. Suppose that the coin is fair, i.e. for each toss, $P[\text{Heads}] = P[\text{Tails}] = 1/2$. Compute $E[X]$. Express it as $\Theta(f(n))$ for some function $f()$ in closed form.

Define I_i to be an indicator variable that is 1 if after i coin flips the number of heads equals the number of tails. Note that I_i cannot be one if i is odd since in this case, the number of heads cannot equal the number of tails. So assuming i is even, we have that $E[I_i] = P[\text{heads equals tails}] = \binom{i}{i/2} \cdot (\frac{1}{2})^i$. From problem four on homework three, we know that $E[I_{2i}]$ is equal to the probability that i heads come up in $2i$ coin flips, which we showed is $= \Theta(1/\sqrt{i})$. The expected value for X can now be obtained by summing over all of these indicators and using linearity of expectation. Note we use I_{2i} to avoid summing over odd numbers of flips, which yield zero anyway:

$$E[X] = E[\sum_{i=1}^{\lfloor n/2 \rfloor} I_{2i}] = \sum_{i=1}^{\lfloor n/2 \rfloor} E[I_{2i}] = \Theta(\sum_{i=1}^{\lfloor n/2 \rfloor} \frac{1}{\sqrt{i}}).$$

We use the following approximation from the notes (page 131):

$$\begin{aligned} \int_0^n \frac{1}{x+1} dx &\leq \sum_{i=1}^n \frac{1}{\sqrt{i}} \leq \int_0^n \frac{1}{\sqrt{x}} dx \\ &\rightarrow -2\sqrt{2} + 2\sqrt{1+n} \leq \sum_{i=1}^n \frac{1}{\sqrt{i}} \leq -2 + 2\sqrt{n} \\ &\rightarrow \sum_{i=1}^n \frac{1}{\sqrt{i}} = \Theta(\sqrt{n}). \end{aligned}$$

Applying this result to our summation, we have:

$$E[X] = \Theta(\sum_{i=1}^{\lfloor n/2 \rfloor} \frac{1}{\sqrt{i}}) = \Theta(\frac{1}{\sqrt{n/2}}) = \Theta(\sqrt{n}).$$

And thus $f(n) = \sqrt{n}$.

2. Suppose the coin is biased such that $P[\text{Heads}] = 1/3$ and $P[\text{Tails}] = 2/3$. X is the same random variable as before. Now compute $E[X]$. Express it as $\Theta(g(n))$ for some function $g()$ in closed form.

Defining I_{2i} exactly as before, and ignoring odd numbers of flips, we now have $E[I_{2i}] = \binom{2i}{i} \cdot (\frac{1}{3})^i \cdot (\frac{2}{3})^i = \binom{2i}{i} \cdot (\frac{2^i}{3^{2i}}) = \binom{2i}{i} \cdot (\frac{2}{9})^i$. We now simplify this quantity using Stirling's approximation:

$$E[I_{2i}] = \binom{2i}{i} \cdot (\frac{2}{9})^i = \Theta(\frac{\sqrt{2i}(2i)^{2i} e^i e^i}{e^{2i} \sqrt{i} \sqrt{i} i^i}) \cdot (\frac{2}{9})^i = \Theta(\frac{(2)^{2i}}{\sqrt{i}} \cdot (\frac{2}{9})^i) = \Theta(\frac{1}{\sqrt{i}} \cdot (\frac{8}{9})^i).$$

Using this, we can now compute $E[X]$ by summing over the I_i and using linearity of expectation as before:

$$E[X] = E[\sum_{i=1}^{\lfloor n/2 \rfloor} I_{2i}] = \sum_{i=1}^{\lfloor n/2 \rfloor} E[I_{2i}] = \Theta(\sum_{i=1}^{\lfloor n/2 \rfloor} \frac{1}{\sqrt{i}} \cdot (\frac{8}{9})^i)$$

Now we show that this sum converges to a constant factor for large n . First, since $\sum_{i=1}^{\lfloor n/2 \rfloor} \frac{1}{\sqrt{i}} \cdot (\frac{8}{9})^i \leq \sum_{i=1}^{\lfloor n/2 \rfloor} (\frac{8}{9})^i$, and the latter converges to 8 (using the formula for a geometric sum), we have $O(1)$ as the upper bound. Further, since all the terms in the sum are ≥ 0 , we have that $\Omega(g(n)) = \Omega(\text{the first term (for large } n)) = \Omega(8/9) \rightarrow \Omega(1)$ as a lower bound for n large (since the quantity only grows larger as we add more terms).

Thus, having shown that the lower and upper bound are constant, we conclude that $E[X] = \Theta(1)$.

ID: 03
Name: Al-Naji, Nader

Collaborator #1: Last Name, First Name
Collaborator #2: Last Name, First Name

Solution to Final Exam, Problem 3

COS 340 - Spring 2012

I pledge my honor that I have not violated the honor code on this exam.

Matching: Given a graph $G(V, E)$, determine whether the graph has a perfect matching i.e. a subset $M \subseteq E$ such that every vertex in V is incident on exactly one edge in M .

Network Design: Given a graph $G(V, E)$, and non negative integers $d_i < |V|$ for all $i \in V$, determine whether there exists a subgraph $G'(V, E')$ ($E' \subseteq E$) such that for all $i \in V$, the degree of i in the subgraph G' is exactly d_i .

Matching is a well known problem that can be solved in polynomial time. Show that Network Design can be solved in polynomial time by a reduction to Matching.

Chose not to do this problem.

Solution to Final Exam, Problem 4

COS 340 - Spring 2012

I pledge my honor that I have not violated the honor code on this exam.

Let a, b, p , and k be integers with p prime. Prove the following:

1. p divides $\binom{p}{k}$ where $1 \leq k \leq p-1$:

The definition of divisibility states that p divides b if $p = k \cdot b$ for some integer k . The choose function, here $\binom{p}{k}$, has a combinatorial interpretation as the number of ways to choose k things from a collection of p things. This implies that it is always an integral value for integral values of the inputs k and p . So we can know that $\binom{p}{k} = \frac{p!}{k!(p-k)!} = p \cdot \frac{(p-1)!}{k!(p-k)!}$ is integral and that, since all we did was pull out a factor of p from the factorial, that $\frac{(p-1)!}{k!(p-k)!}$ is also integral. Thus, we can write $\binom{p}{k}$ as $\binom{p}{k} = p \cdot C$ where $C = \frac{(p-1)!}{k!(p-k)!}$ is integral and thus, by the definition of divisibility, p divides $\binom{p}{k}$.

2. $(ap+1)^p = 1 \pmod{p^2}$.

First, we show that $(ap+1)^k = a \cdot k \cdot p + 1 \pmod{p^2}$ using induction on k .

Let $Q(k) = "(ap+1)^k = a \cdot k \cdot p + 1 \pmod{p^2}"$.

Base: $k=0, k=1$:

If $k=0$, then $(ap+1)^k = 0+1=1$. If $k=1$ then $(ap+1)^k = a \cdot p \cdot 1 + 1$ and the statement holds.

Assume true for k , prove true for $k+1$:

Assuming $(ap+1)^k = a \cdot k \cdot p + 1$, then $(ap+1)^{k+1} = (a \cdot k \cdot p + 1)(a \cdot p + 1) = a^2 \cdot k \cdot p^2 + a \cdot k \cdot p + a \cdot p + 1 = 0 + (k+1)a \cdot p + 1 \pmod{p^2}$ and the statement holds.

We now use this fact to solve the problem: $(ap+1)^p = ap^2 + 1 = 0 + 1 = 1 \pmod{p^2}$.

3. $b^{p(p-1)} = 1 \pmod{p^2}$ where p does not divide b .

$b^{p(p-1)} = (b^{p-1})^p$. Further, by Fermat's Little Theorem, we know that $b^{p-1} = 1 \pmod{p}$ since b is not a multiple of p . This implies that b^{p-1} is equal to a multiple of p plus one and so we can write b^{p-1} as $ap+1$ for some integer a . Now, plugging this into our original formulation and using (2) we have: $b^{p(p-1)} = (b^{p-1})^p = (ap+1)^p = 1 \pmod{p^2}$.

Solution to Final Exam, Problem 5

COS 340 - Spring 2012

I pledge my honor that I have not violated the honor code on this exam.

Consider an arbitrary drawing (with possibly intersecting arcs) of the complete graph on K_n on the plane ($n \geq 5$). Prove that at least $\binom{n}{5}/(n-4)$ pairs of edges have to cross.

Lemma 1: number of crossing edges is greater than or equal to $|E|^3/(64n^2)$ for an arbitrary graph G .

We prove a bound for an arbitrary graph. Denote the number of crossings (same as the number of pairs of intersecting edges) in graph G by $cr(G)$ and consider removing edges until the graph is planar. After this process we should have a graph with a total of $|E| - cr(G)$ edges and n vertices with no intersecting edges. Now, because the graph is planar, we can use Euler's formula to put a lower bound on the number of crossings. Recall that a planar graph must have at least three boundaries per face and that the number of boundaries is equal to $2|E|$:

$$\begin{aligned} 2|E| &\geq 3f \rightarrow 2|E| \geq 3(2 - v + |E|) \text{ by Euler's formula for planar graphs} \\ \rightarrow |E| &\leq 3|V| - 6 \rightarrow |E| \leq 3|V| \text{ for } n \geq 3 \\ \rightarrow |E| - cr(G) &\leq 3|V| \rightarrow cr(G) \geq |E| - 3|V|. \end{aligned}$$

We now make a probabilistic argument to tighten the bound. We construct a random subgraph S of G by putting each vertex in G into S independently with probability p , and connecting the vertices in S with edges if the vertices in the original graph are both in S and were connected by an edge in the original graph. We let e_S denote the number of edges in S , n_S denote the number of vertices in S , and cr_S denote the number of crossings in S . Note that we haven't yet chosen p . This choice will be made later and will affect the tightness of our bound. Recall the above proof that shows $cr_S \geq e_S - 3 \cdot n_S$. We take the expected value of both sides of this expression to provide concrete values and get:

$$E[cr_S \geq e_S - 3 \cdot n_S] \rightarrow E[cr_S] \geq E[e_S] - 3 \cdot E[n_S].$$

Recall that each vertex had probability p of being in S so $E[n_S] = p \cdot n$. The expected number of edges is then $E[e_S] = p^2|E|$ since each edge is in the graph only if both of its adjacent vertices are in the graph. Finally, every intersection has probability of p^4 of being in S since we need the four vertices from the crossing to be in S in order for the intersection to be in S . Thus, $E[cr_S] = p^4 cr(G)$. Finally, plugging these values into the inequality derived yields:

$$p^4 cr(G) \geq p^2|E| - 3p \cdot n \rightarrow cr(G) \geq |E|/p^2 - 3n/p^3.$$

Now, choosing p to be $4n/e$, and assuming the number of edges is at least four times the number of vertices (to ensure $p \leq 1$), we arrive at the final inequality:

$$cr(G) = \text{number of intersecting edges} \geq |E|^3/(64n^2)$$

Note: This lemma was written using an article on Wikipedia about crossing edges as a guide. This was approved by the instructors.

Now we have a minimum number of intersecting edges for an arbitrary graph. We show that this lower bound on the number of edges is stronger than the lower bound above and, therefore, that the lower bound above is satisfied:

For a clique K_n , we have: $|E| = n(n-1)/2$ and thus:
 $cr(G) \geq |E|^3/(64n^2) = n^3(n-1)^3/(64n^2) = n(n-1)^3/64.$

The problem wants us to show that $cr(G) \geq \binom{n}{5}/(n-4) = (n)(n-1)(n-2)(n-3)/120$. We now compare our established bound to this bound. Since each term in the numerator of our bound is larger than each term in the numerator of the problem's bound, and since the denominator of our bound is smaller than the denominator of the problem's bound, we have: $cr(G) \geq n(n-1)^3/64 \geq (n)(n-1)(n-2)(n-3)/120 = \binom{n}{5}/(n-4)$ for $n \geq 5$, thus establishing that there can be no fewer than $\binom{n}{5}/(n-4)$ intersecting edges in a clique K_n .

Solution to Final Exam, Problem 6

COS 340 - Spring 2012

I pledge my honor that I have not violated the honor code on this exam.

A person starts walking home from a bar. The distance is n blocks. Every hour he either manages to walk one block towards home, or gets confused and picks up a cab to the bar (each with probability $1/2$). At any point, if he is at the bar, he always moves ahead one block. What is the expected time it takes him to get home?

Label each consecutive block with a number between zero and n , with the bar being block 0. Let T_i be defined as the amount of time it takes to get from block 0 to block i . We define T_0 to be zero since it takes no time for the man to get to where he already is. Working from this framework, we get the following:

$$T_n = T_{n-1} + 1/2(1 + T_n) + 1/2(1) = 2(T_{n-1} + 1)$$

Let's unpack this equality. What this is basically saying is that the time it takes to get to block n consists of three components. This makes sense since in order to get to block n , the man must first get to block $n - 1$. Then, once arriving at block $n - 1$, he goes back to block 0 with probability $1/2$, and goes to block n with probability $1/2$. If he goes back to block 0, the expected time for him to reach block n becomes the time it takes to get from block 0 to block n plus the cost of the intermediate step. If he goes home, the cost is simply the cost of walking one block. These times are then weighted by the probability that they occur. Of course, this formulation implicitly uses linearity of expectation. Now, working backwards from n we begin to notice a pattern:

$$\begin{aligned} T_n &= 2(T_{n-1} + 1) = 2T_{n-1} + 2 \\ &= 4T_{n-2} + 4 + 2 \\ &= 8T_{n-3} + 8 + 4 + 2 \\ &= 16T_{n-4} + 16 + 8 + 4 + 2 \\ &\dots \end{aligned}$$

So it looks like the closed form is $T_n = 2^{n-1} + \sum_{i=1}^{n-1} 2^i = 2^n + 2^{n-1} - 2$ for $n > 0$ and zero for $n = 0$. We prove this formally using induction on n .

Let $Q(n)$ = "The time it takes to get from the bar to block n is $2^n + 2^{n-1} - 2$ steps for $n > 0$."

Base: $n = 1, n = 2$

It takes one step to get from state zero to state 1 since when the man is at the bar, he automatically walks forward a step. $2^1 + 2^{1-1} - 2 = 1$. $T_2 = 2(T_1 + 1) = 2(2) = 4$ by an argument similar to the one that claimed $T_n = 2(T_{n-1} + 1)$. $2^2 + 2^{2-1} - 2 = 4$. Thus, the statement holds for the base case.

Step: Assume true for n , prove true for $n + 1$.

We assume that it takes $2^n + 2^{n-1} - 2$ steps to get to block n by the inductive hypothesis. The time it takes to get to block $n + 1$ is, as before: $T_{n+1} = 2(T_n + 1) = 2(2^n + 2^{n-1} - 2 + 1) = 2^{n+1} + 2^n + 2 = Q(n + 1)$, which is consistent with the statement.

Thus, having shown that the statement holds for the base case and for $n + 1$ assuming it holds for n , we conclude that the expected time for the man to get home from the bar if his home is n blocks away is $2^n + 2^{n-1} - 2$ if $n > 0$ and 0 if $n = 0$ (though it's probably not a good idea for him to live at the bar given his alcoholic tendencies).

Solution to Final Exam, Bonus

COS 340 - Spring 2012

I pledge my honor that I have not violated the honor code on this exam.

There are a movies in the Netflix library and b registered Netflix users, where $b \geq 3$ is an odd integer. Each user rates each movie as either "like" or "dislike". Suppose k is a number such that, for any two users, their ratings coincide for at most k movies. Prove that $k/a \geq (b-1)/(2b)$.

Consider the above construction as a matrix M with b rows and a columns such that $M_{i,j} = 1$ if user i liked movie j and $M_{i,j} = 0$ if user i disliked movie j . Define a conflict between two users to be when one user has rated a particular movie the same as another user. Formally, let a conflict between user i and user j for a movie k occur when $M_{i,k} = M_{j,k}$. Now consider the minimum number of conflicts in any given column of M . In any particular column, there are at least $(b-1)/2 + 1$ cells that have the same value. This is by the pigeon-hole principle since we are filling each cell with one of two values and there are b cells in a column. In fact, we can break each column up into k likes and $b-k$ dislikes and count the number of conflicts in this fashion. Dividing by two to eliminate double-counting, we obtain that the number of conflicts in any given column is: $(k)(k-1)/2 + (b-k)(b-k-1)/2 = 1/2(b^2 - b - 2bk + 2k^2) \geq 1/4(b-1)^2$ for $b \geq 3$. Observe that the functions of k here achieve their minimum at $k = (b-1)/2$ (the function actually equals $1/4(b-1)^2$ at this minimum) and note that it is ok to take something less than the actual value since we are seeking a lower bound on the minimum number of conflicts. We conclude that the minimum number of conflicts in any given column of M must be greater than or equal to $(b-1)^2/4$. This makes the total number of conflicts overall simply $a \cdot (b-1)^2/4$ since the columns cannot have conflicts with each other. We now look at how this impacts k . The total number of pairs of users is $\binom{b}{2} = (b)(b-1)/2$. The maximum number of per-user conflicts k times this value should yield an upper bound on the total number of conflicts in the matrix. Namely, it should be the case that $k \cdot (\text{number of user pairs}) \geq (\text{total number of conflicts})$ since k is an upper bound on the number of conflicts between two users (ie two rows). Having established this, we now plug in the precomputed lower bound on the total number of conflicts and attain the following:

$$k \cdot (b)(b-1)/2 \geq a \cdot (b-1)^2/4 \rightarrow k \geq a \cdot (b-1)/(b \cdot 2) \rightarrow k/a \geq (b-1)/(b \cdot 2).$$

Hence, the inequality holds.