1. (10 points) **(How good is greedy for Vertex Cover)** This will drive down the reason we study other algorithms for set cover even though in general we know that greedy is optimal. There could be a large family of instances which have structure where we can outperform greedy.

   (a) (10 points) Construct an example where the greedy algorithm has an approximation ratio of $\Omega(\log n)$ for the vertex cover problem where there are $n$ vertices in the graph.

**Solution:**

*Proof.* We will try to construct a bipartite graph, divided into two disjoint vertex sets A and B such that all edges exist between set A and B. And no edge exists among vertices in A. And no two edge exists among vertices in B.
More specifically we design the bipartite graph in such a way that $|A|$ = n and $|B| = n * log(n)/2$.
We consider a graph s.t A is divided into logn groups st size of the ith group is $n/2^i$ i.e let X = partition of A into logn groups st $X_1 = \{A_1, A_2, \ldots, A_{(n/2)}\}$ and $X_2 = \{A_{n/2+1}, A_{n/2+2}, \ldots, A_{n/2+n/4}\}$ and B is divided into logn groups each of size n/2. i.e let Y = partition of B into logn groups st $\forall i, Y_i = \{B_{(i-1)*(n/2)+1}, B_2, \ldots, B_{i*n/2}\}$.
The groups divided are such that for $\forall B_j \in Y_i, degree(B_j) = n/(2^{(i-1)})$. And degrees of A are s.t, $\forall v \in A, degree(v) = n$.
This is achieved by $\forall j = 1$ to $logn$ forming a complete bipartite graph between $\{X_j \cup X_{(j+1)} \ldots\}$ and $Y_j$.
We observe that initially the greedy algorithm picks a max degree of n. Vertices of A and $Y_1$ have degree n. Let it pick $Y_1$ removing all edges connected to $Y_1$, since it was connected to $X_1, X_2, \ldots$ hence all elements in A, since $|Y_1| = n/2$, now degree of each element in A is n/2 and the degree if $Y_2$ is n/2 hence we picked a total of n/2 elements,similarly if we continue, we will end up picking every Y group and hence picking set B and as it is a bipartite graph, this is a valid set cover but $|B| = n * logn/2$. A is the minimum set cover and $|B|/|A| = \Omega(logn)$. Hence a logn approximation.

□

2. (25 points) **(Finishing the Set Cover Rounding Proof)** We'd left the final parts of the proof as homework. You'll now complete this.

   (a) (10 points) We showed the following two properties which our rounding algorithm satisfies (if we repeated the randomized rounding experiment for $T = 2\ln n$ steps: (i) the expected cost is $2\ln n\mathsf{Opt}$ where $\mathsf{Opt}$ is the cost of the optimal LP fractional solution, and (ii) the probability with which all elements are covered is at least $1 - \frac{1}{n}$. Show that there with some constant probability, we will find a solution which has cost at most $O(1)\ln n$ and also covers all the elements. (Hint: Use Markov's inequality and the union bound)

**Solution:**

*Proof.* Let the algorithm procede in T rounds and X be the random variable denoting the number of sets picked. From class we know that,

$$Pr(\text{Element is not chosen in T rounds}) = e^{-T} \tag{1}$$
$$Pr(\exists \text{ an element is not chosen in T rounds}) = \frac{N}{e^T} \tag{2}$$
$$E(X) = T * \mathsf{Opt} \tag{3}$$
$$Pr(X \geq k * \mathsf{Opt}) \leq \frac{E(X)}{k\mathsf{Opt}} \tag{4}$$
$$Pr(X \geq k * \mathsf{Opt}) \leq \frac{T}{k} \tag{5}$$

Taking the union of event in 2 and 5 is the probability of the failure of the algorithm, and applying the union bound inequality on this we get,

$$\begin{aligned}
Pr(2 \cup 5) &\leq Pr(2) + Pr(5) \\
Pr(failure) &\leq Ne^{-T} + \frac{T}{k} \\
\text{Set } T &= 2 * \ln N + 1 \\
Pr(failure) &\leq \frac{1}{e * N} + \frac{2\ln N + 1}{k} \\
\text{Set } k &= 8 * \ln N \\
\forall N > 2, \ln N > 1 \;\&\; \frac{1}{e * N} &< \frac{1}{e} \\
Pr(failure) &< \frac{1}{e} + \frac{2\ln N + \ln N}{8\ln N} \\
Pr(failure) &< \frac{1}{e} + \frac{3}{8} \\
Pr(failure) &< 0.75
\end{aligned}$$

We have shown that with probabilty of success greater than 0.25 we have that the algorithm runs with success where number of sets picked $< 8\ln N\mathsf{Opt}$. $\square$

(b) (10 points) Now if instead of running our rounding $T = 2\ln n$ times, if we had run it a different number (say, $\ln n + C\ln\ln n$) of times. Then try to optimize the parameters and show that we will compute, with some non-trivial probability of $\Omega(\frac{1}{\ln n})$, a solution where the cost is $(\ln n + O(\ln\ln n))\,\mathsf{Opt}$ and all elements are covered.

**Solution:**

*Proof.* From above we conclude that,

$$
\begin{aligned}
Pr(success) &\geq 1 - \frac{n}{e^{-T}} - \frac{T}{k} \\
\text{Set } T &= \ln n + C \ln\ln n \\
\text{Set } k &= \ln n + a \ln\ln n \\
Pr(success) &\geq 1 - \frac{1}{\ln n^{C}} - \frac{\ln n + C\ln\ln n}{\ln n + a \ln\ln n} \\
Pr(success) &\geq -\frac{1}{\ln n^{C}} + \frac{(a-C)\ln\ln n}{\ln n + a\ln\ln n} \\
\text{Using, } \ln x &\geq 1 - \frac{1}{x} \\
\text{Using, } \ln\ln x &\geq 1 - \frac{1}{\ln x} \\
Pr(success) &\geq -\frac{1}{\ln n^{C}} + \frac{(a-C)\ln\ln n}{\ln n + a\ln\ln n}, \text{Assume } a > C \\
Pr(success) &\geq -\frac{1}{\ln n^{C}} + \frac{(a-C)*(1-\frac{1}{\ln x})}{\ln n + a\ln\ln n} \\
Simplifying&, \\
Pr(success) &\geq -\frac{1}{\ln n^{C}} + \frac{(a-C)*\ln(n/e)}{(\ln n)(\ln n + a\ln\ln n)} \\
\text{Using, } \ln x &\leq x \\
\text{Using, } \ln\ln x &\leq \ln x \\
Pr(success) &\geq -\frac{1}{\ln n^{C}} + \frac{(a-C)*\ln(n/e)}{(\ln n)(\ln n + a\ln n)} \\
Pr(success) &\geq -\frac{1}{\ln n^{C}} + \frac{(a-C)*\ln(n/e)}{(\ln n)(\ln n)(1+a)} \\
Pr(success) &\geq -\frac{1}{\ln n^{C}} + \frac{\alpha}{\ln n} \\
Pr(success) &= \Omega(\frac{1}{\ln n})
\end{aligned}
$$

□

(c) (5 points) Finally boost the success probability above by repeating this algorithm some number of times. Roughly how many times do you need to run to get probability of failure to be $e^{-n}$?

**Solution:**

*Proof.* From above, considering 1 round,

$$Pr(success) = \frac{\alpha}{\ln n}$$
$$Pr(failure) = 1 - \frac{\alpha}{\ln n}$$
For R rounds,
$$Pr(failure) = (1 - \frac{\alpha}{\ln n})^R$$
$$\text{Using for large L, } (1 - \frac{1}{L})^L = \frac{1}{e}$$
$$Pr(failure) = e^{-\frac{Rk}{\ln n}}$$

Which implies that R must be $\frac{n \ln n}{k}$ i.e of the order of $n \ln n$. ☐

3. (20 points) **(Integrality Gap for Robust Min-Sum-Set-Cover)** Consider the generalization of min-sum-set-cover where the cover time of an element is defined to be the first time when the element is covered $K$ times, for a given parameter $K$. We will now show that the natural LP has a large integrality gap for this instance.

   (a) (10 points) Write the natural LP for this problem.

**Solution:**

*Proof.* Defining the problem:

$$U = \{e_1, e_2, \ldots, e_n\}$$
$$S = \{s_1, s_2, \ldots, s_m \mid \forall i \in \{1, 2, \ldots, m\}, s_i \subset U\}$$

Goal : Find an ordering s.t.
$\sum_{i=1}^{n}(covertime(e_i))$ is minimum, where cover time for an element e is defined as the first time (K+1) sets have covered that element.
$covertime(e) = min_i(\sum_{j=1}^{i} |S_{\sigma(j)} \cap \{e\}| = K)$.
The IP formulation is as follows:
Let $X_{s,t} = 1|0$, if set s is covered at time t.
Let $Y_{e,t} = 1|0$, is element e is covered K times before time t.

Constraints:

$$\forall t \in \{1, 2, \ldots, m\}, \sum_{i=1}^{m} X_{s_i,t} = 1$$
$$\forall i \in \{1, 2, \ldots, m\}, \sum_{t=1}^{m} X_{s_i,t} = 1$$
$$\forall t \in \{1, 2, \ldots, m\} \& \forall i \in \{1, 2, \ldots, n\}, \sum_{s,e \in s} \sum_{t'=1}^{t-1} X_{s,t'} \geq Y_{e,t} * K.$$

Objective Function:

$Min \sum_e \sum_t (1 - Y_{e,t})$
For the LP relaxation all the X's and Y's are no longer 0|1, rather they lie between 0 and 1.

$$0 \leq X_{s,t} \leq 1, \forall s, t$$
$$0 \leq Y_{e,t} \leq 1, \forall e, t$$

□

(b) (10 points) Consider the following instance, and show that it has a large integrality gap. The universe of elements $U = \{e_1, e_2, \ldots, e_l\}$. The sets are $\mathcal{S} = \{S_1 \equiv \{e_1, e_2, \ldots, e_l\}, S_2 \equiv \{e_1, e_2, \ldots, e_l\}, \ldots, S_n \equiv \{e_1, e_2, \ldots, e_l\}, S_{n+1} = \{e_1\}, S_{n+2} = \{e_2\}, \ldots, S_{n+l} = \{e_l\}\}$. Suppose the coverage requirement $K = (n + 1)$. Show that we can set values of $l$ and $n$ so that the LP solution and integral solutions have a large gap. For this, you need to exhibit some fractional solution of low cost and show that all integral solutions have much larger cost.

**Solution:**

*Proof.* We first show that the minimum of IP is $n * l + l * (l + 1)/2$. Since the cover time is the first time a set covers (n+1) elements, minimum cover time of an element is (n+1), wlog let the element be $e_1$. Since there are only n+1 sets that have $e_1$ in them. These n+1 sets must be picked first, now the remaining $l-1$ sets should be picked. We observe that each new set now produces another element that has a cover time of 1 greater than the previous. Hence summation of cover time $= \sum_{i=1}^{l} n + i = n * l + l * (l + 1)/2$.
To show a large integrality gap, we pick the sets in the same order that we picked them in the integer program that is let $X_{S_i,t} = 1$ if $i = t$ and $X_{S_i,t} = 0$ if $i \neq t$, i.e X is an identity matrix of size $(N + l) * (N + l)$. Now, we observe

what the corresponding Y values are. Since,

$$\forall t \in \{1,2,\ldots,m\} \& \forall i \in \{1,2,\ldots,n\}, \sum_{s,e\in s}\sum_{t'=1}^{t-1} X_{s,t'} \geq Y_{e,t} * K$$

and we have to minimise (1-Y), we should maximise Y. And thus set Y's such that there is equality in our constraint. For our problem,

$$\forall t \in \{1,2,\ldots,(n+l)\} \& \forall i \in \{1,2,\ldots,(n+l)\}, \sum_{s,e\in s}\sum_{t'=1}^{t-1} X_{s,t'}/(n+1) = Y_{e,t}$$

We notice that if $t <= n+1, Y_{e_i,t} = (t-1)/(n+1)$ if $i <= l$ otherwise the Y value is zero, since it is the number of sets covering e in less than time t divided by (n+1), and each set covers e at any time less than equal to n.
Secondly we observe that after $t > n+1, Y_{e_i,t} = 1$ if $i \leq (t-n-1)$ and $Y_{e_i,t} = n/(n+1)$ otherwise. Hence,

$$\sum_e \sum_t (1 - Y_{e,t}) = n*(n+l) - \sum_e \sum_t Y_{e,t}$$
$$\sum_e \sum_t Y_{e,t} = \sum_e \sum_{t=1}^{n+1} Y_{e,t} + \sum_e \sum_{t>n+1} Y_{e,t}$$

The first term exists only for the first l elements, implies,

$$\sum_e \sum_{t=1}^{n+1} Y_{e,t} = l * \sum_{t=1}^{n+1}(t-1)/(n+1) = n*l/2$$

The second term we sum across all e's first. We observe that in the $l$ elements, $(t-n-1)$ of them are 1 and the rest are $n/n+1$. Thus summation across all $t's$ is $(t-n-1)+(l-t+n+1)*\frac{n}{n+1}$. Simplifying we get,

$$Second\_term = \sum_{t>n+1}(t-n-1)+(l-t+n+1)*\frac{n}{n+1}$$
$$Second\_term = \sum_{t>n+1}(l*\frac{n}{n+1}-1)+(t/n+1)$$
$$Second\_term = (l-1)(\frac{nl}{n+1}-1)+\frac{n(l-1)+l*\frac{l+1}{2}-1}{n+1}$$

After some tedious calculations and simplifications we find the ratio of IP:LP at the stage where ,

$$ratio = \frac{2*n*l + l*(l+1)}{n*l + \frac{l^2+l+2*l*n}{n+1}}$$
$$ratio = \frac{2n^2 + nl + 3n + l + 1}{n^2 + 3n + l + 1}$$
$$ratio = \frac{l*(n+1) + 2n^2 + 3n + 1}{l*(1) + n^2 + 3n + 1}$$

Imposing $l >> n$, we get that the algorithm is $\Omega(n)$ approximation under the limit. $\square$

4. (30 points) **(Structure of a fractional optimum for the vertex cover LP relaxation)** Recall in class that we wrote down an integer linear program of two variable inequalities (one per edge) such that a feasible 0-1 solution is a vertex cover. Let VC denote this integer linear program, and let LPVC denote the vertex relaxation. Let $x^*$ an optimum solution to LPVC and let $V_0, V_1, V_h$ be the 3 vertex sets of the graph as discussed in class.

   (a) (5 points) Show that $N(V_0) = V_1$.

**Solution:**

*Proof.*

$$Let(N(V_0)) = X$$
$$X = (X \cap V_0) \cup (X \cap V_1) \cup (X \cap V_h)$$
$$Assume, X \cap V_0 \neq \phi$$
$$\Rightarrow \exists v \in V_0 \;\&\; v \in N(V_0) \Rightarrow \exists u \in V_0 \mid (u, v) \in E$$
$$\text{From LP, } x_v + x_u \geq 1$$
$$x_v < 1/2 \;\&\; x_u < 1/2$$
$$\Rightarrow\Leftarrow$$

Similarly, for $X_h$

$$\begin{aligned}
&Assume, X \cap V_h \neq \phi \\
&\Rightarrow \exists v \in V_h \;\&\; v \in N(V_0) \Rightarrow \exists u \in V_0 \mid (u,v) \in E \\
&\text{From LP, } x_v + x_u \geq 1 \\
&x_v = 1/2 \;\&\; x_u < 1/2 \\
&\Rightarrow\Leftarrow \\
&\Rightarrow X = X \cap V_1 \\
&\Rightarrow X \subset V_1 \\
&\text{To prove: } X = V_1 \\
&\text{To prove: } V_1 \subset X \\
&\Rightarrow \exists v \in V_1
\end{aligned}$$

Since v is in $V_1$, we will prove that it must be in $N(V_0)$, and hence $V_1 \subset X$. If v was not in $N(V_0)$. We could decrease the value of v from the LP optimal we have got by some $\epsilon$, such that value of v still stays above half and hence all edge constraints with vertices in $V_h$ are still satisfied. Since we cannot do so implies a contradiction hence X = $V_1$. □

(b) (10 points) Show that the value of $x^*$ is $|V_1| + \frac{|V_h|}{2}$.

**Solution:**

*Proof.* Let us construct $V_1'$ where it is the set of vertices in $V_1$ that do not have LP optimal value of 1, similarly, construct $V_0'$ where it is the set of vertices in $V_0$ that do not have LP optimal value of 0.
The sum of the LP optimal value of the rest of the vertices are $\frac{|V_h|}{2} + |V_1| - |V_1'|$.
We will prove that giving the set $V_1'$ all 1's and the set $V_0'$ all zeroes is indeed optimal. Since $x^*$ is an optimal solution we construct $x_1, x_2$ such that $x^* = \frac{x_1+x_2}{2}$.
Construct the set $x_1$ by adding $\epsilon$ to each element in $V_0'$ and subtracting $\epsilon$ to each element in $V_1'$ such that all elements in $V_1'$ are still greater than half. Since $N(V_0) = V_1$, all constraints are still satisfied.
Similarly,
Construct the set $x_1$ by adding $\epsilon$ to each element in $V_1'$ and subtracting $\epsilon$ to each element in $V_0'$ such that all elements in $V_0'$ are still greater than 0. Since $N(V_0) = V_1$, all constraints are still satisfied.
As we have shown the possible valid construction of these two sets $x^* = \frac{x_1+x_2}{2}$.

Now we compare the objective function value of them.

$$
\begin{aligned}
Val(x^*) &= Val(\frac{x_1 + x_2}{2}) \\
Val(x^*) &= \frac{Val(x_1) + Val(x_2)}{2} \\
\text{Using optimality,}& \\
Val(x^*) &\leq Val(x_1) \\
Val(x^*) &\leq Val(x_2) \\
\Rightarrow Val(x^*) &= Val(x_1) = Val(x_2)
\end{aligned}
$$

Hence we have constructed a set of optimal values $x_2$, where the elements of $V_1'$ are more closer to 1 and the elements of $V_0'$ are more closer to zero. Performing this operation on the limit, we observe that all elements of $V_0'$ are finally zero. Hence the objective function value from these sets are $|V_1'|$, adding it to the rest of the values we get.

$$
\mathsf{Opt} = |V_1| + \frac{|V_h|}{2}
$$

□

(c) (5 points) Show that all the corner points of the polytope are half-integral.

**Solution:**

*Proof.* We will prove the contrapositive statement, i.e if points are not half-integral then it is not a corner point of a polytope. The proof constructed in the previous theorem holds here as well, $V_0'$ and $V_1'$ are non empty as there are points which are less than half and greater than half and not integers. Hence here also we can construct two sets $x_1$ and $x_2$, such that $x = \frac{x_1+x_2}{2}$.
We construct these sets similarly as above by subtracting and adding a very small value $\epsilon$. As $V_0'$ and $V_1'$ are non empty, implies that $x \neq x_1 \neq x_2$. Hence we have found two points which are feasible and hence inside the polytope and the non integer point is can be represented as a ratio of those two. Hence x is not a corner point. Feasibility of $x_1$ and $x_2$ is guaranteed from our constructions above. □

(d) (10 points) Use the above arguments to compute the minimum vertex cover of a tree.

**Solution:**

*Proof.* We consider a LP optimal solution and construct similar sets $V_0, V_1, V_h$

for the tree as mentioned in the problem statement. We have shown that setting all vertices in $V_1$ to 1 and all vertices in $V_0$ to 0 and all vertices in $V_h$ to $1/2$ is an LP optimal solution. We now consider this solution and expand on it. We use here the property that a subgraph of a tree always has a vertex of degree atmost 1 in it. It is a graph theory theorem that if a graph is acyclic hence it contains a vertex with degree atmost 1. Since a tree is acyclic any subgraph of a tree is acyclic. Hence any subgraph of a tree has a vertex with degree atmost 1 in that subgraph.

We consider the subgraph formed by vertices in $V_h$, let v be the degree atmost 1 vertex in the graph. v cannot have degree 0 as it belongs to $V_h$, if it doesn't have neighbors in $V_h$ it must have neighbors in $V_1$, it can't have neighbors in $V_0$ as it would violate the constraints. Hence all its neighbors are in $V_1$.Now, we can set value of v to be 0 and still satisfy all constraints, but this violates the optimality of the LP hence no such vertex with degree 0 exists. Implies that there is a vertex with degree 1 in our subgraph.

Let u be the neighbor of v.Since both u and v are in $V_h$ there value is $1/2$. Now if we decrease the value of v by half and increase the value of u by half. We have found a new solution that does not violate any constraints as the edge (u,v) is preserved and all other edges of v are in $V_1$. Also since we have just increased the value of u any of its edges won't be violated. Hence we have removed two vertices from $V_h$ and made them have integral values. We can continue applying this idea till there are no vertices left in $V_h$ and all the vertices are in $V_0$ and $V_1$. Hence we have proved that the LP optimal is the answer for the minimum vertex cover for a tree and that we can construct such a set by creating $V_0, V_1, V_h$ and then emptying $V_h$ to get an integer solution. □