CS203 (2023) – Second assignment

Total marks: 25

- **Note.** Answers without clear and concise explanations will not be taken into account. Use of immoral means will get severe punishment.

Name: Geetika

Roll No: 210392

Questions

1. **(5+3+7 marks)** In an undirected graph $G = (V, E)$, a cut is specified by an $S \subseteq V$, where the size of the cut is the number of edges between S and $\bar{S}$.
   A random cut is obtained by keeping each vertex of V in S with probability half.

   (a) Show that a random cut will have expected size $|E|/2$. Justify that each random bit (whether the vertex belongs to S or not) can be pairwise independent.

   (b) Using pairwise independent bits (the generation was shown in class), construct an efficient algorithm to find such a cut.

   Solution:
   A)
   Let v_i denotes the i^{th} vertex of the graph where i $\in [n]$.
   Enumerate the edges of the graph, such that $e_1, e_2....e_m$ where m=|E|.
   Let X_i be the random variable that denotes whether e_i is present in the cut.
   $X_i := \begin{cases} 1, & \text{if } e_i \text{ is present in the cut} \\ 0, & \text{otherwise} \end{cases}$
   For an edge to be present in the cut its both ends should lie in different sets(S and $\bar{S}$).
   Let v_1 and v_2 be the ends to the edge,WLOG.
   P[$e_i \in$ cut]=P[$v_1 \in S$,$v_2 \in \bar{S}$]+P[$v_2 \in S$,$v_1 \in \bar{S}$]
   Given P[$v_i \in$ S]=1/2,P[$v_i \in \bar{S}$]=1/2, which is independent of other vertices present in the graph.
   P[$v_1 \in S$,$v_2 \in \bar{S}$]=P[$v_1 \in$ S].P[$v_2 \in \bar{S}$]=1/2.1/2=1/4
   Similarly P[$v_2 \in S$,$v_1 \in \bar{S}$]=1/2.1/2=1/4
   Hence P[$e_i \in$ cut]=$P[X_i = 1] = 1/2$
   $E[X_i] := \sum\limits_{x \in \{0,1\}} \Pr(X_i = x) \cdot x = \Pr(X_i = 0) \cdot (0) + \Pr(X_i = 1) \cdot (1) = 1/2$
   Define $X = \sum_{i \in [m]} X_i$, which denotes the size of the cut.
   Our task is to find E[X] (size of the cut)

$E[X] = E[\sum_{x\in[m]} X_i] = \sum_{x\in[m]} E[X_i]$(By Linearity of Expectation)
$E[X] = \sum_{x\in[m]} 1/2 = m/2 = |E|/2$
Justification for pairwise independent bits:
In the above proof we used the pairwise independence of bits
P[$v_1 \in S$,$v_2 \in \bar{S}$]=P[$v_1 \in$ S].P[$v_2 \in \bar{S}$]=1/2.1/2=1/4:
We don't need mutual independence to reach to the conclusion, Hence pairwise independence is sufficient.
b)Given:G=(V,E) which is the input to the algorithm.
Let v_i denotes the i^{th} vertex of the graph where i $\in [n]$
The output of algorithm is cut (S, $\bar{S}$).
FindCut(G=(V,E)){
$S = v_1$;
$\bar{S}$=;
for(int i=2;i<=n;i++){
v_i will go in that set $(S, \bar{S})$ where the neighbors are less, this will ensure that i^{th} vertex contributes to the max edges in the cut.
int a=num of neighbors of V_i in S
int b==num of neighbors of V_i in $\bar{S}$
if(a<b){
add the vertex in S;
}
else if(a<b){
add the vertex in $\bar{S}$;
}
else{
place the vertex arbitrarily in S or $\bar{S}$
}
return (S, $\bar{S}$)
}
This algorithm generates the cut having size >= |E|/2, because each vertex is placed such that it contributes to max edges, Which is greater than equal to half the total number of edges passing through that vertex. Hence the expected number of edges in the cut is greater than |E|/2
This algorithm runs in polynomial time. Because here we search for the neighbors which takes linear time, and we are doing this for n vertices.
Hence this is a polynomial time algo. □

2. **(10 marks)** Let U be a set of n elements and $S_1, S_2, \cdots, S_m$ be subsets of U. For a function $f : U \to \{-1, 1\}$, define $f(S_i) := \sum_{x\in S_i} f(x)$. Show that there exist a function f such that for all i, $f(S_i) \le 100\sqrt{n \ln m}$.

   Hint: Assign $1, -1$ randomly to U. Find the probability that a particular S_i is bad. What is the expected number of S_i which are bad?

   Solution:
   Define $g : U \to \{0, 1\}$, st $g(x) = (1 + f(x))/2$ $\forall x \in U$
   Define $g(S_i) := \Sigma_{x\in S_i}$

$g(x) := \Sigma_{x \in S_i}(1 + f(x))/2 := |S_i|/2 + \Sigma_{x \in S_i} f(x)/2 = (|S_i| + f(S_i))/2$
Define X_f be the random variable associated with the value of f(x) where $x \in$ U
Define X_g be the random variable associated with the value of g(x) where $x \in$ U
Define X_{fSi} be the random variable associated with the value of f(S) where $S \subseteq U$
Define X_{gSi} be the random variable associated with the value of g(S) where $S \subseteq U$
Given: Assign $1, -1$ randomly to U
(Assuming uniform assignment $p[X_f = -1] = 1/2, p[X_f = 1] = 1/2$)
$E[X_f] := \sum\limits_{x \in \{-1,1\}} \Pr(X(\omega) = x) \cdot x = \Pr(X_f = -1) \cdot (-1) + \Pr(X_f = 1) \cdot (1) = 0$
Since f is uniformly assigned, g would also be uniformly assigned(because $g(x) = (1 + f(x))/2$)
$p[X_g = 0] = 1/2, p[X_g = 1] = 1/2$
$E[X_g] := \sum\limits_{x \in \{0,1\}} \Pr(X(\omega) = x) \cdot x = \Pr(X_f = 0) \cdot (0) + \Pr(X_f = 1) \cdot (1) = 1/2$
$E[X_{fSi}]$=E$[\sum_{x \in S_i} f(x)] = \sum_{x \in S_i} E[f(x)]$(By Linearity of Expectation)
$E[X_{fSi}]$=$\sum_{x \in S_i} E[X_f] = 0$
$E[X_{gSi}]$=E$[\sum_{x \in S_i} g(x)] = \sum_{x \in S_i} E[g(x)]$(By Linearity of Expectation)
$E[X_{gSi}]$=$\sum_{x \in S_i} E[X_g] = |S_i|/2$
Let Bi be the random variable that denotes S_i is bad and i$\in [m] i.e.$
$$B_i := \begin{cases} 1, & \text{if } f(S_i) > 100\sqrt{n \ln m} \text{ (Bad)} \\ 0, & \text{otherwise} \end{cases}$$
Now, let's find the $P[S_i \text{is bad}] = P[f(S_i) > 100\sqrt{n \ln m}]$
$f(S_i) = 2 * g(S_i) - |S_i|$
$P[f(S_i) > 100\sqrt{n \ln m}] = P[2 * g(S_i) - |S_i| > 100\sqrt{n \ln m}] = P[g(S_i) > 50\sqrt{n \ln m} + |S_i|/2]$
$P[f(S_i) > 100\sqrt{n \ln m}] = P[g(S_i) > |S_i|/2(100\sqrt{n \ln m}/|S_i| + 1)]$
Applying chernoff bound where $nE[X_g] = |S_i|/2$ and $\delta = 100\sqrt{n \ln m}/|S_i|$
$P[g(S_i) > nE[X_g](\delta + 1)] < e^{-nE[X_g]\delta^2/3} = e^{-(|S_i|/2)(100\sqrt{n \ln m}/|S_i|)^2/3} = e^{-(10000 nlnm)/6|S_i|}$
$P[g(S_i) > nE[X_g](\delta + 1)] < e^{-(10000 \ln m)/6}$ $Since |S_i| <= n$
P[g(S_i) $> nE[X_g](\delta + 1)] < e^{-(1666 \ln m)} = m^{-1666}$
$P[B_i = 1] < m^{-1666}$
$E[B_i] := \sum\limits_{x \in \{0,1\}} \Pr(B_i = x) \cdot x = \Pr(B_i = 0) \cdot (0) + \Pr(B_i = 1) \cdot (1) < m^{-1666}$
Define Random Variable B=$\Sigma_{i \in [m]} B_i$
E[B]=E$[\sum_{i \in [m]} B_i] = \sum_{i \in [m]} E[B_i] < m.m^{-1666} = m^{-1665}$(By Linearity of Expectation)
For m>1 (number of subsets of U), $m^{-1665} < 1$ or even negligible.
This implies E[B]<1
Now let's assume there doesn't exists a function f such that for all i, $f(S_i) \leq 100\sqrt{n \ln m}$
That means $\exists j \in [m]$ such that $f(S_i) > 100\sqrt{n \ln m}$ or $B_j = 1 or B >= 1$
$E[B] = \sum\limits_{j \in range(B)} \Pr(B = j) \cdot j$
Since j>1, $E[B] >= \sum\limits_{j \in range(B)} \Pr(B = j)$=1
but from above we have $E[B] < 1$
Hence there doesn't exits such a j, which is a contradiction.
Hence, there exists a function such that for all i, $f(S_i) \leq 100\sqrt{n \ln m}$

$\square$