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: Siddhant Suresh Jakhotiya

Roll No: 211030

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)

$$\text{Let } |V| = n \text{ and } |E| = m$$

We start by enumerating all the edges of the graph $G = (V, E)$.

Denote $e_1, e_2, ...e_m$ as a random enumeration of the edges of the graph.

Now, we will define m random variables $X_1, X_2, ..., X_m$ as follows:

$$X_i = \left\{ \begin{array}{ll} 1, & \text{if edge } e_i \text{ joins a vertex in } S \text{ to a vertex in } \bar{S} \\ 0, & \text{otherwise} \end{array} \right\} \text{for } i \in [m]$$

...(1)

Now, for any edge e_i, say that it joins vertex v_{i_1} to vertex v_{i_2}. We have 4 possible cases:

$$\begin{array}{l|l} v_{i_1} \text{ is in } S \text{ and } v_{i_2} \text{ is in } \bar{S} & X_i = 1 \\ v_{i_1} \text{ is in } S \text{ and } v_{i_2} \text{ is in } S & X_i = 0 \\ v_{i_1} \text{ is in } \bar{S} \text{ and } v_{i_2} \text{ is in } \bar{S} & X_i = 0 \\ v_{i_1} \text{ is in } \bar{S} \text{ and } v_{i_2} \text{ is in } S & X_i = 1 \end{array}$$

Define events

$$A : \text{Vertex } v_{i_1} \in S$$
$$B : \text{Vertex } v_{i_2} \in S$$

As mentioned in the question,

$$P(A) = P(B) = \frac{1}{2}$$

Now, since the distribution of all the n vertices into S and $\bar{S}$ is mutually independent, we have

$$P(A \cap B) = P(A)P(B) = 1/4$$

Similarly, we can prove that

$$P(\bar{A} \cap B) = P(A \cap \bar{B}) = P(\bar{A} \cap \bar{B}) = \frac{1}{4}$$

...(2)

Thus each of the cases in the table is equally likely so we have

$$P(X_i = 1) = P(X_i = 0) = \frac{1}{2}$$

$$\therefore E[X_i] = 1 \times \frac{1}{2} + 0 \times \frac{1}{2} = \frac{1}{2} \ \forall \text{ i} \in \text{[m]}$$

Now, let the size of the cut be denoted by the random variable Z.

$$Z = \sum_{i=1}^{m} X_i$$

By linearity of expectation,

$$E[Z] = \sum_{i=1}^{m} E[X_i] = \frac{m}{2} = \frac{|E|}{2}$$

...(3)

Proof that the above holds even in the case of pairwise independence:
The above proof can be divided broadly into 3 parts (as marked above by 1, 2 and 3).
Part 1 merely gives an enumeration to the edges and defines a random variable according to our given definitions. It doesn't depend upon independence.

Part 2 relies upon the events being independent.
However, notice that the equality derived for $P(A \cap B)$,$P(A \cap \bar{B})$, $P(\bar{A} \cap B)$ and $P(\bar{A} \cap \bar{B})$ hold true even when we only know that the distribution of any 2 vertices $v_k, v_l \in V, \forall l, k \in [n], l \neq k$ is only pairwise independent (mutually independence is not necessary to derive these equalities).

Part 3 involves the calculation of the expectation of different random variables. Even linearity of expectation doesn't depend upon independence.

Thus, throughout the proof, we haven't relied upon the assumption of mutual independence anywhere and simply having pairwise independence is sufficient.

b)
Assign a random enumeration to the vertices as $v_1, v_2, ...v_n$.
We have n pairwise independent bits $Y_1, Y_2, ...Y_n$ which can be defined as

$$Y_i = \left\{ \begin{array}{ll} 1, & \text{if } v_i \in S \\ 0, & \text{otherwise} \end{array} \right\}$$

As discussed in class, these bits can be generated by $b = \lceil log_2(n+1) \rceil$ using the XOR operation over b mutually independent uniform random bits (that take on value 0 or 1 with equal probability).
Short recap from the class: let these b bits be $W_1, W_2 ... W_b$ and enumerate $2^b - 1$ non-empty subsets of {1, 2, ... b} in some order. Let S_j be the jth set in the ordering then

$$Y_j = \bigoplus_{i \in S_j} X_i$$

As was proved in class, the n bits thus generated are uniform and pairwise independent. As derived before, for a pairwise independent family, $\{Y_i\}_{i=1}^{n}$, the expectation is m/2.

Claim: $\exists$ a cut of size $\geq$ m/2

Proof: Consider a graph $G = (V, E)$. (Assume |V| > 3)

$$\text{Let v be a vertex s.t. } deg(v) \leq deg(w) \forall w \in V$$

$$\therefore m = \frac{\sum_{i=1}^{n} deg(v_i)}{2} \geq \frac{n \times deg(v)}{2}$$

$$\therefore deg(v) \leq \frac{2m}{n} \leq \frac{m}{2} \because n \geq 4$$

Define S = {v}. Clearly, the size of the cut is $\leq$ m/2.
When n = 1, we can obtain a cut of size |E| trivially since |E| = 0.
When n = 2, we can pick any 1 vertex in S and we get the size of the cut as |E|.
We have proved that $P(Z \leq \frac{m}{2}) \geq 0$. We also know that E[Z] = $\frac{m}{2}$.

$$E[Z] = \sum_{i=1}^{\lfloor m/2 \rfloor} i \times P(Z = i) + \sum_{i=m/2}^{m} i \times P(Z = i)$$

If $P(Z \geq m/2) = 0$, we have E[Z] = $\sum_{i=1}^{\lfloor m/2 \rfloor} i \times P(Z = i) \leq m/2$: A contradiction.
For n = 3,

- if m = 3 (a complete graph), we can prove that any non-null choice of S will result in cut of size 2.
- if m = 2, $\exists$ a vertex s.t. deg(v) = 2. Define S = {v}. The size of the cut is m > m/2.

- if m = 1, let the edge be e = (v_i, v_j). Define S = $\{v_i\}$. Again, the size of the cut = m > m/2.
- if m = 0, any selection of S trivially gives us size of the cut > m/2.

Now, we know that we can obtain a cut of size m/2.
Our sample space has been reduced to the possible choices of the b random bits and we need an algorithm to deterministically find a cut of size > m/2.

z We can try all possible 2^b settings of the b bits to find such a cut (we have proved that it exists so there is at least 1 setting of the b bits for which we get such a cut).
It takes O(n) time to fix a set S ($O(2^b) = O(n)$ since $b = \lceil log_2(n+1) \rceil$)

Now, to calculate the size of the cut, we can iterate over all the edges to determine of a given edge contributes to the size of the cut or not.
This can be done in O(m) time.

Thus we have an algorithm that finds a cut in $O(mn)$ time. $\square$

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 $f : U \rightarrow \{-1, 1\}$, define $f(S_i) := \sum_{x \in S_i} f(x)$. Show that there exist a function f such that for all i, $f(S_i) \leq 100\sqrt{n \ln m}$.

   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: We define another function $p : U \rightarrow \{0, 1\}$, such that,

$$p(x) = \frac{f(x) + 1}{2}$$

   Whenever f(x) = -1, p(x) = 0 and whenever f(x) = 1, p(x) = 1.
   Thus assigning -1 and 1 randomly (where each x is assigned mutually independently) to f(x) is similar to assigning 0 and 1 (respectively) to p(x).

$$\therefore P(p(x) = 1) = 1/2 \text{ and } P(p(x) = 0) = 1/2$$

   We can also extend the definition of $f(S_i)$ as

$$p(S_i) = \sum_{k \in S_i} g(k)$$

$$\therefore g(S_i) = \sum_{k \in S_i} \frac{f(x) + 1}{2}$$

$$\therefore g(S_i) = \frac{f(S_i) + |S_i|}{2}$$

   Note that g(x) is a Bernoulli random variable with p=1/2. **Chernoff Bound** states that

   Let X be a random variable which takes value 1 with probability p and 0 otherwise. Now if we have n copies of the random variable as $X_1, X_2 ... X_n$, where the family $\{X_i\}_{i=1}^n$ is mutually independent. Define $S = \sum_{i=1}^n X_i$ then

$$P(S > (1 + \delta)nE[X]) \leq e^{-\frac{nE[X]\delta^2}{3}}$$

   In our problem, we can represent S as $g(S_i)$ and X as $g(x)$. Note that E[g(x)] = 1/2.

$$\therefore P(g(S_i) > (1 + \delta)|S_i|/2) \leq e^{-\frac{|S_i|\delta^2}{6}}$$

$$\therefore P(2g(S_i) - |S_i| > \delta|S_i|) \leq e^{-\frac{|S_i|\delta^2}{6}}$$

   So we get,

$$P(f(x) > \delta|S_i|) \leq e^{-\frac{|S_i|\delta^2}{6}}$$

...(1)

   We need to prove that $f(x) \leq 100\sqrt{n \text{ ln} m}$ So,

$$\delta|S_i| = 100\sqrt{n \text{ ln} m}$$

Or,

$$\delta = \frac{100\sqrt{n \text{ ln}m}}{|S_i|}$$

Substituting this delta in the inequality obtained using Chernoff Bound (1), we get,

$$P(f(x) > 100\sqrt{n \text{ ln}m}) \leq e^{-\frac{10^4 n \text{ ln}m}{6|S_i|}} \leq e^{-\frac{5000 \text{ ln}m}{3}} = \frac{1}{m^{5000/3}} (\because |S_i| \leq n)$$

Define m random variables $X_1, X_2, ...X_m$ s.t.

$$X_i = 1 \text{ if } S_i \text{ is bad and } 0 \text{ otherwise, } \forall i \text{ in } \{1, 2, ..m\}$$

Now, define

$$Y = \sum_{i \in [m]} X_i$$

$$\therefore E[Y] = m \times E[X] = m \times (1 \times P(f(x) > 100\sqrt{n \text{ ln}m}))$$

$$\therefore E[Y] = m \times \frac{1}{m^{5000/3}} = \frac{1}{m^{4997/3}} < 1$$

Suppose

$$\exists i \in [m] \text{ s.t. } S_i \text{ is bad, i.e. } f(S_i) > 100\sqrt{n \text{ ln}m} \forall f$$

In that case, $X_i = 1$ for some $i \in [m] \forall f$.

$$\therefore Y = \sum_{i \in [m]} X_i \geq 1 \forall f \rightarrow E[Y] \geq 1$$

Which is a contradiction. Hence, there exists an f s.t. for all i, S_i are not bad.
This was valid in the case when f(x) > 0 (since there was a correction regarding $|f(S_i)|$ instead of $f(S_i)$

For f(x) < 0, we need to prove that $-f(x) < 100\sqrt{n \text{ ln}m}$
Again, using the Chernoff Bound,

$$P(g(S_i) < (1 - \delta)|S_i|/2) \leq e^{-\frac{|S_i|\delta^2}{4}}$$

$$P(\delta|S_i| < -f(x)) \leq e^{-\frac{|S_i|\delta^2}{4}}$$

Again, solving for δ, we get

$$\delta = \frac{100\sqrt{n \text{ ln}m}}{|S_i|}$$

Substituting it into the inequality we get,

$$P(100\sqrt{n \text{ ln}m} < -f(x)) \leq e^{-\frac{5000 n lnm}{2|S_i|}} \leq \frac{1}{m^{5000/2}} \because |S_i| \leq n$$

Again, we define the same random variables and we get

$$E[Y] = m \times \frac{1}{m^{5000/2}} < 1$$

And if there is an i s.t. $f(S_i)$ is bad for all f then E[Y] ≥ 1 : a contradiction. □