1 Problem 1

1.1 Modified Algorithm

Given weighted graph $\mathbf{G}(\mathbf{V}, \mathbf{E}), |\mathbf{V}| = \mathbf{n}$. Assuming the min-cut is $\mathbf{C}$, and the weights of all edges in $\mathbf{C}$ is $\mathbf{W}$. We define the weight of the edge $\mathbf{e_k}$ to $\mathbf{w_k}, k = 1, 2, \cdots, |E|$. The original contraction algorithm can be modified as below:

Algorithm 1 Modified Karger's Contraction Algorithm

Input: G(V,E)

Output: Min-cut C

  1: **function Modified-Contraction**(G)

  2:      **while $|\mathbf{V}| > \mathbf{2}$ do**

  3:          Select edge $\mathbf{e_k} \in \mathbf{E}$ with probability $\frac{\mathbf{w_k}}{\sum_{\mathbf{j=1}}^{|\mathbf{E}|} \mathbf{w_j}}, k = 1, 2, \cdots, |E|$;

  4:          Merge the endpoints of $\mathbf{e_k}$, producing graph $\mathbf{G'}$;

  5:          **Modified-Contraction(G')**;

  6:      **return** sets of edges remaining in the graph as $\mathbf{C}$

1.2 Proof for the success probability of algorithm

In the i-th ground contraction, the probability to choose an edge which is in the weighted min-cut $\mathbf{C}$ is **at most**:

$$\frac{W}{\frac{W(n-i)}{2}} = \frac{2}{n-i}$$

The algorithm perform $n-2$ contractions on the graph, we get this lower bound:

$$Pr\big[\text{return weighted min-cut } \mathbf{C}\,\big] \geq \prod_{i=0}^{n-3}(1 - \frac{2}{n-i})$$

$$= \prod_{i=0}^{n-3} \frac{n-i-2}{n-i} \qquad (1)$$

$$= \frac{2}{n(n-1)}$$

2 Problem 2

2.1 Calculate the expected value

For any $k \in \{0, 1, 2, \cdots, m\}$, the probability of $|E(S,T)| = k$ should be:

$$Pr\big[|E(S,T)| = k\big] = \frac{\binom{m}{k}}{2^m} \tag{2}$$

The expected value can be calculated as below:

$$
\begin{aligned}
\mathbb{E}\Big[|E(S,T)|\Big] &= \sum_{k=0}^{m} Pr\big[|E(S,T)| = k\big] \cdot k \\
&= \sum_{k=0}^{m} \frac{\binom{m}{k}}{2^m} \cdot k \\
&= \frac{1}{2^m} \sum_{k=0}^{m} \binom{m}{k} \cdot k \\
&= \frac{1}{2^m} 2^{m-1} m \\
&= \frac{m}{2}
\end{aligned}
\tag{3}
$$

2.2 Algorithm to generate subset

Algorithm 2 Subroutine to generate Subset

Input: $\mathcal{R}(\cdot)$, **V**

Output: S

 1: **function** GENERATE($\mathcal{R}(\cdot)$, **V**)

 2:      Initialize $\mathbf{S} = \emptyset$

 3:      **while** $|\mathbf{S}| < \frac{n}{2}$ **do**

 4:          **for** $v \in \mathbf{V}$ **do**

 5:             $p = \mathcal{R}(v)$

 6:             **if** $p \geq \frac{1}{2}$ **then**

 7:                 $\mathbf{S} = \mathbf{S} \cup \{v\}$;

 8:      **return S**

To analysis the correctness of the algorithm, we need to prove that: $\forall S \in \mathcal{F}$, the uniformal probability is given below:

$$\Pr[\text{ Choose } \mathbf{S}] = \frac{1}{\binom{n}{\frac{n}{2}}}, \qquad \forall S \in \mathcal{F}.$$

Given random variables generated by the algorithm, $\Pr[X_i = 1] = \frac{1}{2}, i = 1, 2, \cdots, n$, For any $\mathbf{S}$ is chosen only and if only when there exits $\frac{n}{2}$ random variables is equal to 1. Define $Y_i = \sum_{k=0}^{n} X_k, i = 1, 2, \cdots, \binom{n}{\frac{n}{2}}$.

$$
\begin{aligned}
\mathbb{E}\big[Y_i = \frac{n}{2}\big] &= \mathbb{E}\big[\sum_{k=0}^{n} X_k = \frac{n}{2}\big] \\
&= \binom{n}{\frac{n}{2}} \mathbb{E}[X_i] \\
&= \frac{1}{2}\binom{n}{\frac{n}{2}}
\end{aligned}
\tag{4}
$$

This is exactly the same with uniform distribution.

In the procedure of the algorithm, if we have generate a qualified subset S, the algorithm finish immediately, expected number of times to call the $\mathcal{R}(\cdot)$ subroutine can be derived as below:

$$
\begin{aligned}
\mathbb{E}\big[N\big] &= \sum_{k=\frac{n}{2}}^{n} \Pr[N_k = k] \cdot k \\
&= \sum_{k=\frac{n}{2}}^{n} \binom{k-1}{\frac{n}{2}-1}\big(\frac{1}{2}\big)^{k-1} \cdot \frac{1}{2} \cdot k \\
&= \sum_{k=\frac{n}{2}}^{n} \binom{k-1}{\frac{n}{2}-1}\big(\frac{1}{2}\big)^{k} \cdot k
\end{aligned}
\tag{5}
$$

3 Problem 3

As vertices in the rooted tree has recursive structure.We consider to assign a polynomial to each vertex in the tree. For earch vertex $v \in T$ with height h, the polynomial is defined as :

$$
P_v = \begin{cases} x_0, & h = 0, \\ \prod_{i=1}^{k}\left(x_h - P_{u_i}\right), & h \geq 1 \end{cases}
$$

Given rooted tree T with height h and n vertices, according to the polynomial defined above, there are $h+1$ variables in $P_{root(T)} \rightarrow \{x_0, x_1, \ldots, x_h\}$, it's necessary for us to figure out the degree of $P_{root(T)}$. Expand $P_{root(T)}$, we get:

$$P_{root(T)} = (x_h - P_{v_1})(x_h - P_{v_2}) \cdots (x_h - P_{v_k})$$
$$= x_h^k + \ldots + \prod_{i=1}^{k} P_{v_i} \tag{6}$$

$P_{root(T)}$ is then divided into 3 parts. We analysis the degree one by one.

$$\mathbf{Deg}\left(x_h^k\right) = k$$

$$\mathbf{Deg}\left(\ldots\right) = \max_{1 \le i \le k-1} \sum_{k=1}^{i} \mathbf{Deg}\left(P_{v_k}\right) + k - i$$

$$\mathbf{Deg}\left(\prod_{i=1}^{k} P_{v_i}\right) = \sum_{i=1}^{k} \mathbf{Deg}\left(P_{v_i}\right) = n$$

As $\forall i \in \{1, 2, \ldots, k\}, \mathbf{Deg}\left(P_{v_i}\right) \ge 1$. It's easy to prove that:

$$n = \mathbf{Deg}\left(\prod_{i=1}^{k} P_{v_i}\right) \ge \mathbf{Deg}\left(\ldots\right) \ge \mathbf{Deg}\left(x_h^k\right) = k$$

where n is the number of vertices in rooted tree T.

Now it's clear that $P_{root(T)}$ is a $h+1$ multivariate polynomial with degree n. We are going to construct polynomial for rooted trees T_1 and T_2, and prove that T_1 and T_2 are isomorphic is equivalent to $P_{root(T_1)} \equiv P_{root(T_2)}$.

Applying induction on the height of rooted trees.

1 If $h = 0$, it's trivial that $P_{root(T_1)} = x_0$ and $P_{root(T_2)} = x_0$, it's obvious T_1 and T_2 are isomorphic.

2 Assume that $\forall h \le k$, we have $P_{root(T_2)} \equiv P_{root(T_2)}$ equivalent to the isomorphism between T_1 and T_2.

3 For $h = k + 1$, expand two polynomials, we have all terms are children vertices of $root(T_1)$ and $root(T_2)$, they are with heights $h = k$, can be verified by our hypothesis. Thus, it still works.

Combining all of above, **we can test the isomorphism of rooted trees by test their polynomial identity**.

We construct a polynomial as the subtraction of two polynomials:

$$Q\left(x_0, x_1, \ldots, x_h\right) = P_{root(T_1)} - P_{root(T_2)} \tag{7}$$

We can test its identity on field $\mathbb{Z}_p$, where prime $p \geq 2n$.

According to **Schwartz-Zippel Theorem**, the error probability:

$$\Pr\left[Q(r_0, r_1, \ldots, r_h) = 0\right] \leq \frac{n}{p} \leq \frac{n}{2n} = \frac{1}{2}$$

We can independently perform such identity test for $\log n$ times, thus we can reduce the error probability to:

$$\left(\frac{1}{2}\right)^{\log n} = O\left(\frac{1}{n}\right)$$

4 Problem 4

Consider we have a bijection $\sigma \in \mathbf{S_n}$, where $\mathbf{S_n}$ is the universe of permutations of set $\{1, 2, \ldots, n\}$, σ maps a_i to b_{σ_i}, $1 \leq i \leq n$. That's to say, if $a_1, \ldots, a_n$ is a permutation of $b_1, \ldots, b_n$, a σ should be exist. Therefore, we can define **Edmonds** Matrix $\mathbf{A}$:

$$\mathbf{A}_{ij} = \begin{cases} x_{ij}, & if \ a_i = b_{\sigma_i} \left(\sigma_i = j\right), \\ 0, & otherwise. \end{cases}$$

Matrix $\mathbf{A}$ has determinant:

$$det(A) = \sum_{\sigma \in \mathbf{S_n}} sgn(\sigma) \prod_{i=1}^{n} A_{i,\sigma_i}$$

The determinant is a summation over $n!$ terms. It is helpful to analysis the result of $det(A)$, it is non-zero if and only if when $a_1, a_2, \ldots, a_n$ is a permutation of $b_1, b_2, \ldots, b_n$.

Thus, We can construct a polynomial $\mathbf{Q}(x_{11}, x_{12}, \ldots, x_{nn}) = det(\mathbf{A})$. The problem of determination of permutation now transform to determine if $\mathbf{Q}(x_{11}, x_{12}, \ldots, x_{nn})$ is identically zero or not.

To solve the problem, we first determine the degree of n^2 multivariate polynomial $\mathbf{Q}(x_{11}, x_{12}, \ldots, x_{nn}) \leq n$. We can test its identity on the field $\mathbb{Z}_p$, where prime $p \geq 2n$.

According to **Schwartz-Zippel Theorem**, the error probability:

$$\Pr\left[\mathbf{Q}(r_{11}, r_{12}, \ldots, r_{nn}) = 0\right] \leq \frac{n}{p} \leq \frac{n}{2n} = \frac{1}{2}$$

The computation time complexity is dominated by the calculation of determinant, it's $O\left(n^3\right)$.

We can independently perform such identity test for $\log n$ times, thus we can reduce the error probability to:

$$\left(\frac{1}{2}\right)^{\log n} = O\left(\frac{1}{n}\right)$$

Combining all of above together, we can solve this problem in $O\left(n^3 \log n\right)$ with error probability $O\left(\frac{1}{n}\right)$.

5 Problem 5

The original problem can be transformed to the following form:

$$\Pr\left[e^{-\epsilon}Z \leq \hat{Z} \leq e^{\epsilon}Z\right] \geq 1 - \delta$$
$$\Longleftrightarrow \Pr\left[\hat{Z} \leq e^{-\epsilon}Z \text{ or } \hat{Z} \geq e^{\epsilon}Z\right] \leq \delta \tag{8}$$
$$\Longleftrightarrow \Pr\left[|\ln \hat{Z} - \ln Z| \geq \epsilon\right] \leq \delta$$

It's obvious that:

$$\mathbb{E}\left[\ln \hat{Z}\right] = \mathbb{E}\left[\sum_{i=1}^{n} \hat{\rho}_i\right]$$

$$= \sum_{i=1}^{n} \mathbb{E}\left[\hat{\rho}_i\right]$$

$$= \sum_{i=1}^{n} \mathbb{E}\left[\frac{1}{s}\sum_{j=1}^{s} X_i^{(j)}\right] \tag{9}$$

$$= \sum_{i=1}^{n} \rho_i$$

$$= \ln Z$$

Applying Chebyshev's Inequality, we obtain:

$$\Pr\left[|\ln \hat{Z} - \mathbb{E}\left[ln\hat{Z}\right]| \geq \epsilon\right] \leq \frac{\mathbf{Var}\left[\ln \hat{Z}\right]}{\epsilon^2}$$

Therefore, we need to calculate the variance of $\ln \hat{Z}$.

$$\mathbf{Var}\left[\ln \hat{Z}\right] = \mathbb{E}\left[\left(\ln \hat{Z}\right)^2\right] - \mathbb{E}^2\left[\ln \hat{Z}\right]$$

$$= \mathbb{E}\left[\left(\sum_{i=1}^{n} \hat{\rho}_i\right)^2\right] - \left(\sum_{i=1}^{n} \rho_i\right)^2 \tag{10}$$

The first term can be expanded as below:

$$\mathbb{E}\left[\left(\sum_{i=1}^{n} \hat{\rho}_i\right)^2\right] = \mathbb{E}\left[\sum_{i=1}^{n} \hat{\rho}_i^{\ 2} + 2\sum_{1\leq i<j\leq n} \hat{\rho}_i \cdot \hat{\rho}_j\right]$$

$$= \sum_{i=1}^{n} \mathbb{E}\left[\hat{\rho}_i^{\ 2}\right] + 2\sum_{1\leq i<j\leq n} \mathbb{E}\left[\hat{\rho}_i \cdot \hat{\rho}_j\right]$$

$$= \sum_{i=1}^{n} \left(\frac{1}{s}\mathbf{Var}\left[X_i\right] + \mathbb{E}^2\left[\rho_i\right]\right) + 2\sum_{1\leq i<j\leq n} \mathbb{E}\left[\hat{\rho}_i\right] \cdot \mathbb{E}\left[\hat{\rho}_j\right] \tag{11}$$

$$= \sum_{i=1}^{n} \left(\frac{1}{s}\rho_i \cdot (1-\rho_i) + \rho_i^2\right) + 2\sum_{1\leq i<j\leq n} \rho_i \cdot \rho_j$$

7

Combining all of above:

$$\Pr\left[|\ln\hat{Z} - \mathbb{E}\left[ln\hat{Z}\right]| \geq \epsilon\right] \leq \frac{\mathbf{Var}\left[\ln\hat{Z}\right]}{\epsilon^2}$$

$$= \frac{\sum_{i=1}^{n}\left(\frac{1}{s}\rho_i \cdot (1-\rho_i) + \rho_i^2\right) + 2\sum_{1 \leq i < j \leq n}\rho_i \cdot \rho_j - \left(\sum_{i=1}^{n}\rho_i\right)^2}{\epsilon^2}$$

$$= \frac{\frac{1}{s}\sum_{i=1}^{n}\rho_i \cdot (1-\rho_i)}{\epsilon^2}$$

$$\leq \frac{n}{4\epsilon^2}$$

$$\leq \delta \tag{12}$$

Solve the equation above, we get s:

$$s \geq \frac{n}{4\delta\epsilon^2}$$

So, we take $s = \left\lceil \frac{n}{4\delta\epsilon^2} \right\rceil$.

6 Problem 6

6.1 Question 1

Feel difficult

6.2 Question 2

As the first $\frac{n}{2}$ balls thrown into bins using two-choice paradigm, the maximum load is expected to $\Theta\left(\log\log\frac{n}{2}\right) = \Theta\left(\log\log n\right)$, the remaining $\frac{n}{2}$ balls thrown into bins independently and dependently, the maximum load is $\Theta\left(\frac{\log n}{\log\log n}\right)$, the maximum load is expected to be:

$$\Theta\left(\log\log n + \frac{\log n}{\log\log n}\right) = \Theta\left(\log\log n\right)$$

6.3 Question 3

Feel difficult