CS648 project

Team Binary Expectation
Aryan Bhardwaj | Sanat Goel | Tushaar Ranganathan
Expected Value of Weight of MST

2023-24 2nd Semester

1 Problem Description

What is the expected weight of the MST in a complete graph where each edge is assigned weight randomly uniformly in the range [0,1]?

2 Method 1

Our Algorithm

Step 1: We construct a spanning tree for the first $\frac{n}{2}$ vertices following Prim's algorithm,with the last $(\frac{n}{2})$ $^{\text{th}}$ edge having an expected weight of $\frac{2}{n}$.

Let these nodes be called Red.

Step 2: Among the remaining $\frac{n}{2}$ vertices,we connect $\frac{n}{3}$ of these vertices to our spanning tree of $\frac{n}{2}$ vertices created in Step 1.

Let the $\frac{n}{3}$ vertices added in this step be called orange vertices.

This step is done as follows, go to these orange vertices one by one, for a 1^{st} orange node consider all the edges connecting this orange node to all the green nodes, select the minimum of it and add it in the spanning tree.The green nodes are part of the red nodes in step 1, whose connected outgoing edge value in spanning tree is less than or equal to $\frac{2}{n}$. In other words, if the i^{th} edge of spanning tree is less than $\frac{2}{n}$, then $(i-1)^{th}$ node is a green node.

For subsequent orange nodes, consider the edges connecting this orange node to all the green nodes and to other orange nodes already added to our spanning tree, select the minimum of it and add it to the spanning tree. This way the orange nodes are connected.

Step 3: We are left with $\frac{n}{6}$ vertices, which we connect to the $\frac{n}{3}$ vertices added to spanning tree in Step 2.

Let's call these remaining $\frac{n}{6}$ nodes as yellow vertices.

This step is done as follows, go to these yellow vertices one by one, for a given yellow node consider all the edges connecting this yellow node to all the orange nodes, and select the minimum of it.

Notice that none of the edge connecting these orange and yellow vertices were revealed or compared to with other edges in previous steps, so these edges follow $Unif(0,1)$ distribution.

Note: Unif(a,b), where $a < b$, is the continuous random variable that is equally likely to take any value between a and b, and has zero probability of taking a value less than a or greater than b.

We will now be providing the analysis for Step 2.

2.1 Analysis

First of all, we will prove that the median weight of all the edges added in step 1, is less than $\frac{2}{n}$.

Note: The median of a random variable X is the number m such that $P(X \leq m) = \frac{1}{2}$.

Note: the ith edge added in Step 1 is min of (n-i+1) Unif[0, 1] random variables

For simplicity lets consider the case with n Unif(0, 1) random variable.

Let X1, X2, ... Xn be n Unif(0, 1) random variable

let $X_{min} = min(X_1, X_2, \ldots, X_n)$

$$\text{Prob}(X_{\text{min}} > t) = \text{Prob}(X_1 > t, X_2 > t, \ldots, X_n > t)$$

$$= \text{Prob}(X_1 > t) \cdot \text{Prob}(X_2 > t) \cdot \ldots \cdot \text{Prob}(X_n > t)$$
$$= (1-t)^n$$

So

$$\text{prob}(X_{\text{min}} \leq t) = 1 - (1-t)^n$$

Let x_{med} denote the median of X_{min} random variable.

Since

$$\text{prob}(X_{\text{min}} \leq x_{\text{med}}) = 1 - (1 - x_{\text{med}})^n$$

but

$$\text{prob}(X_{\text{min}} \leq x_{\text{med}}) = \frac{1}{2}$$

so

$$1 - (1 - x_{\text{med}})^n = \frac{1}{2}$$

solving this we get

$$x_{\text{med}} = 1 - 2^{-\frac{1}{n}}$$

Lets prove $1 - 2^{-x} < x$ for positive x, let's consider the function $f(x) = 1 - 2^{-x} - x$. We will show that this function is negative for positive x.

1. Substitute $x = 0$:

$$f(0) = 1 - 2^{-0} - 0 = 1 - 1 - 0 = 0$$

2. Take the derivative of $f(x)$:

$$f'(x) = \ln(2) \cdot 2^{-x} - 1$$

3. Analyze the behavior of the derivative:

$$\begin{aligned}
\text{For } x = 0, 2^{-x} &= 1 \\
\text{For } x > 0, 2^{-x} &< 1 \\
\text{This means that } \ln(2) \cdot 2^{-x} &< \ln(2) \text{for all } x > 0 \\
\text{Hence f(x)= } \ln(2) \cdot 2^{-x} - 1 &< 0 \text{ for all } x > 0.
\end{aligned}$$

Since $f(x)$ is continuous and decreasing for $x > 0$, $f(x) < 0$ for all x. Therefore, $1 - 2^{-x} < x$ for positive x.

Hence

$$x_{\text{med}} = 1 - 2^{-\frac{1}{n}} \leq \frac{1}{n}$$

Hence for the edges the all the edges added in step 1

Probability that their weight is less than $\frac{2}{n}$ is more than $\frac{1}{2}$.

We choose $\frac{2}{n}$ since $\frac{n}{2}$th edges added in step 1 is minimum of $\frac{n}{2}$ unif$(0, 1)$ random variables, and for all $i \leq \frac{n}{2}$

$\frac{1}{n-i+1} < \frac{2}{n}$.

Using the result that if we toss a fair coin t times then the probability of getting less than $\frac{t}{4}$ heads is inverse exponential in t.

Let's prove it using Chernoff bound. Consider tossing a fair coin t times. Let X be the number of heads obtained in t tosses. Each toss is a Bernoulli trial with success probability $p = 0.5$. Thus, the expected value of X is $\mu = tp = \frac{t}{2}$.

We want to find the probability that X is less than half of its expected value, i.e., $X < \frac{t}{4}$.

Chernoff bounds provide an upper bound on the probability of such deviations. For any $\delta > 0$, Chernoff bound states:

$$P(X < (1 - \delta)\mu) \leq e^{-\frac{\delta^2 \mu}{2}}$$

Let's choose $\delta = \frac{1}{2}$ (half), since we're interested in $X < \frac{t}{4} = (1 - \frac{1}{2})\mu$.

Substituting $\delta = \frac{1}{2}$ and $\mu = \frac{t}{2}$:

$$P(X < \frac{t}{4}) \leq e^{-\frac{\left(\frac{1}{2}\right)^2 \cdot \frac{t}{2}}{2}} = e^{-\frac{t}{16}}$$

So, for t fair coin tosses, the probability of getting fewer than one-fourth heads is upper-bounded by $e^{-\frac{t}{16}}$, which indeed decreases exponentially with t.

This result implies that at least $\frac{n}{8}$ of the $\frac{n}{2}$ vertices added in Step 1 will have a weight less than $\frac{2}{n}$ with high probability, since the probability of having less than $\frac{n}{8}$ vertices taking a value less than $\frac{2}{n}$ is inverse exponential in n.

Let's find the expected weight of the first edge in Step 2.

Let $w_1, w_2, \ldots, w_{\frac{n}{2}}$ be the weight of the edges added in step 1. Then consider the following $\frac{n}{2}$ random variables $Y_1, Y_2, \ldots, Y_{\frac{n}{2}}$, where Y_i follows a $\text{Unif}(w_i, 1)$ distribution.

Let $Y_{\min} = \min(Y_1, Y_2, \ldots, Y_{\frac{n}{2}})$. Then for 1^{st} edge added in step 2, it will follow the same distribution as $Y_{\min}$.

Let m of the $\frac{n}{2}$ edges added in Step 1 be less than $\frac{2}{n}$. Let the indices of these m edges be denoted by $k_1, k_2, \ldots, k_m$. We have proved above that with high probability m is at least $\frac{n}{8}$.

Notice $\min(Y_1, Y_2, \ldots, Y_{\frac{n}{2}}) \leq \min(Y_{k_1}, Y_{k_2}, \ldots, Y_{k_m})$.

Since each of the w_{k_i} is less than $\frac{2}{n}$, now consider m independent random variables $T_1, T_2, \ldots, T_m$ following a $\text{Unif}(\frac{2}{n}, 1)$ distribution.

Intuitively, it is easy to see that

$$E\left[\min(Y_{k_1}, Y_{k_2}, \ldots, Y_{k_m})\right] \leq E\left[\min(T_1, T_2, \ldots, T_m)\right]$$

Since for all $c \in [0, 1]$

$$\text{Prob}(Y_{k_i} \leq c) \geq \text{Prob}(T_i \leq c)$$

We proved in our previous report when we consider n independent random variables $X_1, X_2, \ldots, X_n$ following $\text{Unif}(\alpha, 1)$ distribution. The expected value of the minimum among them is given by:

$$E[\min(X_1, X_2, \ldots, X_n)] = \frac{1 + n\alpha}{1 + n}$$

Using this result:

$$E[\min(T_1, T_2, \ldots, T_m)] = \frac{1 + m \cdot \left(\frac{2}{n}\right)}{1 + m}$$

Let w_{2_i} denote the ith edge in step 2. Since m is at least $\frac{n}{8}$ with high probability, so:

$$E[w_{2_1}] \leq \frac{(1 + \frac{n}{8}\left(\frac{2}{n}\right))}{1 + \frac{n}{8}} = \frac{\frac{5}{4}}{\frac{n}{8} + 1} \leq \frac{\frac{5}{4}}{\frac{n}{8}} = \frac{10}{n}$$

Since, for the ith vertex added in step 2, we are also considering the edge connecting this node and $i - 1$ edges added in step 2, which follows $\text{Unif}(0, 1)$ distribution, since it was not revealed until now:

$$E[w_{2_i}] \leq \frac{1 + \left(\frac{n}{8} + i - 1\right)\left(\frac{2}{n}\right)}{1 + \frac{n}{8} + i - 1 + 1} \leq \frac{1}{\frac{n}{8} + i - 1 + 1} + \frac{2}{n} \leq \frac{1}{\frac{n}{8} + i} + \frac{2}{n}$$

Total expected weight of Step 2 using Linearity of Expectation:

$$E[w_{2_1} + w_{2_2} + \ldots + w_{2_{\frac{n}{3}}}]$$

$$= E[w_{2_1}] + E[w_{2_2}] + \ldots + E[w_{2_{\frac{n}{3}}}]$$

$$\leq \sum_{i=1}^{\frac{n}{3}} \left(\frac{1}{\frac{n}{8} + i - 1 + 1} + \frac{2}{n} \right)$$

$$\leq \frac{1}{\frac{n}{8}} + \frac{1}{\frac{n}{8} + 1} + \cdots + \frac{1}{\frac{n}{8} + \frac{n}{3}} + \frac{2}{n} * \frac{n}{3}$$

$$\leq \log\left(\frac{11n}{24}\right) - \log\left(\frac{n}{8}\right) + 0.667$$

$$\leq \log(\frac{11}{3}) + 0.667$$

$$\approx 1.96$$

As we have already calculated the expected weight of Step 1 and Step 3 in our previous report, we will directly use the results.

$$E[\text{Total weight of edges in Step 1}] \leq \ln 2$$

$$E[\text{Total weight of edges in Step 3}] \leq \ln 3 - \ln 2$$

$$E[Weight of MST] \leq E[\text{Edges in Step 1}] + E[\text{Edges in Step 2}] + E[\text{Edges in Step 3}]$$

$$\leq (\ln 2) + (\ln 11 - \ln 3 + 0.667) + (\ln 3 - \ln 2)$$

$$\leq \ln 11 + 0.667$$

$$\approx 3.06456$$

Thus the expected weight of MST via Prim's algorithm is less than equal to 3.06456.