Homework #3

Due: 2025-5-8 23:59 | 6 Questions, 100 Pts

Name: Jiacong Fang, ID: 2200017849

Note: The total points of this homework is 10 + 15 + 5 + 30 + 10 + 30 = 100.

Question 1 (10') (Balanced Chatting Groups). There are n students, each is assigned one of the two language to speak, Chinese or English. A chatting group is a non-empty subset of n students. There are m different chatting groups. One student must speak the same language in all chatting groups he or she participates. Prove that there exists an assignment of the n student's languages, such that in all the m chatting groups, the difference of speaker numbers of the two languages is $\mathcal{O}\left(\sqrt{n \log m}\right)$. ◀

Answer. Denote the m different chatting groups as $G_1, G_2, \ldots, G_m$. For each student i, let X_i be

$$X_i = \begin{cases} 1 & \text{if student } i \text{ speaks Chinese} \\ -1 & \text{if student } i \text{ speaks English} \end{cases}$$

We randomly assign $X_i \in \{0, 1\}$ for each student i independently with equal probability. For each chatting group G_j, suppose $|G_j| = k_j$, let the difference of speaker numbers of the two languages in G_j as Y_j,

$$Y_j = \left| \sum_{i \in G_j} X_i \right| = |2Y_{c,j} - k_j|,$$

where $Y_{c,j}$ is the number of Chinese speakers in G_j. Notice that $Y_{c,j} \sim \text{Binomial}(k_j, 1/2)$, thus we have ($c$ is a constant waited to be determined)

$$\begin{aligned}
\Pr\left[\exists G_j, Y_j > c\sqrt{n \log m}\right] &\leq m \cdot \Pr\left[|2Y_{c,j} - k_j| > c\sqrt{n \log m}\right] \\
&= m \cdot \Pr\left[\left|Y_{c,j} - \frac{k_j}{2}\right| > \frac{c\sqrt{n \log m}}{2}\right] \\
\text{(Chernoff Bound)} \quad &\leq 2m \cdot \exp\left(-\frac{2}{n} \cdot \frac{c^2 n \log m}{4}\right) \\
\text{(For } c \text{ large enough)} \quad &= 2m^{1-c^2/2} < 1
\end{aligned}$$

Therefore, there exists an assignment of the n student's languages, such that in all the m chatting groups, the difference of speaker numbers of the two languages is $\mathcal{O}\left(\sqrt{n \log m}\right)$. **Q.E.D.** ◁

Question 2 (15') (Distinguish Sets by Intersections). Consider $k \le \frac{1.99n}{\log_2 n}$ and assume n is sufficiently large, prove that for any collection of subsets $S_1, \cdots, S_k \subseteq \{1, \cdots, n\}$, there exist two distinct subsets $X, Y \subseteq \{1, \cdots, n\}$ such that $|X \cap S_i| = |Y \cap S_i|$ for all $i \in \{1, \cdots, k\}$.

[Hint: Introduce proper randomization process when choosing X, and analyze the concentration of $|X \cap S_i|$.] ◀

Answer. Given any collection of subsets $S_1, \cdots, S_k \subseteq \{1, \cdots, n\}$, let $k_i = |S_i|$. We randomly choose $X \subseteq \{1, \cdots, n\}$ by including each element with probability $p = 1/2$. Given X and for any $i \in \{1, 2, \cdots, k\}$, denote $N_i = |X \cap S_i|$. Then N_i follows a binomial distribution, $N_i \sim \text{Binomial}(k_i, 1/2)$. Therefore, by the Chernoff bound, we have

$$\Pr\left[\left|N_i - \frac{k_i}{2}\right| \ge \varepsilon\right] \le 2\exp\left(-\frac{2\varepsilon^2}{k_i}\right)$$

where ε is a constant waited to be determined. Then by the union bound, we have

$$\begin{aligned}\Pr\left[\exists i \text{ s.t. } \left|N_i - \frac{k_i}{2}\right| \ge \varepsilon\right] &\le \sum_{i=1}^{k} \Pr\left[\left|N_i - \frac{k_i}{2}\right| \ge \varepsilon\right] \le 2\sum_{i=1}^{k} \exp\left(-\frac{2\varepsilon^2}{k_i}\right) \\ &\le 2k \cdot \exp\left(-\frac{2\varepsilon^2}{n}\right) := p(\varepsilon)\end{aligned}$$

We choose the X that satisfies $|N_i - k_i/2| < \varepsilon$ for all $i \in \{1, 2, \cdots, k\}$, then the number of such X is at least $N_{\text{valid}} := 2^n \cdot (1 - p(\varepsilon))$. Notice that the number of valid $v = (N_1, N_2, \cdots, N_k)$ is at most $N_{\text{vec}} := (2\varepsilon)^k$. Then our goal is to show that $N_{\text{valid}} > N_{\text{vec}}$, which implies the existence of two distinct satisfied X and Y.

Let $\varepsilon = \sqrt{n \ln(2k^2)}/\sqrt{2}$. Then we have

$$p(\varepsilon) \le 2k \cdot e^{-\ln\left(2k^2\right)} = \frac{1}{k}. \quad N_{\text{vec}} = \left(2n \ln\left(2k^2\right)\right)^{k/2}$$

Therefore, we only need to show that

$$\begin{aligned}2^n \cdot \frac{k-1}{k} < \left(2n \ln\left(2k^2\right)\right)^{k/2} &\iff \frac{k}{2} \log_2\left[2n \ln\left(2k^2\right)\right] < n + \log_2 \frac{k-1}{k} \\ &\iff k < \frac{2n + 2\log_2[(k-1)/k]}{\log_2\left[2n \ln(2k^2)\right]}\end{aligned}$$

Notice that $k \le \frac{1.99n}{\log_2 n}$, so we need to show that

$$\frac{1.99n}{\log_2 n} \le \frac{2n + 2\log_2[(k-1)/k]}{\log_2\left[2n \ln(2k^2)\right]} = \frac{2n + 2\log_2[(k-1)/k]}{\log_2 n + 1 + \log_2[\ln(2k^2)]} \tag{1}$$

which is equivalent to

$$0.01n \log_2 n \ge 1.99n + 1.99n \cdot \log_2\left[\ln\left(2k^2\right)\right] + 2\log_2 n \cdot \log_2\left[\frac{k}{k-1}\right] \tag{2}$$

Since we have $2\log_2 n \cdot \log_2\left[\frac{k}{k-1}\right] = \mathcal{O}(\log n \cdot \log k) = \mathcal{O}\left(\log^2 n\right)$, $0.01n \log_2 n = \mathcal{O}(n \log n)$, and

$$1.99n \cdot \log_2\left[\ln\left(2k^2\right)\right] = \mathcal{O}(n \log(\log k)) = \mathcal{O}(n \log(\log n))$$

Therefore, for sufficiently large n, Eq. (2) holds, i.e., $N_{\text{valid}} > N_{\text{vec}}$, which implies the existence of two distinct X and Y that satisfy the requirements. **Q.E.D.** ◁

Question 3 (5') (Problematic Proof in Balls in Bins). Consider the following balls–and–bins problem: we throw n balls in n bins, and want to know how many bins contain at least 2 balls. Alice proposes the following solution, which is problematic. Please point out what are the problem(s) with Alice's proof.

`Let indicator variable` $X_i = 1$ `if there are at least` 2 `balls in bin` i, `and let` $X = \sum X_i$ `be the number of bins with at least two balls. Consider the probability of there being at least two balls in bin` i,

$$\Pr[X_i = 1] = \binom{n}{2} \cdot \left(\frac{1}{n}\right)^2 \geq \frac{1}{4}$$

`Thus, we have`

$$\mu = \mathbb{E}[\mathbb{X}] \geq \frac{n}{4}$$

`We now use a Chernoff Bound to prove that there are at least` $\frac{n}{8}$ `bins that contain` 2 `balls, with high probability:`

$$\begin{aligned} \Pr\left[X \leq \frac{n}{8}\right] &\leq \Pr\left[X \leq \frac{1}{2}\mu\right] \\ &\leq \exp\left(-\frac{\mu}{8}\right) \leq \frac{1}{n} \end{aligned}$$

◀

Answer. The Chernoff Bound requires the random variables to be independent, but in this case, the indicator variables X_i are not independent. Therefore, Alice cannot use Chernoff Bound on $\sum_{i=1}^{n} X_i$ directly which is the main problem. **Q.E.D.** ◁

Question 4 (30') (Random Graphs Proof of Power of Two Choices). In this problem, we will show another way to prove that the maximium load is $O(\log\log n)$, though our argument cannot provide the coefficient.

There are $\frac{n}{512}$ balls and n bins. We implement the two-choice procedure and record the procedure using a graph with n nodes:

Algorithm 1 Power of Two Choices

Output: Load, G

```
G ← (V, ∅)
∀i, load(i) ← 0
for each ball do
    Randomly pick two bins u ≠ v
    Add edge (u, v) to G
    if load(u) ≤ load(v) then
        load(u) ← load(u) + 1                ▷ throw the ball into bin u
    else
        load(v) ← load(v) + 1                ▷ throw the ball into bin v
    end if
end for
```

We use the graph G to analyze the loads.

a. (10') Show that there exists a constant $K > 0$ such that, for all subsets S of the vertex set with $|S| \geq K$, the induced graph $G[S]$ contains at most $5|S|/2$ edges, and hence has average degree at most 5, w.h.p.
   [Hint: You may find this useful: $\binom{n}{d} \leq \left(\frac{ne}{d}\right)^d$.]
   [Hint: When you attempt to bound a sum, try to break it into two parts and bound them separately.]

b. (10') Given graph G, we recursively remove all vertices of degree ≤ 10 in G until there are no more vertices of degree ≤ 10. Prove that this procedure ends after $O(\log\log n)$ rounds w.h.p., and the number of remaining vertices in each remaining component is at most K w.h.p.

c. (10') $\forall i$, if a node u survives i rounds, show that $\mathrm{load}(u) \leq 10i$. If a node u is never deleted, show that $\mathrm{load}(u) \leq 10i^* + cK$ w.h.p., where i^* is the total number of rounds and c is a constant.
   [Hint: Use Induction.]

(b) shows w.h.p. that $i^* = O(\log\log n)$. Thus, if we throw $n/512$ balls into n bins using the best-of-two-bins method, then w.h.p. the maximum load of any bin is $O(\log\log n)$. Hence for the case of n balls and n bins, the maximum load would be at most $512 * O(\log\log n) = O(\log\log n)$.

◀

Answer. a). For a fixed set S_0 with k nodes, we have (denote $m = n/512$):

$$\Pr\left[S_0 \text{ has } > \frac{5k}{2} \text{ edges}\right] \leq \binom{m}{5k/2}\left(\frac{k}{n}\right)^{2.5k\times 2} = \binom{m}{5k/2}\left(\frac{k}{n}\right)^{5k}$$

since each edge has 2 endpoints, and the probability of each vertex being selected into S is k/n while the choices are independent. Then by the union bound, we have:

$$\begin{aligned}
\Pr\left[\exists S \text{ that contains } > \frac{5|S|}{2} \text{ edges}\right] &\leq \sum_{k\geq K} \binom{n}{k}\binom{m}{5k/2}\left(\frac{k}{n}\right)^{5k} \\
\text{(by hint 1)} &\leq \sum_{k\leq K} \left(\frac{ne}{k}\right)^{k}\left(\frac{ne}{512(5k/2)}\right)^{5k/2}\left(\frac{k}{n}\right)^{5k} \\
&= \sum_{k\leq K} \left(\frac{k}{n}\right)^{3k/2} \frac{e^{7k/2}}{1280^{5k/2}} \\
&= \sum_{k\leq K} \left(\frac{k}{n}\right)^{3k/2} \left(\frac{e^{3.5}}{1280^{2.5}}\right)^{k}.
\end{aligned}$$

Let $a = \frac{e^{3.5}}{1280^{2.5}} < 1/2$, then for large enough k, a^k will be exponentially small, and for small k, we can bound the sum by the other terms, i.e. $(k/n)^{3k/2}$. Therefore, we bound the sum by dividing it into two parts:

$$\begin{aligned}
\Pr\left[\exists S \text{ that contains } > \frac{5|S|}{2} \text{ edges}\right] &\leq \sum_{k=K}^{2\log_2 n} \left(\frac{k}{n}\right)^{3k/2} a^k + \sum_{k=2\log_2 n}^{n} \left(\frac{k}{n}\right)^{3k/2} a^k \\
&\leq \sum_{k=K}^{2\log_2 n} \left(\frac{k}{n}\right)^{3k/2} + \sum_{k=2\log_2 n}^{n} a^k \\
&\leq 2\log_2 n \cdot \left(\frac{2\log_2 n}{n}\right)^{3K/2} + \sum_{k=2\log_2 n}^{n} \frac{1}{n^2} \\
&\leq \mathcal{O}\left(\frac{1}{n}\right)
\end{aligned}$$

The last inequality holds when we choose K to be a large enough constant. Therefore, we have shown that the statement holds with high probability.

b). We first prove that the size of G's largest connected component is $O(\log n)$ with high probability. Consider $k = \Theta(\log n)$, then we have:

$$\begin{aligned}
\Pr[k+1 \text{ vertices are connected}] &\leq \Pr[\text{at least } k \text{ edges within the } k+1 \text{ vertices}] \\
&\leq \binom{m}{k}\left(\frac{\binom{k+1}{2}}{\binom{n}{2}}\right)^{k} = \binom{m}{k}\left(\frac{k(k+1)}{n(n-1)}\right)^{k} \\
(n(n-1) \geq n^2/2, k(k+1) \leq 2k^2) \quad &\leq \binom{m}{k}\left(\frac{2k^2}{n^2/2}\right)^{k} = \binom{m}{k}\left(\frac{4k^2}{n^2}\right)^{k}
\end{aligned}$$

Therefore, by the union bound, we have:

$$\begin{aligned}
\Pr[\exists \text{ a connected component of size } k+1] &\leq \binom{n}{k+1}\Pr[k+1 \text{ vertices are connected}] \\
&\leq \binom{n}{k+1}\binom{m}{k}\left(\frac{4k^2}{n^2}\right)^{k} \\
\left(\binom{n}{d} \leq \left(\frac{ne}{d}\right)^{d}\right) \quad &\leq \left(\frac{ne}{k+1}\right)^{k+1}\left(\frac{ne}{512k}\right)^{k}\left(\frac{4k^2}{n^2}\right)^{k}
\end{aligned}$$

$$\begin{aligned}
\Pr\left[\exists \text{ a connected component of size } k+1\right] &\le \binom{m}{k}\left(\frac{2k^2}{n^2/2}\right)^k = \binom{m}{k}\left(\frac{4k^2}{n^2}\right)^k \\
&= \frac{en}{k+1}\left(\frac{k}{k+1}\right)^k\left(\frac{e^2}{128}\right)^k \\
(k=\Theta(\log n)) \quad &\le \frac{en}{k+1}\left(\frac{e^2}{128}\right)^k \le \mathcal{O}\left(\frac{1}{n}\right).
\end{aligned}$$

Therefore, we have shown that the size of G's largest connected component is $O(\log n)$ with high probability.

When there exists a component with more than K vertices, by the result of part (a), we know that the average degree of the component is at most 5. Then by Markov's inequality, for any vertex u in the component, we have:

$$\Pr\left[\deg(u) \ge 11\right] \le \frac{\mathbb{E}_u\left[\deg(u)\right]}{11} \le \frac{5}{11} < \frac{1}{2}.$$

Therefore, in expectation, we have at least $K/2$ vertices with degree at most 10 and will be deleted in current round.

Combining these two results, we observe that in each round, the number of vertices in any component with size larger than K is reduced by at least half in expectation. Since the size of the largest component is initially at most $O(\log n)$, the number of rounds is at most $O(\log\log n)$ with high probability. And when the procedure ends, the number of remaining vertices in each component is at most K with high probability.

c). We prove it by induction on i. For the base case, if a node u is deleted in the $i=1$ round, then $\deg(u) \le 10$, so the load of u is at most 10. The statement holds for $i=1$.

Suppose the statement holds for $i \in \{1, 2, \cdots, k-1\}(k \le i^*)$, now consider any vertex deleted in the k-th round. Denote the set of vertices that are deleted in the first $k-1$ rounds as S_{k-1}, then by the induction hypothesis, we have:

$$\forall s \in S_{k-1}, \quad \text{load}(s) \le 10(k-1).$$

Since each ball is thrown into the bin with the smaller load, the load of u that relates to S_{k-1} is at most $10(k-1)$. And u is deleted in the k-th round which means that $\deg(u) \le 10$, so the total laod of u is at most $10k$. Therefore, the statement holds for $i=k$.

Now we consider the case when u is never deleted, denote the set of vertices that are deleted in the first i^* rounds as S_{i^*}. Similar to the previous case, we know that the load of u that relates to S_{i^*} is at most $10i^*$. Besides, due to the result of part (b), we know that the size of the largest component is at most K with high probability. Therefore, the load of u is at most $10i^* + cK$ with high probability.

Combining these two results, we have shown that $\forall i$, if a node u survives i rounds, then $\text{load}(u) \le 10i$. If a node u is never deleted, then $\text{load}(u) \le 10i^* + cK$ with high probability.

Q.E.D. $\triangleleft$

Question 5 (10') (Binomial Branching Process). Consider a binomial Galton-Watson branching process. At each step, every node from the previous step gives rise to $X \sim \mathrm{Bin}(m, p)$ children independently. Denote the number of newly born nodes at time t by Z_t, with $Z_0 = 1$ being the initial node. The process does not die out if $Z_t > 0$ for all $t \geq 1$. Let

$$y(m, p) := \Pr[Z_t > 0, \forall t \geq 1].$$

a. (5') Give an implicit equation for $y(m, p)$. For example, $u(x) = x(1+u(x))^{u(x)}$ is an implicit equation. Prove your result.

b. (5') With $m \geq 2$ fixed and $mp = 1 + \varepsilon, \varepsilon > 0$, compute

$$\lim_{\epsilon \to 0^+} \frac{y(m, p)}{\varepsilon}.$$

You do not need to prove the existence of the limit.

◀

Answer. a). In the lecture notes, we already know that the extinction probability q^* is the solution of $f(x) = x$, where $f(x)$ is the probability generating function of X, i.e.

$$f(x) = \sum_{i \geq 0} \Pr[X = i] x^i = \sum_{i \geq 0} \binom{m}{i} p^i (1-p)^{m-i} x^i = (1 - p + px)^m.$$

Notice that $y(m, p)$ is actually the probability that the process does not die out, *i.e.*, $1 - q^*$. Therefore, we have:

$$q^* = (1 - p + p \cdot q*)^m \implies y(m, p) = 1 - (1 - p \cdot y(m, p))^m. \tag{3}$$

b). Let $y(m, p) := k\varepsilon, p = (1 + \varepsilon)/m$, then we have:

$$1 - k\varepsilon = \left[1 - \left(\frac{k\varepsilon}{m} + \frac{k\varepsilon^2}{m}\right)\right]^m$$

By Taylor expansion, we have:

$$\left[1 - \left(\frac{k\varepsilon}{m} + \frac{k\varepsilon^2}{m}\right)\right]^m = 1 - m\left(\frac{k\varepsilon + k\varepsilon^2}{m}\right) + \frac{m(m-1)}{2}\left(\frac{k\varepsilon + k\varepsilon^2}{m}\right)^2 + o(\varepsilon^2).$$

When $\varepsilon \to 0^+$, we have:

$$1 - k\varepsilon = 1 - k\varepsilon(1 + \varepsilon) + \frac{m-1}{2m} k^2 (\varepsilon + \varepsilon^2)^2,$$

$$\frac{m-1}{2m} k^2 \varepsilon^2 (1 + \varepsilon)^2 - k\varepsilon^2 = 0 \implies \frac{m-1}{2m} k(1 + \varepsilon)^2 - 1 = 0 \implies k = \frac{2m}{(m-1)}.$$

Therefore, we have:

$$\lim_{\varepsilon \to 0^+} \frac{y(m, p)}{\varepsilon} = \frac{2m}{(m-1)}.$$

Q.E.D. ◁

Question 6 (30') (Offline Balls and Bins). In the balls and bins model we have discussed in class, the balls are given *online*. What will happen if they are given *offline*?

a. (8') There are m balls and n bins. We are given graph $G = (V, E)$ with n vertices and m edges. Each vertex v correponds to a bin, and each edge e correponds to the two choices of a ball. Each ball can only be allocated to one of the two choices. Prove that the minimal max load

$$\text{Load(G)} = \max_{S \subseteq V} \left\lceil \frac{|E(S)|}{|S|} \right\rceil$$

where $\lceil \cdot \rceil$ denotes the ceiling functions, S is any subset of V and $E(S)$ consists of all the edges in E that have both endpoints in S. Note that G may not be simple; that is, there can be multiple edges, meaning that two or more balls must be placed in the same pair of bins.

[Hint: The $\geq$ side is easy. To prove the equality, formulate the minimal max load problem either as a linear programming, perfect matching or maximum flow problem. You can consider its duality to derive the result.]

[Hint: This result is for (b) only and the proof does not involve any randomization arguments. Skip it if you have a tight schedule or have no idea.]

b. (12') Now we consider n balls and n bins; that is, let $m = n$. Show that if the two choices are picked randomly, the minimal max load is at most 2 with high probability.

[Hint: Bound the probability that a subgraph contains too many edges. To show the sum is $o(1)$, you may divide its size into three intervals: $[2, n/6]$, $[n/6, 5n/12]$ and $[5n/12, n/2]$. What are the dominating coefficients in each interval?]

[Hint: Given any $\alpha \in [1/6, 5/6]$ and n sufficiently large (independent of α), there is $\binom{n}{\alpha n} < 2^{H(\alpha)n}$, where $H(\alpha) = -\alpha \log \alpha - (1-\alpha)\log(1-\alpha)$ (log to base 2).]

[Hint: You may find the following numerical results useful:

$$\text{(i) } 2^{H(\alpha)+H(1/2)+4\alpha\log\alpha} < 0.96 \text{ for any } \alpha \in [1/6, 5/12].$$

$$\text{(ii) } 2^{H(a)+H(5/6)+4\alpha\log\alpha} < 0.8 \text{ for any } \alpha \in [5/12, 1/2].$$

Be careful when simplifying the binomial coefficients.]

c. (10') Continuing with n balls and n bins, show that the minimal max load is not 1 with high probability. Hence, we conclude that the load is almost surely 2.

You can use a known result: If $m = cn/2, c > 1$, then w.h.p. random graph G_m consists of a unique giant component, with $(1 - \frac{x}{c} + o(1))n$ vertices and $(1 - \frac{x^2}{c^2} + o(1))\frac{cn}{2}$ edges. Here $0 < x < 1$ is the solution of the equation $xe^{-x} = ce^{-c}$. The remaining components are of order at most $\mathcal{O}(\log n)$.

◀

Answer. a). We first prove the $\geq$ side, denote $\text{Load}(G) = k^*$. Then $\forall S \subseteq V$, we have:

$$|E(S)| \leq \sum_{v \in S} \text{load}(v) \leq \sum_{v \in S} k^* = k^*|S| \implies \frac{|E(S)|}{|S|} \leq k^* = \text{Load}(G). \tag{4}$$

For the other side, we construct a network flow model $G' = (V', E')$ as follows:

- Add a source node s and a sink node t.
- For each $e \in E$, add a node x_e and add an edge (s, x_e) with capacity 1.
- For each $v \in V$, add a node x_v and add an edge (x_v, t) with capacity L.
- For each edge $(u, v) \in E$, add edges (x_e, x_u) and (x_e, x_v) with capacity 1.

where

$$L =: \max_{S \subseteq V} \left\lceil \frac{|E(S)|}{|S|} \right\rceil$$

Then the max flow of G' is $|E|$ if and only if there is a valid assignment of edeges to vertices with max load at most L.

And the Max-Flow Min-Cut Theorem states that the maximum flow in a flow network is equal to the capacity of the minimum s-t cut. Therefore, we only need to prove that the minimum s-t cut of G' is at most $|E|$. Suppose we devide V' into A and B, and let $s \in A$ and $t \in B$. For any subset $S \subseteq V$, we select A and B as follows:

- $A = \{s\} \cup \{x_v : \forall v \in S\} \cup \{x_e : e = (u, v) \wedge (u \in S \vee v \in S)\}$.
- $B = V' \setminus A$.

Notice that for any $x_e \in B$, if we add x_e to A, then the capacity of the cut will be increased by 1. For any $e = (u, v), x_e \in A$ and $x_u, x_v \in A$, if we add x_e to B, then the capacity of the cut will be increased by 3. For any $e = (u, v), x_e \in A$ and x_u or $x_v \in B$, if we add x_e to B, then the capacity of the cut will increased by 1. Therefore, the selected cut is the minimal one given the subset S, and we have:

$$\begin{aligned}\text{capacity}(A, B) &= L * |S| + 2|\{e : x_e \in B\}| + |\{(u, v) : x_{(u,v)} \in A \text{ and only 1 endpoint in } A\}| \\ &\geq L * |S| + |E \setminus E(S)| = L * |S| + |E| - |E(S)|\end{aligned}$$

And we have:

$$L = \max_{S \subseteq V} \left\lceil \frac{|E(S)|}{|S|} \right\rceil \implies L \geq \frac{|E(S)|}{|S|}, \forall S \subseteq V$$

Therefore, for all subset $S \subseteq V$, we have:

$$\text{capacity}(A, B) \geq L * |S| + |E| - |E(S)| \geq |E|.$$

which implies that the capacity of any s-t cut is as least $|E|$. Since the max flow of G' is equal to the minimal cut, the max flow is at least $|E|$. Due to any feasible flow of G' is at most $|E|$, the max flow of G' is exactly $|E|$.

Therefore, there exists a valid assignment of edges to vertices with max load at most L, *i.e.,*

$$\text{Load}(G) \leq L = \max_{S \subseteq V} \left\lceil \frac{|E(S)|}{|S|} \right\rceil \tag{5}$$

Combine with the result in Eqs. (4) and (5), we have:

$$\text{Load(G)} = \max_{S \subseteq V} \left\lceil \frac{|E(S)|}{|S|} \right\rceil$$

Note: We can also use Lagrange duality, or perfect matching with Hall's Marriage Theorem (**Lemma 1**) to prove the result.

b). Using the result in part (a), we only need to prove that with high probability, any subset $S \subseteq V$ satisfies $|E(S)| \le 2|S|$. We try to bund the probability that there exists a subset S such that $|E(S)| > 2|S|$. For a fixed set A of t nodes, we have:

$$\Pr[E(A) > 2t] \le \binom{m}{2t}\left(\frac{t}{n}\right)^{2t\times 2} = \binom{n}{2t}\left(\frac{t}{n}\right)^{4t}$$

since each edge has 2 endpoints, and the probability of each vertex being selected into S is t/n while the choices are independent. Notice that when $t > n/2$, $\Pr[E(A) > 2t] = 0$. Then by the union bound, we have:

$$\Pr[\exists \text{ a bad set}] \le \sum_{t=2}^{n/2}\binom{n}{t}\binom{n}{2t}\left(\frac{t}{n}\right)^{4t}$$

Follow the hint, we can bound the sum by dividing it into three parts.

- For $t \in [5n/12, n/2]$, $\alpha := t/n \in [5/12, 1/2]$, then we have:

$$\begin{aligned}\sum_{t=5n/12}^{n/2}\binom{n}{t}\binom{n}{2t}\left(\frac{t}{n}\right)^{4t} &\le \sum_{t=5n/12}^{n/2}\binom{n}{\alpha n}\binom{n}{5n/6}\alpha^{4\alpha n}\\ &\le 2^{H(\alpha)n + H(5/6)n + 4\alpha n\log\alpha}\\ &\le \frac{n}{12}\cdot 0.8^n\end{aligned}$$

- For $t \in [n/6, 5n/12)$, $\alpha := t/n \in [1/6, 5/12)$, then we have:

$$\begin{aligned}\sum_{t=n/6}^{5n/12}\binom{n}{t}\binom{n}{2t}\left(\frac{t}{n}\right)^{4t} &\le \sum_{t=n/6}^{5n/12}\binom{n}{\alpha n}\binom{n}{n/2}\alpha^{4\alpha n}\\ &\le 2^{H(\alpha)n + H(1/2)n + 4\alpha n\log\alpha}\\ &\le \frac{n}{4}\cdot 0.96^n\end{aligned}$$

- For $t \in [2, n/6)$, then we have:

$$\begin{aligned}\sum_{t=2}^{n/6}\binom{n}{t}\binom{n}{2t}\left(\frac{t}{n}\right)^{4t} &\le \sum_{t=2}^{n/6}\left(\frac{ne}{t}\right)^{t}\left(\frac{ne}{2t}\right)^{2t}\left(\frac{t}{n}\right)^{4t}\\ &= \sum_{t=2}^{n/6}\left(\frac{e^3 t}{4n}\right)^{t}\end{aligned}$$

  Notice that when $t = n/6$, $e^3t/(4n) = e^3/24 < 1$. Therefore, we can bound the sum by dividing it into two parts, $[1, c\sqrt{n})$ and $[c\sqrt{n}, n/6)$, where c is a constant waited to be determined. In details,

$$\begin{aligned}
\sum_{t=2}^{n/6}\left(\frac{e^3 t}{4n}\right)^t &\le \sum_{t=2}^{c\sqrt{n}}\left(\frac{e^3 t}{4n}\right)^t + \sum_{t=c\sqrt{n}}^{n/6}\left(\frac{e^3 t}{4n}\right)^t \\
&\le \sum_{t=2}^{c\sqrt{n}}\left(\frac{e^3\cdot c\sqrt{n}}{4n}\right)^t + \sum_{t=c\sqrt{n}}^{n/6}\left(\frac{e^3\cdot n/6}{4n}\right)^t \\
&\le \sum_{t=2}^{c\sqrt{n}}\left(\frac{e^3\cdot c\sqrt{n}}{4n}\right)^2 + \sum_{t=c\sqrt{n}}^{n/6}\left(\frac{e^3\cdot n/6}{4n}\right)^{c\sqrt{n}} \\
&\le c\sqrt{n}\cdot\frac{e^6c^2}{16n} + \frac{n}{6}\cdot\left(\frac{e^3}{24}\right)^{c\sqrt{n}} \\
(c \text{ is a constant}) \quad &= \mathcal{O}\left(\frac{1}{\sqrt{n}}\right).
\end{aligned}$$

Combining these three parts, we have:

$$\begin{aligned}
\Pr\left[\exists \text{ a bad set}\right] &\le \sum_{t=2}^{n/2}\binom{n}{t}\binom{n}{2t}\left(\frac{t}{n}\right)^{4t} \\
&\le \frac{n}{12}\cdot 0.8^n + \frac{n}{4}\cdot 0.96^n + \mathcal{O}\left(\frac{1}{\sqrt{n}}\right) = o(1).
\end{aligned}$$

for sufficiently large n. Therefore, we have shown that the minimal max load is at most 2 with high probability.

c). According to the known result, w.h.p there is a unique giant component S such that $|S| = (1-\frac{x}{c}+o(1))n$ and $|E(S)| = (1-\frac{x^2}{c^2}+o(1))\frac{cn}{2}$. Then we have:

$$\begin{aligned}
\frac{|E(S)|}{|S|} &= \frac{(1-x^2/c^2+o(1))\cdot cn/2}{(1-x/c+o(1))\cdot n} = \frac{c}{2}\cdot\frac{1-x^2/c^2+o(1)}{1-x/c+o(1)} \\
(\text{Let } c=2) \quad &= 1+\frac{x}{2}+o(1) \\
&> 1.2+o(1) > 1. \quad \hookleftarrow x\in(0.4,0.42) \text{ for } xe^{-x} = 2e^{-2}.
\end{aligned}$$

Therefore, combining with the result in part (a). , with high probability, the minimal max load is not 1.

Lemma 1. ***(Hall's Marriage Theorem)*** *For a bipartite graph G on the parts X and Y, then following conditions are equivalent:*

(i) There is a perfect matching of X into Y.

(ii) For each subset $T \subseteq X$, the inequality $|T| \le |N_G(T)|$ holds.

where $N_G(T)$ denote the set of vertices of G that are adjacent to some vertex in T, that is,

$$N_G(T) := \{v \in V(G) \mid (v,w) \in E(G) \text{ for some } w \in T\}.$$

The proof of this theorem can be found in the Course Notes of MIT, ignoring the details here.

Q.E.D. ◁