PROBLEM 1

a) The probability of failure for the fingerprinting algorithm, when the prime number for fingerprinting is chosen uniformly at random between 2 and T, and the numbers being compared are at most $n$bit numbers is,

$$\Pr[\text{Failure}] \leq \frac{n \log T}{T}.$$

In the algorithm there are $n - m + 1$ iterations. If there is a false match in any one of the iterations, the output will be false.

$$\Pr[\text{Error}] = (n - m + 1) \cdot \frac{m \log T}{T}.$$

Using the assumption that $n \ll m$ as n grows, we can approximate this as,

$$\Pr[\text{Error}] \leq \frac{mn \log T}{T}.$$

Now, substituting $T = mn^2$ we can write,

$$\Pr[\text{Error}] \leq \frac{mn \log(mn^2)}{mn^2}.$$

$$\Pr[\text{Error}] = \frac{\log(m) + 2\log(n)}{n} = O\left(\frac{\log(n)}{n}\right).$$

As n grows, $\Pr[\text{Error}] \to 0$ because $n \gg \log(n)$ as n grows.

b) The computation in step 2 is performed only once and requires $O(m)$ time. The iteration in step 3 runs $n - m + 1$ times. Each iteration takes $O(m)$. If is for calculating the number from the binary string . The total running time is,

$$O(m) \cdot (n - m + 2) = O(m) \cdot O(n) = O(mn).$$

c) We can write,

$$X(j) = 2^{m-1}x_j + 2^{m-2}x_{j+1} + \cdots + 2x_{j+m-2} + x_{j+m-1}.$$

$$X(j+1) = 2^{m-1}x_{j+1} + \cdots + 4x_{j+m-2} + 2x_{j+m-1} + x_{j+m} = 2\left(2^{m-2}x_{j+1} + \cdots + 2x_{j+m-2} + x_{j+m-1}\right) + x_{j+m}.$$

Substituting for $X(j)$ we get,

$$X(j+1) = 2X(j) - 2^{m-1}x_j + x_{j+m}.$$

We have, $X(j+1) \mod p = (2X(j) \mod p) - (2^{m-1}x_j \mod p) + (x_{j+m} \mod p)$. and, $F(X(j+1), p) = F(X(j), p) - (2^{m-1}x_j \mod p) + (x_{j+m} \mod p)$.

d) We eliminate the $O(m)$ time complexity per iteration since $F(X(j+1), p)$ can now be computed in constant time with simple modulo operations. Consequently, the time complexity of step 3 is reduced to $O(n)$. When combined with the $O(m)$ time required for step 2, the total runtime of the optimized algorithm becomes $O(m + n)$.

PROBLEM 2

We will prove it using induction. Let Y_n be the probability that X_n is even. We are going to to show that $Y_n = \frac{1}{2}$ for all $n \geq 1$.

When $n = 1$, $Y_1 = \frac{1}{2}$ as X is a Binomial random variable with $(1, \frac{1}{2})$. so $P(X = 0) = \frac{1}{2}$ and $P(X = 1) = \frac{1}{2}$. This means that the probability of X being even $(X = 0)$ is $\frac{1}{2}$, so, $Y_1 = \frac{1}{2}$. Let $Y_k = \frac{1}{2}$ for some $k \geq 1$. We have to prove that $Y_{k+1} = \frac{1}{2}$. Z_{k+1} is the probability that the $(k+1)$ th Bernoulli trial is 1. To find $P(Y_{k+1})$, we consider two scenarios. They are X_k is even, $Z_{k+1} = 0$ and X_k is odd, $Z_{k+1} = 1$.

So, we can write,

$$P[Y_{k+1}] = P[Y_k] \cdot P[Z_{k+1} = 0] + (1 - P[Y_k]) \cdot P[Z_{k+1} = 1]$$

putting the value, $P[Y_k] = \frac{1}{2}$, $P[Z_{k+1} = 0] = \frac{1}{2}$, and $P[Z_{k+1} = 1] = \frac{1}{2}$, we get the following,

$$P[Y_{k+1}] = \left(\frac{1}{2}\right) \cdot \left(\frac{1}{2}\right) + \left(\frac{1}{2}\right) \cdot \left(\frac{1}{2}\right) = \frac{1}{2}$$

So, $Y_{k+1} = \frac{1}{2}$. By induction, we can prove that the probability that X_n is even is $\frac{1}{2}$.

PROBLEM 3

Consider 2 states. State s_0 is the state where the last roll was not a six. State s_1 is the state where the last roll was a six. From s_0, if a six is rolled, then the new state is s_1 and probability $\frac{1}{6}$. From s_1, if six is not rolled then the state remains same with probability $\frac{5}{6}$. From s_1, if a six is rolled the process ends with probability $\frac{1}{6}$. Froms_1, if six is not rolled, the next state is s_0 and the probability is $\frac{5}{6}$. Now we consider, E_0 is the expected number of rolls to get consecutive sixes starting from State s_0 and E_1 is the Expected number of rolls to get consecutive sixes starting from State s_1.

$$E_0 = 1 + \frac{5}{6}E_0 + \frac{1}{6}E_1$$

we have to make a roll, so we need to start with 1. Then ,it can stay in s_0 or go to s_1.

$$E_1 = 1 + \frac{5}{6}E_0$$

Here, either the process ends or go back to the state s_0.

Solving the eqn, we get

$$E_0 = 1 + \frac{5}{6}E_0 + \frac{1}{6}\left(1 + \frac{5}{6}E_0\right)$$

$$E_0 = 1 + \frac{5}{6}E_0 + \frac{1}{6} + \frac{5}{36}E_0$$

$$E_0 = \frac{7}{6} + \frac{5}{6}E_0$$

$$\frac{1}{6}E_0 = \frac{7}{6}$$

$$E_0 = 42$$

The result is 42.

PROBLEM 4

p(n) is the probability that no two people have the same birthday. The probability of first person having unique birthday is $\frac{365}{365} = 1$. The second person should have a different birthday than the first, so the probability is $\frac{364}{365}$. For the third person it will be $\frac{363}{365}$.
So, we can write,

$$p(n) = 1 \times \left(1 - \frac{1}{365}\right) \times \left(1 - \frac{2}{365}\right) \times \cdots \times \left(1 - \frac{n-1}{365}\right)$$

Using the Taylor series expansion of $e^x = 1 + x + \frac{x^2}{2!} + \cdots$ and we can approximate $e^x \approx 1 + x$ for $|x| \ll 1$,so $1 - x \approx e^{-x}$

Thus, we can rewrite ,

$$p(n) \approx 1 \times e^{-\frac{1}{365}} \times e^{-\frac{2}{365}} \times \cdots \times e^{-\frac{n-1}{365}}$$

$$= e^{-\frac{1}{365}(1+2+\cdots+(n-1))}$$

$$p(n) \approx e^{-\frac{n(n-1)}{730}} \approx e^{-\frac{n^2}{730}}$$

Now, we want $1 - p(n) \geq 0.9$. This gives, $e^{-\frac{n^2}{730}} \leq 0.1$

we have,

$$-\frac{n^2}{730} \leq \ln(0.1)$$

$$n^2 \geq 730 \times \ln(0.1)$$

$$n^2 \geq 1680.88$$

$$n \geq \sqrt{1680.88} \approx 40.99 \approx 41$$

So, $n \geq 41$.

PROBLEM 5

The selection of $k = 10 \log n$ random elements takes $O(k \log n) = O(10 \log n \cdot \log n) = O((\log n)^2)$ time. Sorting the selected elements requires $O(k \log k)$ comparisons. Each comparison takes $O(\log k)$ time Therefore, the sorting step takes $O(k(\log k)^2) = O(\log n \cdot (\log \log n)^2) = O(\log n \cdot (\log n)^2) = O((\log n)^3)$.. Thus, the total time complexity is,

$$O((\log n)^2) + O((\log n)^3) = O((\log n)^3).$$

The algorithm fails if the median of the k chosen elements does not fall within the middle half of the array. There can be two cases.

If the returned median is in the left quarter, then all $k/2$ elements on the left of the median were selected from the left quarter of the array. The probability of choosing any element from the left quarter is $1/4$. The probability of selecting $k/2$ or more elements from the left quarter is,

$$P_L = \sum_{i=k/2}^{k} \binom{k}{i} \left(\frac{1}{4}\right)^i \left(\frac{3}{4}\right)^{k-i}.$$

Using a bound for the binomial coefficient and simplifying, we find,

$$P_L \leq \binom{k}{k/2} \left(\frac{3}{4}\right)^k \sum_{i=k/2}^{k} \left(\frac{1}{3}\right)^i.$$

$$P_L \leq \left(\frac{3}{4}\right) \cdot \left(\frac{1}{2}\right)^{k/5}.$$

Since $k = 10 \log n$, we can write,

$$P_L \leq \left(\frac{1}{2}\right)^{2 \log n} = \frac{1}{n^2}.$$

Similarly, if the median is in the right quarter, all $k/2$ elements on the right of the median must have been chosen from the right quarter. The probability of picking any element from the right quarter is also $1/4$. So, in this case,

$$P_R \leq \frac{1}{n^2}.$$

From both cases, the error probability P is,

$$P = P_L + P_R \leq \frac{2}{n^2} = O\left(\frac{1}{n^2}\right).$$

So, the probability of error is $O\left(\frac{1}{n^2}\right)$.

PROBLEM 6

a) If the initial list is sorted in descending order, and the pivot is always chosen as the first element, then in each iteration, the list is partitioned into two sub lists. one has only the pivot and the other has remaining elements. This ensures that each element is selected as the pivot once, and the i-th element will be compared against $(n - i + 1)$ other elements. So, the total number of comparisons is

$$(n-1) + (n-2) + (n-3) + \cdots + 1$$

Simplifying, we get

$$n(n-1) - (1 + 2 + \cdots + (n-1)) = \frac{n(n-1)}{2} = \Omega(n^2)$$

The total number of comparisons is $\Omega(n^2)$.

b) Let $y_1, y_2, \ldots, y_n$ be the elements sorted in increasing order. Let X represent the total number of comparisons when the pivot is chosen uniformly at random . $X_{i,j}$ is an indicator random variable, where $X_{i,j} = 1$ if y_i and y_j are compared during the execution of Quicksort, and $X_{i,j} = 0$ otherwise.

$$X = \sum_{i=1}^{n} \sum_{j=i+1}^{n} X_{i,j}$$

$$E[X] = E\left[\sum_{i=1}^{n} \sum_{j=i+1}^{n} X_{i,j}\right] = \sum_{i=1}^{n} \sum_{j=i+1}^{n} E[X_{i,j}]$$

The value of $X_{i,j}$ is 1 if either y_i or y_j is chosen as the pivot before any element in between them. This probability is equivalent to choosing either y_i or y_j as the pivot among the $j - i + 1$ elements $y_i, y_{i+1}, \ldots, y_j$, since pivots chosen outside this range do not affect $X_{i,j}$. Since pivots are chosen uniformly at random, the probability of selecting y_i or y_j as the pivot among these $j - i + 1$ elements is,

$$P(X_{i,j} = 1) = \frac{2}{j - i + 1}$$

So, $E[X_{i,j}] = \frac{2}{j-i+1}$, and we can write,

$$E[X] = \sum_{i=1}^{n} \sum_{j=i+1}^{n} E[X_{i,j}] = \sum_{i=1}^{n} \sum_{j=i+1}^{n} \frac{2}{j - i + 1} = \sum_{i=1}^{n} \sum_{k=1}^{n-i} \frac{2}{k+1}$$

As $\sum_{k=1}^{n-i} \frac{1}{k+1}$ is a harmonic sum, for sufficiently large n, this sum is $O(\log n)$. So,

$$E[X] = \sum_{i=1}^{n-1} O(\log n) = O(n \log n)$$

PROBLEM 7

Y is the number of boxes one needs to buy to claim the prize. Y_j is the number of days required to obtain one of the coupons in the jth new pair, having already collected at least one coupon from the $j-1$ pairs. Here, $P(Y_j = 0) = 0$ and $P(Y_j = 1) = 1 - \frac{j-1}{m}$ because we must avoid selecting any of the $2(j-1)$ coupons from the previously collected $j-1$ pairs, out of a total of $2m$ coupons.

$$P(Y_j = 2) = \frac{j-1}{m} \cdot \left(1 - \frac{j-1}{m}\right)$$

$$P(Y_j = 3) = \left(\frac{j-1}{m}\right)^2 \cdot \left(1 - \frac{j-1}{m}\right)$$

After generalizing,

$$P(Y_j = k) = \left(\frac{j-1}{m}\right)^{k-1} \cdot \left(1 - \frac{j-1}{m}\right),$$

Y_j is a geometric random variable with success probability $1 - \frac{j-1}{m} = \frac{m-j+1}{m}$.

$$E[Y_j] = \frac{m}{m-j+1}.$$

$$E[Y] = \sum_{j=1}^{m} E[Y_j] = \sum_{j=1}^{m} \frac{m}{m-j+1} = m \sum_{j=1}^{m} \frac{1}{j} = m \cdot \text{Harmonic}(m) = m \cdot O(\log m) = O(m \log m).$$

So, the expected number of boxes to buy is $O(m \log m)$

PROBLEM 8

1) X is the expected number of draws required to collect all n coupons and X_i is the number of draws required to get the ith new coupon. The probability p of collecting any particular coupon in a single draw is approximately the ratio of the number of subsets containing that coupon to the total number of subsets. So we can write,

$$p = \frac{\sum_{i=1}^{d} \binom{n-1}{i-1}}{\sum_{i=1}^{d} \binom{n}{i}} \approx \frac{\sum_{i=0}^{d} \binom{n-1}{i}}{\sum_{i=0}^{d} \binom{n}{i}}.$$

we know,

$$\frac{\binom{n-1}{d-1}}{\binom{n}{d}} \cdot \frac{n-d+1}{n-2d+1} \leq p \leq \frac{\binom{n-1}{d-1}}{\binom{n}{d}} \cdot \frac{n-d+1}{n-2d-1}.$$

we get,

$$\frac{d}{n} \cdot \frac{n-2d+1}{n-d+1} \leq p \leq \frac{d}{n} \cdot \frac{n-d+1}{n-2d-1}.$$

If $d = O(1)$, then $\frac{n-2d+1}{n-d+1} \to 1$ and $\frac{n-d+1}{n-2d-1} \to 1$ as n grows. From the above inequality,

$$p = \Theta\left(\frac{d}{n}\right).$$

Since the probability of collecting a new coupon among the $i-1$ already collected ones is approximately $(i-1) \cdot \Theta\left(\frac{d}{n}\right)$, X_i is a geometric random variable with success probability $1-(i-1) \cdot \Theta\left(\frac{d}{n}\right)$. Therefore,

$$E[X_i] = \frac{1}{1-(i-1) \cdot \Theta\left(\frac{d}{n}\right)}.$$

The total expected number of draws $E[X]$ is,

$$E[X] = \sum_{i=1}^{n} E[X_i] = \sum_{i=1}^{n} \frac{1}{1-(i-1) \cdot \Theta\left(\frac{d}{n}\right)}.$$

Using the approximation $E[X] = \Theta\left(\frac{n}{d}\right)\sum_{i=1}^{n-1}\frac{1}{i}$, we get,

$$E[X] = \Theta\left(\frac{n}{d}\right) \cdot O(\log n) = O\left(\frac{n}{d}\log n\right).$$

The expected number of draws is,

$$E[X] = O\left(\frac{n}{d}\log n\right).$$

So, we find that the expected number of draws is $O(n\log n)$.

2) With $d = O(\log n)$, we find that, $\frac{n-2d+1}{n-d+1} \to 1 \quad \text{and} \quad \frac{n-d+1}{n-2d-1} \to 1$. from the previous problem , we have, $p = \Theta\left(\frac{d}{n}\right)$. and the expected number of draws , $E[X] = O\left(\frac{n}{d}\log n\right)$.

using $d = O(\log n)$, we get,

$$E[X] = O\left(\frac{n}{\log n} \cdot \log n\right) = O(n).$$

3) let $d = \Theta(n)$. In this case , we have, $\frac{n-2d+1}{n-d+1} = \Theta\left(\frac{n}{n}\right) \to 1 \quad \text{and} \quad \frac{n-d+1}{n-2d-1} = \Theta\left(\frac{n}{n}\right) \to 1$. From the previous problem , we have, $p = \Theta\left(\frac{d}{n}\right)$.

and the expected number of draws is, $E[X] = O\left(\frac{n}{d}\log n\right)$.

Using $d = \Theta(n)$, we get,

$$E[X] = O\left(\frac{n}{n} \cdot \log n\right) = O(\log n).$$

PROBLEM 9

Let Y_1 is a binomial random variable. It represents the total number of balls that fall into bin 1. Here each ball has a probability of $\frac{1}{2}$ of falling into bin 1. So, $Y_1 \sim \text{Binomial}(2n, \frac{1}{2})$. Y_2 is the number of balls falling into bin 2. It contains all the remaining balls not falling into bin 1. So, $Y_2 = 2n - Y_1$. The difference between the number of balls in the two bins is

$$Y_D = |Y_1 - Y_2| = |Y_1 - (2n - Y_1)| = |2Y_1 - 2n| = 2|Y_1 - n|$$

Now, for a binomial random variable, we have, $E[Y_1] = 2n \cdot \frac{1}{2} = n$ $\text{Var}(Y_1) = 2n \cdot \frac{1}{2} \cdot \left(1 - \frac{1}{2}\right) = \frac{n}{2}$

We apply Chebyshev's inequality,

$$\Pr[|Y_1 - n| \geq k \cdot \frac{\sqrt{n}}{\sqrt{2}}] \leq \frac{1}{k^2}$$

$$\Pr[2|Y_1 - n| \geq 2k \cdot \frac{\sqrt{n}}{\sqrt{2}}] \leq \frac{1}{k^2}$$

Since $Y_1 < Y_2$ would imply $\Pr(Y_1 - Y_2 \geq c\sqrt{n}) = 0 \leq \epsilon$ for any $\epsilon > 0$, we can assume $Y_1 \geq Y_2$. So,

$$\Pr[Y_1 - Y_2 \geq 2k \cdot \frac{\sqrt{n}}{\sqrt{2}}] \leq \frac{1}{k^2}$$

$$\Pr[Y_1 - Y_2 \geq k\sqrt{2} \cdot \sqrt{n}] \leq \frac{1}{k^2}$$

Now, setting $c = k\sqrt{2}$, we get, $\Pr[Y_1 - Y_2 \geq c\sqrt{n}] \leq \frac{2}{c^2}$. For any $\epsilon > 0$, we can choose $c = \sqrt{\frac{2}{\epsilon}}$ to obtain,

$$\Pr[Y_1 - Y_2 \geq c\sqrt{n}] \leq \epsilon$$

PROBLEM 10

1)

$$X = \sum_{C \in \Omega_4} X_C,$$

Here Ω_4 is the set of all 4-vertex subsets in the graph. $X_C = 1$ if the vertices in subset form a clique, and 0 otherwise. As X_C is a Bernoulli random variable, it is 1 if all possible edges between the 4 vertices in C are present. There are $\binom{4}{2} = 6$ edges required for a 4-clique. Each existing independently with probability p. Therefore, $X_C = 1$ with probability p^6 and 0 with probability $1 - p^6$.

$$E[X] = \sum_{C \in \Omega_4} E[X_C] = \binom{n}{4} p^6.$$

$$E[X] = \frac{n(n-1)(n-2)(n-3) \cdot p^6}{24} \approx \frac{n^4 p^6}{24} + \Theta(n^3).$$

Now, consider the cases based on the behavior of $pn^{2/3}$ If $pn^{2/3} \to 0$, Then $n^4 p^6 \to 0$ (since $(pn^{2/3})^6 = n^4 p^6$), meaning $E[X] \to 0$. If $pn^{2/3} \to \infty$, Then $n^4 p^6 \to \infty$, and thus $E[X] \to \infty$.

These cases determine the expected count of 4-cliques in the graph.

2)

We use Markov's inequality. Here X is nonnegative, we set $a = 1$.

$$\Pr(G \text{ has a 4-clique}) = \Pr(X \geq 1) \leq \frac{E[X]}{1} = E[X]$$

If $pn^{2/3} \to 0$, then we get $E[X] \to 0$, consequently $\Pr(G \text{ has a 4-clique}) \to 0$.

3)

$$X^2 = \left(\sum_{C\in\Omega_4} X_C\right)^2 = \sum_{C\in\Omega_4} X_C^2 + \sum_{C,D\in\Omega_4:D\neq C} X_C X_D$$

$$E[X^2] = \sum_{C\in\Omega_4} E[X_C^2] + \sum_{C,D\in\Omega_4:D\neq C} E[X_C X_D]$$

$$(E[X])^2 = \left(\sum_{C\in\Omega_4} E[X_C]\right)^2 = \sum_{C\in\Omega_4} (E[X_C])^2 + \sum_{\substack{C,D\in\Omega_4\\ D\neq C}} E[X_C]E[X_D].$$

We can write,

$$\mathrm{Var}(X) = E[X^2] - (E[X])^2.$$

$$\mathrm{Var}(X) = \left(\sum_{C\in\Omega_4} E[X_C^2] + \sum_{\substack{C,D\in\Omega_4\\ D\neq C}} E[X_C X_D]\right) - \left(\sum_{C\in\Omega_4} (E[X_C])^2 + \sum_{\substack{C,D\in\Omega_4\\ D\neq C}} E[X_C]E[X_D]\right).$$

$$\mathrm{Var}(X) = \left(\sum_{C\in\Omega_4} E[X_C^2] - \sum_{C\in\Omega_4} (E[X_C])^2\right) + \left(\sum_{\substack{C,D\in\Omega_4\\ D\neq C}} E[X_C X_D] - \sum_{\substack{C,D\in\Omega_4\\ D\neq C}} E[X_C]E[X_D]\right).$$

$$\mathrm{Var}(X) = \sum_{C\in\Omega_4} \mathrm{Var}(X_C) + \sum_{\substack{C,D\in\Omega_4\\ D\neq C}} \mathrm{Cov}(X_C, X_D).$$

4) We have

$$\sum_{C\in\Omega_4} \mathrm{Var}(X_C) = \sum_{C\in\Omega_4} \left(\mathbb{E}[X_C^2] - (\mathbb{E}[X_C])^2\right).$$

X_C is a Bernoulli random variable, we know that $Y_C = X_C^2$ and $\mathbb{E}[X_C^2] = \mathbb{E}[X_C] = q^6$

$$\begin{aligned}\sum_{C\in\Omega_4} \mathrm{Var}(X_C) &= \sum_{C\in\Omega_4} \left(\mathbb{E}[X_C^2] - (\mathbb{E}[X_C])^2\right) \\ &= \sum_{C\in\Omega_4} \left(q^6 - (q^6)^2\right) \\ &= O(m^4)\cdot O(q^6) \\ &= O(m^4q^6).\end{aligned}$$

$$\sum_{C\in\Omega_4} \mathrm{Var}(X_C) = O(m^4q^6). \quad [|\Omega_4| = O(m^4)]$$

5)

$$\sum_{\substack{C,D\in\Omega_4\\ D\neq C}} \mathrm{Cov}(X_C, X_D) = \sum_{\substack{C,D\in\Omega_4\\ D\neq C}} \left(\mathbb{E}[X_C X_D] - \mathbb{E}[X_C]\mathbb{E}[X_D]\right).$$

Here, X_C and X_D are Bernoulli random variables with probability p^6, so $\mathbb{E}[X_C]\mathbb{E}[X_D] = p^6 \cdot p^6 = p^{12}$. We consider 3 cases here,

- If C and D are either completely disjoint or share only one vertex, then having a clique in C does not affect D, as each clique requires 6 edges to form. So, for $X_C X_D = 1$, we need a total of 12 edges (6 for each clique). In this case, disjoint cases do not contribute to the sum.

$$\mathrm{Cov}(X_C, X_D) = p^{12} - p^{12} = 0.$$

- If C and D share two vertices, a total of 11 edges are needed to form a clique in both C and D. 1 edge to connect the common vertices and 5 edges for each clique with the common and unique vertices of C and D. So,

$$\Pr(X_C X_D = 1) = p^{11}, \quad \mathbb{E}[X_C X_D] = p^{11}.$$

$$\mathrm{Cov}(X_C, X_D) = p^{11} - p^{12} = p^{11}(1-p) = O(p^{11}).$$

The number of ways to choose cliques C and D that share exactly 2 vertices is $O(n^6)$. We can choose 2 common vertices in $O(n^2)$ ways, 2 unique vertices for C in $O(n^2)$ ways, and 2 unique vertices for D in $O(n^2)$ ways. So the contribution to the sum is,

$$O(n^6) \cdot O(p^{11}) = O(n^6 p^{11}).$$

- If C and D share three vertices, 9 edges are needed to form a clique in both C and

D. 3 edges to connect the common vertices and 3 edges for each clique with the common vertices and the unique vertex of C and D.

$$\Pr(X_C X_D = 1) = p^9, \quad \mathbb{E}[X_C X_D] = p^9.$$

$$\mathrm{Cov}(X_C, X_D) = p^9 - p^{12} = p^9(1 - p^3) = O(p^9).$$

The number of ways to choose cliques C and D that share exactly 3 vertices is $O(n^5)$. We can choose 3 common vertices in $O(n^3)$ ways, 1 unique vertex for C in $O(n)$ ways, and 1 unique vertex for D in $O(n)$ ways. So, the contribution to the sum is:

$$O(n^5) \cdot O(p^9) = O(n^5 p^9).$$

The total the sum is,

$$O(n^6 p^{11}) + O(n^5 p^9).$$

6)

As $E[X] = O(n^4p^6)$, we can write, $(E[X])^2 = O(n^8p^{12})$. For the variance , we know $\text{Var}(X) = \sum_{C \in \Omega_4} \text{Var}(X_C) + \sum_{\substack{C,D \in \Omega_4 \\ D \neq C}} \text{Cov}(X_C, X_D) = O(n^4p^6) + O(n^6p^{11}) + O(n^5p^9)$,

So, we can write,

$$\frac{\text{Var}(X)}{(E[X])^2} = \frac{O(n^4p^6) + O(n^6p^{11}) + O(n^5p^9)}{O(n^8p^{12})}.$$

Breaking down each term, we get,

$$\frac{\text{Var}(X)}{(E[X])^2} = O\left(\frac{1}{n^4p^6}\right) + O\left(\frac{1}{n^2p}\right) + O\left(\frac{1}{n^3p^3}\right).$$

Since we assume $pn^{2/3} \to \infty$ as $n \to \infty$, each term $O\left(\frac{1}{n^4p^6}\right)$, $O\left(\frac{1}{n^2p}\right)$, and $O\left(\frac{1}{n^3p^3}\right)$ tends to zero. So,

$$\frac{\text{Var}(X)}{(E[X])^2} \to 0.$$

By Chebyshev's inequality,

$$\Pr\left(|X - E[X]| \geq k \cdot \sigma\right) \leq \frac{1}{k^2}.$$

We set $k = \frac{E[X]}{\sqrt{\text{Var}(X)}}$, we get,

$$\Pr\left(|X - E[X]| \geq E[X]\right) \leq \frac{\text{Var}(X)}{(E[X])^2} \to 0.$$

$\Pr(|X - E[X]| \geq E[X]) \to 0$ implies that X is concentrated around $E[X]$.

$$\Pr(G \text{ has a 4-clique}) = 1 - \Pr(X = 0) \approx 1 - \Pr(|X - E[X]| \geq E[X]) \to 1.$$

As $n \to \infty$ and $pn^{2/3} \to \infty$, we have $\Pr(G \text{ has a 4-clique}) \to 1$.