SUBGRAPH EDGE COUNTS IN EXTREMELY DENSE ERDŐS–RÉNYI GRAPHS

RUIYANG LUO, ALLEN WANG, AND DAVID WEI

Abstract. Define $\rho(G) := \{|E(G')| : G' \text{ is an induced subgraph of } G\}$. An area of study in graph theory and statistics literature is analyzing the asymptotic behavior of $|\rho(G)|$ for Erdős–Rényi graphs. Our paper studies the domain of graphs $G(n, p(n))$ where $p(n) = 1 - o(1)$. We give asymptotic results on $|\rho(G(n, 1 - n^{-d}))|$ for $d > 1$. Furthermore, we provide an upper bound on $|\rho(G(n, 1 - n^{-d}))|$ when $d \in (0, 1)$ and conjecture that it is $(1 - o(1))\binom{n}{2}$ with high probability.

1. Introduction

For a graph G, let $\rho(G)$ be the set of edge counts attained by induced subgraphs of G. This quantity measures how "flexibly" its induced subgraphs realize different edge counts. For example, although the complete graph K_n on n vertices has the maximum possible number of edges, its induced subgraphs have only the triangular edge counts $0, 1, 3, 6, \cdots, \binom{n}{2}$, meaning that $|\rho(K_n)| = n$. Thus, denser graphs do not necessarily have larger values of $|\rho(G)|$.

This example illustrates a tradeoff between density and flexibility. Adding edges to a graph may increase the maximum possible induced subgraph edge count, but it may also make possible edge counts more rigid. Thus, the regime $p = 1 - o(1)$ is especially natural, where the graph G is close to complete, but the relatively few missing edges found in the sparse complement graph $\overline{G}$ may allow for many additional attainable edge counts.

There is some existing work in literature regarding the study of $\rho(G)$. [CFM92] introduces the problem, calling it "Range of Subgraph Sizes (RSS)". The paper proves results on random graphs that have a "full RSS", i.e. $\rho(G) = \{0, 1, \ldots, |E(G)|\}$ and proves tight bounds on the size of ρ for dense Erdős–Rényi graphs.

Theorem 1.1 ([CFM92]). *For p constant, we have*

$$|\rho(G(n,p))| = |E(G(n,p))| - \Theta\left(\frac{n^{3/2}}{\sqrt{\log n}}\right)$$

almost surely.

The proof of Theorem 1.1 given by [CFM92] is a constructive one. For

$$\frac{(\log_b(n))^2}{3} \leq \ell \leq p\binom{n}{2} - n^{3/2+\varepsilon}$$

where $b = 1/\min(p, 1-p)$, and arbitrarily small $\varepsilon > 0$, it is possible to construct a subgraph with ℓ edges with high probability. The authors first take a subgraph

---

Ruiyang Luo: rayluo@mit.edu, Allen Wang: alllllen@mit.edu, David Wei: djwei@mit.edu.

which has

$$k = \left\lfloor \sqrt{\frac{2\ell}{p}} - \frac{5}{\sqrt{\varepsilon p^{3/2}(1-p)}} \cdot \sqrt{\frac{\sqrt{\ell}}{\log \ell}} \right\rfloor$$

vertices. This value is chosen so that the number of edges in this graph is less than ℓ but sufficiently close. Then, over the course of three "phases", they iteratively add vertices to the graph to create induced subgraphs whose edge counts get closer and closer to the desired target until it finally reaches ℓ.

Other papers such as [AK09], [BZ22], [Yar25] prove additional results in the constant p domain.

Theorem 1.2 ([CFM92][1]). *Let $G(n,m)$ be a graph with n vertices and m edges that is selected uniformly at random from all graphs with n vertices and m edges. Let $m = \frac{n}{2}(\log n + \log\log n + c_n)$ and c be a constant. We have*

$$\lim_{n\to\infty} \mathbb{P}(G(n,m) \text{ has full RSS}) = \begin{cases} 1, & c_n \to -\infty \\ 1 - e^{-e^{-c}} & c_n \to c \\ 0 & c_n \to \infty. \end{cases}$$

The results of this theorem are applicable when $p = (\log n + \log\log n + c_n)/n$, since results from the model $G(n,m)$ apply to $G(n,p)$ when $|m - \binom{n}{2}p| = O(n\sqrt{p(1-p)})$. Please see [Łuc90] for a more detailed discussion of this so-called "equivalence".

2. RESULTS

We present an upper bound to $|\rho(G(n, 1-n^{-d}))|$ when $d \in (0,1)$ and asymptotic results for $d > 1$.

Theorem 2.1. *For each $d \in (0,1)$, we have*

$$\left|\rho(G(n, 1-n^{-d}))\right| \le \frac{n^2}{2} - \frac{1}{1000} \max\left\{ n^{2-d}, \frac{n^{(3+d)/2}}{\sqrt{\log n}} \right\}$$

with high probability, for all sufficiently large n.

Proof. Let $G = (V,E) \sim G(n, 1-n^{-d})$. Clearly we have $\rho(G) \subseteq \{0, 1, \ldots, |E|\}$. Since $|E| \sim \text{Bin}(\binom{n}{2}, 1-n^{-d})$, by the Chernoff bound we have

$$\begin{aligned} &\mathbb{P}\left(|E| \ge \binom{n}{2}(1-n^{-d}) + 10n\sqrt{\log n} \right) \\ \le\, &\mathbb{P}\left(\left| |E| - \binom{n}{2}(1-n^{-d}) \right| \ge 10n\sqrt{\log n} \right) \\ \le\, &2\exp\left(-\frac{1}{3}\left(10n\sqrt{\log n}\right)^2 \binom{n}{2}^{-1} (1-n^{-d})^{-1} \right) \\ \le\, &2\exp\left(-10\log n\right) = \frac{2}{n^{10}}. \end{aligned}$$

Thus with high probability, for all sufficiently large n,

$$\left|\rho(G(n, 1-n^{-d}))\right| \le \binom{n}{2}(1-n^{-d}) + 10n\sqrt{\log n} \le \frac{n^2}{2} - \frac{1}{1000} n^{2-d}.$$

---

[1] The source of this theorem incorrectly writes $1 - e^{e^{-c}}$ instead of $1 - e^{-e^{-c}}$. The former is absurd, as $e^{e^{-c}} > 1$.

Now for $v \in G$ let $d(v)$ denote the degree of v in the complement graph $\overline{G}$. Fix some $t \in [n]$ and consider an induced subgraph $H \subseteq G$ with $|H| = t$. Let H^C denote the subgraph of G induced by $V \setminus V(H)$. We may write

$$\left|E(H^C)\right| = \binom{n-t}{2} - \left|E(\overline{H^C})\right| = \binom{n-t}{2} - \left|E(\overline{G})\right| + \sum_{v \in V(H)} d(v) - \left|E(\overline{H})\right|.$$

In the rightmost expression, the first two terms are fixed. Let us now bound (with high probability) the last two within an interval.[2] For each vertex $v \in V$, we have $d(v) \sim \text{Bin}(n-1, n^{-d})$, so by Chernoff

$$\begin{aligned}&\mathbb{P}\left(\left|d(v) - (n-1)n^{-d}\right| \geq 10\sqrt{n^{1-d}\log n}\right)\\ &\leq 2\exp\left(-\frac{1}{3}\left(10\sqrt{n^{1-d}\log n}\right)^2\left((n-1)n^{-d}\right)^{-1}\right)\\ &\leq 2\exp\left(-10\log n\right) = \frac{2}{n^{10}}.\end{aligned}$$

Via union bound across all n choices of v, this means that with high probability all $d(v)$ fall lie within $10\sqrt{n^{1-d}\log n}$ of $(n-1)n^{-d}$, i.e.

$$\max_{v \in V} d(v) - \min_{v \in V} d(v) \leq 20\sqrt{n^{1-d}\log n}.$$

Consequently,

$$\begin{aligned}\max_{|H|=t} \sum_{v \in V(H)} d(v) - \min_{|H|=t} \sum_{v \in V(H)} d(v) &\leq \sum_{v \in V(H)} \left(\max_{v \in V} d(v) - \min_{v \in V} d(v)\right)\\ &\leq 20t\sqrt{n^{1-d}\log n}.\end{aligned}$$

Thus for $t \leq \frac{1}{100}\sqrt{\frac{n^{1+d}}{\log n}}$, with high probability the range of $\sum_{v \in V(H)} d(v)$ is at most $\frac{1}{5}n$.

Let $K = \frac{1}{100}\sqrt{\frac{n^{1+d}}{\log n}}$ for convenience. We now bound the range of $\left|E(\overline{H})\right|$, for a fixed choice of $t \leq K$. For a given H, we have $\left|E(\overline{H})\right| \sim \text{Bin}(\binom{t}{2}, n^{-d})$. By Chernoff,

$$\mathbb{P}\left(\left|E(\overline{H})\right| \geq \frac{1}{5}n\right) \leq \left(\frac{e\binom{t}{2}n^{-d}}{n/5}\right)^{n/5} \leq \left(\frac{1}{100\log n}\right)^{n/5}.$$

There are at most $\binom{n}{t} \leq n^K$ choices of H. Hence by union bound, the probability that some $\left|E(\overline{H})\right|$ does not lie in the interval $[0, \frac{1}{5}n]$ is at most

$$\frac{n^K}{(\log n)^{n/5}} = \exp\left(K\log n - \frac{n}{5}\log\log n\right) \leq \exp\left(-\frac{n}{5}(\log\log n - 1)\right),$$

since $K\log n \leq \frac{n}{5}$.

As a result, the range of

$$\sum_{v \in V(H)} d(v) - \left|E(\overline{H})\right|$$

is at most $\frac{2}{5}n$ with high probability, for each fixed $t \leq K$.

---

[2] Some metacommentary: the "point" of this rewrite is that it is much easier to focus on the task of bounding sums of $d(v)$, of which we only have n of, as our bottleneck, than it is to efficiently bound an expression such as $\left|E(\overline{H^C})\right|$ across all exponentially-many choices of H^C.

Now we bound the number of possible values of $\left|E(H^C)\right|$, across all induced $H \subseteq G$ (with no restriction on $|H|$). If $|H| \geq K$, then

$$\left|E(H^C)\right| \leq \binom{n-K}{2} < \frac{n^2}{2} - \frac{1}{2}nK,$$

so there are at most this many values. On the other hand, there are K choices for $|H| < K$, and for each of these, the range of $\left|E(H^C)\right|$ is at most $\frac{2}{5}n$ with high probability. Hence across all such choices of $|H|$, we get at most $\frac{2}{5}nK$ distinct values. Therefore, the total number of values attained by $\left|E(H^C)\right|$, across all induced $H \subseteq G$, is at most

$$\frac{n^2}{2} - \left(\frac{1}{2} - \frac{2}{5}\right)nK = \frac{n^2}{2} - \frac{1}{1000}\frac{n^{(3+d)/2}}{\sqrt{\log n}}.$$

Since the range of H^C is all induced subgraphs of G, our desired result follows.

□

For $d \in (1, 2)$, we see a clear change in the asymptotic behavior.

Theorem 2.2. *For $d \in (1, 2)$, we have*

$$|\rho(G(n, 1 - n^{-d}))| = (1 + o(1))\frac{n^{3-d}}{2}$$

with high probability for sufficiently large n.

Proof. Let $G = (V, E) \sim G(n, 1 - n^{-d})$. We first bound $|\rho(G)|$ from above. For every induced subgraph $G' = (V', E')$, we have

$$|E'| = \binom{|V'|}{2} - |E(\overline{G'})|.$$

Since $|E(\overline{G})| \sim \mathrm{Bin}\left(\binom{n}{2}, n^{-d}\right)$, by Chernoff, we have

$$\mathbb{P}\left(\left|E(\overline{G}) - n^{-d}\binom{n}{2}\right| \geq \varepsilon n^{2-d}/2\right) \leq 2\exp\left(-\frac{\varepsilon^2 n^{2-d}}{6}\right)$$

so $E(\overline{G}) = (1 + o(1))\frac{n^{2-d}}{2}$ with high probability since $d < 2$. Therefore, since there are at most $|V| + 1 = n + 1$ values of $|V'|$ and at most $E(\overline{G}) + 1 = (1 + o(1))\frac{n^{2-d}}{2}$ values of $E(\overline{G'})$, there can only be at most

$$(n + 1) \cdot (1 + o(1))\frac{n^{2-d}}{2} = (1 + o(1))\frac{n^{3-d}}{2}$$

values of $|E'|$.

Now, we show that in fact, $|\rho(G(n, 1 - n^{-d}))| \geq (1 - o(1))\frac{n^{3-d}}{2}$. The key idea is to utilize the fact that $\overline{G}$ is a very "isolated" graph. Concretely speaking, let

$$X = \sum_{v \in V(\overline{G})} \binom{\deg(v)}{2}$$

be the number of pairs of edges that share a vertex. The expected value of X is

$$\mathbb{E}[X] = n\binom{n-1}{2}n^{-2d} = \Theta(n^{3-2d}) = o(n^{2-d})$$

since $d > 1$. Thus, for any fixed $\varepsilon > 0$, by Markov, we have

$$\mathbb{P}(X \geq \varepsilon n^{2-d}) \leq \frac{1}{\varepsilon}\Theta(n^{1-d}) = o(1).$$

This clearly implies $X = o(n^{2-d})$ with high probability, as desired. Moreover, the number of non-isolated edges is at most

$$\sum_{v \in V(\overline{G}), \deg(v) \geq 2} \deg(v) \leq 2 \sum_{v \in V(\overline{G})} \binom{\deg(v)}{2} = 2X = o(n^{2-d}).$$

Since $\overline{G}$ has $m = (1+o(1))\frac{n^{2-d}}{2}$ edges, this implies that $\overline{G}$ also contains $m' = (1+o(1))\frac{n^{2-d}}{2}$ isolated edges. Moreover, we have that at least $n - 2m$ isolated vertices of $\overline{G}$, since there are at most m edges.

Let m' denote the number of isolated edges and m denote the number of edges of $\overline{G}$. Choose $2m' \leq k \leq n - 2m$. For $t = 0, 1, \ldots, m'$, choose t isolated edges of $\overline{G}$. Then pick $k - 2t$ isolated vertices of $\overline{G}$. This is clearly possible because there are at least $n - 2m \geq k \geq k - 2t$ isolated vertices in $\overline{G}$. This produces an induced graph S with k vertices and $\binom{k}{2} - t$ edges, so we have

$$\binom{k}{2}, \ldots, \binom{k}{2} - m'$$

are all in $\rho(G)$.

The rest is simple counting. Firstly, we need to make sure that $\left[\binom{k}{2} - m', \binom{k}{2}\right]$ and $\left[\binom{k+1}{2} - m', \binom{k+1}{2}\right]$ are disjoint. This is clear because

$$\binom{k+1}{2} - m' > \binom{k}{2} \iff k > m',$$

which clearly holds since $k \geq 2m'$. Therefore, we have

$$\begin{aligned} |\rho(G)| &\geq (n - 2m - 2m') \cdot (m' + 1) = (n - o(n))(m' + 1) \\ &= (n - o(n))\frac{(1 - o(1))n^{2-d}}{2} = (1 - o(1))\frac{n^{3-d}}{2}. \end{aligned}$$

As a result, since we have achieved matching upper and lower bounds, we have $|\rho(G)| = (1 + o(1))\frac{n^{3-d}}{2}$, as desired. ☐

Interestingly, there is yet another a distinctive phase shift at $d = 2$.

Theorem 2.3. *We have*

$$\frac{|\rho(G(n, 1 - n^{-2}))|}{n} \xrightarrow{d} 1 + M,$$

where $M \sim \mathrm{Pois}(\frac{1}{2})$ *is a Poisson variable.*

Proof. Consider the complement graph $\overline{G} \sim G(n, n^{-2})$. Let m be the number of edges in $\overline{G}$. Clearly, $m \sim \mathrm{Bin}(\binom{n}{2}, n^{-2})$, which is well-known to converge to the Poisson distribution with parameter $1/2$ as $n \to \infty$.

Now, we claim that $\overline{G}$ is a matching with high probability. In other words, every vertex is part of *at most* one edge in $\overline{G}$. The desired statement follows immediately from a first moment computation. Let X be the number of adjacent edge-pairs in $\overline{G}$. Then, the expected number of triples of vertices $u, v, w \in V(\overline{G})$ for which $\{u, v\}, \{u, w\} \in E(\overline{G})$ is

$$\mathbb{E}[X] = n\binom{n-1}{2}(n^{-2})^2 = O(n^{-1}),$$

since we have n ways to select the shared vertex, $\binom{n-1}{2}$ ways to select two other vertices to connect u to, and a $(n^{-2})^2$ probability that both edges exist. By Markov's inequality, we have

$$\mathbb{P}(X \geq 1) \leq \mathbb{E}[X] \to 0,$$

showing that $\overline{G}$ is a matching with probability at least $1 - O(n^{-1})$.

Consider a set of vertices $S \subseteq V(G)$ such that $|S| = t$. We may write

$$|E(S)| = \binom{t}{2} - r,$$

where r is the number of matching edges fully contained inside S. We will now compute the range of possible values of r, which would tell us the set of possible edge counts that can be attained with a subset of t vertices in G.

Recall that we previously computed that $\overline{G}$ will contain $m \sim \text{Pois}(\frac{1}{2})$ matching edges. Assume $m > 0$ now for the sake of avoiding edge cases. Clearly, the r matched edges in S contribute $2r$ vertices. Also, there are up to $n - 2m$ vertices that are not part of any matched edges that can be part of S as well. Finally, there are $m - r$ matched edges not in S, and we can take at most one vertex from each matched edge to put into S. Thus, the maximum possible number of vertices that S can possibly have is

$$2r + (n - 2m) + (m - r) = n - m + r,$$

meaning that we have the bound $t \leq n - m + r$. Equivalently, we have

$$r \geq m - n + t,$$

which provides a lower bound for r.

We now compute an upper bound for r. Since S contains r matched edges, it is easy to see that S must contain at least $2r$ vertices. Thus, $2r \leq t$. Moreover, $r \leq m$ since there are m matched edges total. As a result, $r \leq \min(m, \lfloor \frac{t}{2} \rfloor)$, giving a desired upper bound.

Combining everything together, we have

$$\max(0, m - n + t) \leq r \leq \min\left(m, \left\lfloor \frac{t}{2} \right\rfloor\right).$$

Note that it is not hard to show that every r in this range is indeed attainable simply through a constructive argument.

Now, restrict $t \in [2m, n - 2m]$. When t is in this range, we have $0 \leq r \leq m$. In fact, every such r is attainable through the following construction. Take r edges in $\overline{G}$ and $t - 2r$ isolated vertices of $\overline{G}$. The vertices attached to the edges combined with the $2r$ vertices form an induced graph in G that is missing r edges, as desired.

So we have that for $2m \leq t \leq n - 2m$,

$$\left[\binom{t}{2} - m, \binom{t}{2}\right] \subseteq \rho(G).$$

Conveniently, these intervals are all disjoint for all t in the range when $m \geq 1$. Indeed, we have

$$\left(\binom{t+1}{2} - m\right) - \binom{t}{2} = t - m \geq m \geq 0.$$

Note that when $m = 0$, the only overlap occurs between the intervals for $t = 0$ and $t = 1$, which both contribute an edge count of zero (causing a duplicate edge count).

We have $(n - 2m) - 2m + 1 = n - 4m + 1$ values of t in that special range, and each such t contributes $m + 1$ distinct edge counts (modulo the single duplicate arising when $m = 0$), giving the lower bound

$$|\rho(G)| \geq (m+1)(n - 4m + 1) - 1 \geq (m+1)n - (m+1)(4m-1) - 1.$$

Note that we accounted for the special case of overlap (as discussed previously) when $m = 0$ by subtracting one from the above lower bound.

For $t \notin [2m, n-2m]$, we have either $t < 2m$ or $t > n - 2m$. In total there are around $4m + O(1)$ values of t, and each value of t contributes at most $m+1$ possible values of r (it would be less in these cases of t, but that is irrelevant). The total number of new edge counts contributed here is at most $4m \cdot (m+1)$, meaning that we have the upper bound

$$|\rho(G)| \leq (m+1)n - m(4m+1) + 4m(m+1) = (m+1)n + 4m.$$

Noting that $m = O(1)$ with high probability, we may conclude that

$$|\rho(G)| = (m+1)n + O(1),$$

so dividing both sides by n, we get that

$$\frac{|\rho(G(n, 1-n^{-2}))|}{n} - (m+1) \xrightarrow{p} 0.$$

Since $m \xrightarrow{d} \text{Poisson}(\frac{1}{2})$ as $n \to \infty$, by Slutsky's Theorem, we have

$$\frac{|\rho(G(n, 1-n^{-2}))|}{n} \xrightarrow{d} m + 1,$$

as desired. □

The "phase change" at $d = 2$ is confirmed once we consider the case when $d > 2$, which is considerably easier than the above cases.

Corollary 2.4. *Fix $d > 2$. We have*

$$G(n, 1 - n^{-d}) = K_n$$

with high probability for sufficiently large n. Consequently, we also have

$$|\rho(G(n, 1-n^{-d}))| = n$$

with high probability.

Proof. Let $\overline{G} \sim G(n, n^{-d})$ denote the associated complement graph. Let $M = E(\overline{G})$ be the number of edges present in $\overline{G}$. Then, we have

$$\mathbb{E}[M] = \binom{n}{2} n^{-d} \sim \frac{1}{2} n^{2-d},$$

which goes to zero when $d > 2$. By Markov's inequality, we get

$$\mathbb{P}(M \geq 1) \leq \mathbb{E}[M] \to 0,$$

meaning that $\overline{G}$ is the empty graph, and G is the complete graph K_n, with high probability. □

3. Future Directions

We were not able to prove a lower bound on $|\rho(G(n, 1-n^{-d}))|$ for $d \in (0,1)$, but we conjecture that it asymptotically matches the upper bound provided by Theorem 2.1.

Conjecture 3.1. *We have*

$$\left|\rho(G(n, 1-n^{-d}))\right| = \frac{n^2}{2} - \tilde{\Theta}\left(\max\{n^{2-d}, n^{(3+d)/2}\}\right)$$

for $d \in (0,1)$ and sufficiently large n.

Here $\tilde{\Theta}$ is the "soft-O" notation commonly used in computer science, which suppresses (here, possibly negative) powers of logarithms.

There are many different ideas once can take to prove this statement. One such idea we tried was utilizing the so-called "second moment method". The second moment method gives an upper bound on the probability of a nonnegative random variable vanishing in terms of the first and second moments. Thus, if $N_{k,\ell}$ denotes the number of induced subgraphs with k vertices and ℓ edges, one could possibly try and show that $N_{k,\ell} \geq 1$ occurs with high probability for a suitable choice of (k, ℓ). The problem we ran into was that it was difficult to get good estimates for the second moment. One could perhaps do this by obtaining very precise estimates of certain binomial coefficients, but it is not clear to us if this is feasible.

Another such direction one could possibly take is to modify the proof given by [CFM92] to adapt it to $p = 1 - n^{-d}$. Of course, this method will require a lot of attention to detail to make sure the proof remains not only valid but strong enough to prove the desired statement, as the constructive method is very lengthy and involved.

We also remark that the choice to consider only edge removal probabilities of the form n^{-d} is not particularly important, and that the relevant results should hold for any $G \sim G(n, 1-f(n))$ whenever $f(n) = o(1)$. For $f(n) = n^{-1}$ in particular, which is not included in our results, the approach in 2.1 should apply with some minor tweaks (which produce some additional log factors). Therefore, if 3.1 holds, the following is true as well:

Conjecture 3.2 (maximum of ρ for Erdős–Rényi Graphs)**.** *For each n, we have*

$$\max_{p_n \in [0,1]} \mathbb{E}\left[|\rho(G(n, p_n)|)\right] = \frac{n^2}{2} - \tilde{\Theta}\left(n^{5/3}\right),$$

attained when

$$p_n = 1 - \tilde{\Theta}\left(n^{-1/3}\right).$$

Note that 2.1 alone is essentially enough to prove that this maximum is *at most* $\frac{n^2}{2} - \tilde{\Theta}(n^{5/3})$, and that if equality is achievable then it must occur at $p_n = 1 - \tilde{\Theta}(n^{-1/3})$.

We may compare this result to a sort of deterministic counterpart from [LY], which establishes that

$$\frac{n^2}{2} - C_1 n \log n \leq \max_{|V(G)|=n} |\rho(G)| \leq \frac{n^2}{2} - C_2 n \log\log n.$$

Therefore, though we are probably able to achieve a $\rho(G)$ on the order of $(1-o(1))\frac{n^2}{2}$ for random graphs, in some sense we still do much worse than the deterministic maximum. This is not surprising, given that the known construction for the lower bound is highly structured and therefore such configurations are "unlikely" to arise from a random graph.

4. Acknowledgments

This paper is our writeup for the final project of the class 18.619 (Discrete Probability and Stochastic Processes) at the Massachusetts Institute of Technology. We thank Professor David Gamarnik and the TAs for giving us this amazing opportunity to do research.

References

[AK09] Noga Alon and A. V. Kostochka. Induced subgraphs with distinct sizes. *Random Structures & Algorithms*, 34(1):45–53, 2009.

[BZ22] József Balogh and Maksim Zhukovskii. On the sizes of large subgraphs of the binomial random graph. *Discrete Mathematics*, 345(2):112675, 2022.

[CFM92] Neil Calkin, Alan Frieze, and Brendan D. McKay. On subgraph sizes in random graphs. *Combinatorics, Probability and Computing*, 1(2):123–134, 1992.

[Łuc90] Tomasz Łuczak. On the equivalence of two basic models of random graphs. In Andrzej Ruciński, Jerzy Jaworski, and Michał Karoński, editors, *Random Graphs '87: Based on Proceedings of the 3rd International Seminar on Random Graphs and Probabilistic Methods in Combinatorics, June 27–July 3, 1987, Poznań, Poland*, pages 151–157. John Wiley & Sons, Chichester, 1990.

[LY] Juisi Li and Yuanchen Yang. Puzzling contest 2026/3/p6. `https://artofproblemsolving.com/community/c6h3799823`. To the best of our knowledge, this is a monthly Chinese student-written mock contest. The user IAmTheHazard is coauthor Allen Wang of this paper.

[Yar25] Yu. N. Yarovikov. On the sizes of k-subgraphs of the binomial random graph. *Doklady Mathematics*, 111(3):199–201, 2025.