



RESEARCH ARTICLE

Two-Point Concentration of the Independence Number of the Random Graph

Tom Bohman¹ and Jakob Hofstad²

¹Department of Mathematical Sciences, Carnegie Mellon University, Pittsburgh, PA, 15213, USA;
E-mail: tbohman@math.cmu.edu.

²Department of Mathematical Sciences, Carnegie Mellon University, Pittsburgh, PA, 15213, USA;
E-mail: jhofstad@andrew.cmu.edu.

Received: 9 December 2022; Accepted: 12 December 2023

2020 Mathematics subject classification: Primary – 05c80; Secondary – 60c05

Abstract

We show that the independence number of $G_{n,p}$ is concentrated on two values if $n^{-2/3+\epsilon} < p \leq 1$. This result is roughly best possible as an argument of Sah and Sawhney shows that the independence number is not, in general, concentrated on two values for $p = o((\log(n)/n)^{2/3})$. The extent of concentration of the independence number of $G_{n,p}$ for $\omega(1/n) < p \leq n^{-2/3}$ remains an interesting open question.

1. Introduction

An independent set in a graph $G = (V, E)$ is a set of vertices that contains no edge of the graph. The independence number of G , denoted $\alpha(G)$, is the maximum number of vertices in an independent set in G . The binomial random graph $G_{n,p}$ has a vertex set of cardinality n , and each potential edge appears independently with probability p , where $p = p(n)$ can vary with n . We say that an integer-valued random variable X defined on $G_{n,p}$ is concentrated on $k = k(n)$ values if there is a function $f(n)$ such that

$$\lim_{n \rightarrow \infty} \mathbb{P}(X \in \{f(n), \dots, f(n) + k - 1\}) = 1.$$

In this work, we address the following question: For which probabilities $p = p(n)$ is $\alpha(G_{n,p})$ concentrated on two values?

The independence number of $G_{n,p}$ has been a central issue in the study of the random graph since the beginning. In the 1970s, Bollobás and Erdős [5] and Matula [17] independently showed that $\alpha(G_{n,p})$ is concentrated on two values when p is a constant. In the late 1980s, Frieze [9] showed that if $\omega(1/n) < p < o(1)$, then

$$\alpha(G_{n,p}) = \frac{2}{p} [\log(np) - \log \log(np) + \log(e/2) \pm o(1)] \quad (1)$$

with high probability. This celebrated result combines the second moment method with a large deviation inequality in a surprising way. Dani and Moore [8] give the best-known lower bound on the independence number of $G_{n,p}$ when p is a constant over n . For further discussion of $\alpha(G_{n,p})$, see the canonical random graphs texts [4] [10] [14].

© The Author(s), 2024. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike licence (<https://creativecommons.org/licenses/by-nc-sa/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the same Creative Commons licence is included and the original work is properly cited. The written permission of Cambridge University Press must be obtained for commercial re-use.

The algorithmic question of finding a large independent set in a random graph is also a central issue. An influential work of Karp and Sipser [15] introduced a simple randomized algorithm that determines the independence number of $G_{n,p}$ for $p = c/n$ such that $c < e$. However, the computational problem of finding a maximum independent set in $G_{n,p}$ for larger p appears to be a very difficult problem; for example, Coja-Oghlan and Efthymiou [6] showed that the space of independent sets in $G_{n,p}$ that are larger than roughly half the independence number form an ‘intricately ragged landscape’ that foils local search algorithms.

Interest in the concentration of the independence number of $G_{n,p}$ is also motivated by the study of the chromatic number of $G_{n,p}$. An influential series of works starting with Shamir and Spencer [19] and culminating in the result of Alon and Krivelevich [2] establish two point concentration of the chromatic number of $G_{n,p}$ for $p < n^{-1/2-\epsilon}$ (also see [16], [1], [7], [20]). More recently, Heckel [12] showed that the chromatic number of $G_{n,1/2}$ is **not** concentrated on any interval of length smaller than $n^{1/4-\epsilon}$. Heckel and Riordan [13] then improved this lower bound on the length of the interval on which the chromatic number is concentrated to $n^{1/2-\epsilon}$. They also state a fascinating conjecture regarding fine detail of the concentration of the chromatic number of $G_{n,1/2}$. Determining the extent of concentration of the chromatic number of $G_{n,p}$ for $n^{-1/2} \leq p < o(1)$ is another central and interesting question. Two point concentration of the domination number of $G_{n,p}$ has also been established for a wide range of values of p [11].

Despite this extensive interest in two point concentration of closely related parameters, two point concentration of $\alpha(G_{n,p})$ itself has not been addressed since the early pioneering results of Bollobás and Erdős and Matula. Our main result is as follows.

Theorem 1. *Let $\epsilon > 0$. If $n^{-2/3+\epsilon} < p \leq 1$, then $\alpha(G_{n,p})$ is concentrated on two values.*

Soon after the original version of this manuscript was posted to the ArXiv, Sah and Sawhney observed that Theorem 1 is roughly best possible.

Theorem 2 (Sah and Sawhney [18]). *Suppose $p = p(n)$ satisfies $p = \omega(1/n)$ and $p < (\log(n)/(12n))^{2/3}$. Then there exists $q = q(n)$ such that $p \leq q \leq 2p$ and $\alpha(G_{n,q})$ is not concentrated on fewer than $n^{-1}p^{-3/2} \log(np)/2$ values.*

Note that this anti-concentration statement is somewhat weak as it is compatible with two point concentration for some functions $p < n^{-2/3}$. On the other hand, it is natural to suspect that the extent of concentration of the independence number is monotone in the regime. See the Conclusion for further discussion.

Now, the second moment method is the main tool in the proof of Theorem 1. Let X_k be the number of independent sets of size k in $G_{n,p}$. We have

$$\mathbb{E}[X_k] = \binom{n}{k} (1-p)^{\binom{k}{2}},$$

and if k is large enough for this expectation to vanish, then there are no independent sets of size k with high probability. This simple observation provides a general upper bound on $\alpha(G_{n,p})$. As far as we are aware, this is the only upper bound on $\alpha(G_{n,p})$ that appears in the literature. It turns out this upper bound is sufficient to establish two point concentration when $n^{-1/2+\epsilon} < p \leq 1$. (This was recently observed by the authors [3] for $n^{-1/3+\epsilon} < p < o(1)$.) However, for smaller values of p this upper bound turns out not to be sufficient for two point concentration as the independence number is actually somewhat smaller than the largest value of k for which $\mathbb{E}[X_k] = \omega(1)$ with high probability.

In the regime $n^{-2/3+\epsilon} < p < n^{-1/2+\epsilon}$, we need to consider a more general object. We define an *augmented independent set of order k* to be a set S of $k+r$ vertices such that the graph induced on S is a matching with r edges and every vertex $v \notin S$ has at least two neighbors in S . Clearly, such an augmented independent set contains 2^r independent sets of size k , and so $\mathbb{E}[X_k]$ can be quite large even when the expected number of augmented independent sets of order k vanishes. As the appearance of a single independent set of size k implies the appearance of an augmented independent set of order at least

k , it follows from this observation that the appearance of augmented independent sets containing large induced matchings can create a situation in which $\alpha(G_{n,p})$ is less than k with high probability (whp) even though $\mathbb{E}[X_k]$ tends to infinity. Furthermore, it turns out the variance in the number of augmented independent sets of order k is small enough to allow us to establish two point concentration using the second moment method in the regime in question.

The remainder of this paper is organized as follows. In the following section, we make some preliminary calculations. We then prove our main result, Theorem 1, in Section 3. Theorem 2 is proved in Section 4. In Section 5, we observe that if $p = c/n$ for some constant c , then there is a constant K such that $\alpha(G_{n,p})$ is not concentrated on any interval of length $K\sqrt{n}$. We conclude by making some remarks regarding the concentration $\alpha(G_{n,p})$ for $\omega(1/n) < p \leq n^{-2/3}$ in Section 6.

We mentioned above that the second moment method applied to X_k is sufficient to establish two point concentration down to $p = n^{-1/3+\epsilon}$ (see [3] for this proof). It turns out that the second moment method applied to the number of maximal independent sets of size k suffices to establish two point concentration down to $p = n^{-1/2+\epsilon}$. We include this proof in an appendix.

2. Preliminary calculations

Throughout Sections 2 and 3, we restrict our attention to $p = p(n)$, where $p = n^{-2/3+\epsilon} < p < 1/\log(n)^2$ for some $\epsilon > 0$. This is valid as the classical results of Bollobás and Erdős and Matula treat the case $p > 1/\log(n)^2$. It is a classical result of Frieze that for all such p we have

$$\alpha(G_{n,p}) = \frac{2[\log(np) - \log(\log(np)) + \log(e/2) \pm o(1)]}{p} \tag{2}$$

with high probability. As we are interested in refinements of this result, we will need some estimates regarding numbers in this range.

Recall that X_k is the number of independent sets of size k in $G_{n,p}$. The upper bound in Frieze’s result is achieved by simply identifying a value of k such that $\mathbb{E}[X_k] = o(1)$. In order to make comparisons with this bound, we define k_x to be the largest value of k for which

$$\mathbb{E}[X_k] = \binom{n}{k}(1-p)^{\binom{k}{2}} > n^{2\epsilon}.$$

We note in passing that the exact value in the exponent of this cut-off function is not crucial; we simply need it to be small (in particular, we need $\mathbb{E}[X_{k_x+2}] = o(1)$) and greater than the exponent in the polynomial function bound that we use to define the variable k_z below.

Now, suppose $k = k_x \pm o(1/p)$. In other words, consider

$$k = \frac{2[\log(np) - \log(\log(np)) + \log(e/2) \pm o(1)]}{p}. \tag{3}$$

(Note that impact of the 2ϵ in the definition of k_x is absorbed in the $o(1)$ term.) Our first observation is that, for such a k , we have

$$\begin{aligned} \frac{ne}{k}(1-p)^{k/2} &= (1+o(1))\frac{ne}{k} \exp\{-\log(np) + \log(\log(np)) - \log(e/2) + o(1)\} \\ &= (1+o(1))\frac{\log(np)}{\log(np) - \log \log(np) + \log(e/2) - o(1)} = 1 + o(1). \end{aligned}$$

So, we have

$$(1-p)^{k/2} = (1+o(1))\frac{k}{en}. \tag{4}$$

It follows that we have

$$\left(1 - (1 - p)^k\right)^{n-k} = \left(1 - (1 + o(1))\left(\frac{k}{en}\right)^2\right)^{n-k} = e^{-(1+o(1))\frac{k^2}{e^2n}}. \tag{5}$$

Similarly,

$$\left[1 - (1 - p)^k - pk(1 - p)^{k-1}\right]^{n-k} = e^{-(1+o(1))\frac{pk^3}{e^2n}}. \tag{6}$$

3. Two-point concentration

In this section, we show that two point concentration of $\alpha(G_{n,p})$ persists for p as low as $n^{-2/3+\epsilon}$ but at a value that does not equal k_x for small p . Our main technical result is as follows. We write $f(n) \sim g(n)$ if $\lim_{n \rightarrow \infty} f/g = 1$.

Theorem 3. *If $n^{-2/3+\epsilon} < p < 1/\log(n)^2$ for some $\epsilon > 0$, then there exists an integer $k_z = k_z(n)$ such that $\alpha(G_{n,p}) \in \{k_z, k_z + 1\}$ whp. Furthermore, we have*

$$k_x - k_z \begin{cases} = 0 & \text{if } p = \omega(\log(n)n^{-1/2}) \\ = \xi_n & \text{if } p = C \log(n)n^{-1/2} \\ \sim \frac{k_x^2}{e^2n} \sim \frac{4\log(np)^2}{e^2p^2n} & \text{if } p = o(\log(n)n^{-1/2}), \end{cases}$$

where $\xi_n \in \left(\frac{1}{e^2C^2} - \frac{5}{2}, \frac{1}{e^2C^2} + \frac{3}{2}\right)$.

We emphasize that the sequence ξ_n for the case $p = C \log(n)n^{-1/2}$ does not converge to a particular value; rather, it ranges over the specified interval. Indeed, ξ_n is more precisely given by the following expression

$$\xi_n = \left[\frac{1}{e^2C^2} - \frac{2 \log(\mathbb{E}[X_{k_x}])}{\log(n)} + 2\epsilon \pm o(1) \right].$$

As $\mathbb{E}[X_{k_x}]$ can take a wide range of values as n varies, we have persistent variation of ξ_n over a small list of values. The remainder of this section is a proof of Theorem 3.

Recall that an *augmented independent set of order k* in a graph G is defined to be a set of vertices S for which there exists some $r \geq 0$ such that $|S| = k + r$ and S contains exactly r edges that form a matching (i.e., these edges are pairwise disjoint) and all vertices outside of S have at least two neighbors in S . To motivate this definition, first note that such a set S contains 2^r independent sets of size k , so this structure is well suited for isolating large variations in the number of independent sets of size k . The fact that we are interested in studying the number of independent sets of size k where k is close to the independence number of G motivates the condition regarding vertices outside of S . Indeed, suppose S is a set such that the induced graph on S is a (not necessarily perfect) matching and v is a vertex outside of S that has one neighbor in S . If v is adjacent to a vertex u that is not in the matching then adding v to S creates an augmented independent set of order k with an additional edge in its matching. So in this situation we would want to include v in S in order to isolate as much variation in the number of independent sets of k as possible. On the other hand, if v is adjacent to a vertex u that is part of the matching, then $S \setminus \{u\} \cup \{v\}$ contains an independent set of size $k + 1$.

Lemma 4. *For any graph G , $\alpha(G) = k$ if and only if G has an augmented independent set of order k but no augmented independent set of any larger order.*

Proof. Let $\hat{\alpha}(G)$ be the largest k for which there is an augmented independent set of order k . As an augmented independent set of order k contains an independent set on k vertices, we have

$$\hat{\alpha}(G) \leq \alpha(G).$$

Now, suppose that G has a maximum independent set S of size k . Let T be a maximum set of vertices that contains S and has the additional property that the graph induced on T is a matching. We claim that T is an augmented independent set; that is, every vertex outside of T has two neighbors in T . Clearly, every vertex outside T has a neighbor in T (since S is a maximum independent set). If v is a vertex outside of T with only one neighbor in T , then this neighbor u is in one of the matching edges (since T is chosen to be a maximum set with the given property), but in this situation the set $(T \setminus \{u\}) \cup \{v\}$ contains an independent set that is larger than S , which is a contradiction. So, T is an augmented independent set of order at least k . Therefore,

$$\alpha(G) \leq \hat{\alpha}(G). \quad \square$$

We prove Theorem 3 by applying the second moment method to the random variable Z , which counts the number of augmented independent sets of order k_z with a prescribed number of edges r_z . This requires some relatively delicate estimates to determine the optimal values of the parameters k_z and r_z . We begin with those calculations. We then move on to the first moment and second moment calculations for Z that suffice to prove Theorem 3.

3.1. Defining and estimating k_z and r_z

In this section, we define k_z and r_z , and establish a number of estimates of these quantities.

To start, let $E(n, k, r)$ denote the expected number of augmented independent sets of order k with exactly r edges in $G_{n,p}$. There are $\binom{n}{k+r}$ ways to choose the location of our set, and given this set there are $\frac{(k+r)!}{(k-r)!2^r r!}$ ways to choose the locations of the r disjoint edges. The probability of choosing internal edges accordingly is

$$p^r (1-p)^{\binom{k+r}{2}-r},$$

and finally, the probability that every vertex outside our set has at least two neighbors within the set is $F(n, k, r)^{n-k-r}$ where we set

$$F(n, k, r) := 1 - (1-p)^{k+r} - (k+r)p(1-p)^{k+r-1}. \tag{7}$$

Thus, we have

$$E(n, k, r) = \binom{n}{k+r} \frac{(k+r)!}{(k-r)!2^r r!} p^r (1-p)^{\binom{k+r}{2}-r} F(n, k, r)^{n-k-r}. \tag{8}$$

Now, we define $k_z = k_z(n)$ to be the largest $k \leq k_x$ for which there exists an r such that $E(n, k, r) > n^\epsilon$.¹

Our first observation is that, while k_z is not equal to k_x when p is sufficiently small, the difference between the two numbers is relatively small.

Lemma 5. *If $p = \omega(\log(n)/\sqrt{n})$, then $k_z = k_x$. Otherwise, we have*

$$k_z \geq k_x - O\left(\frac{k_x^2}{n}\right) \geq k_x - o(k_x^{1/2}).$$

¹We impose the condition $k \leq k_x$ to avoid situations in which $k_z > k_x$. In particular, there are certain values of n and p for which we would have $k_z = k_x + 1$. For example, this would happen if $n^{3\epsilon/2} < \mathbb{E}[X_{k_x+1}] < n^{2\epsilon}$.

Proof. This observation follows from considering $r = 0$ and $k = k_x - \frac{Ck_x^2}{n}$, where C is a sufficiently large constant and applying Equations (4) and (6). We first note that if $p = \omega(\log(n)/\sqrt{n})$, then the expression in Equation (6) is subpolynomial. In this case, we have $E(n, k_x, 0) > n^\epsilon$ and $k_z = k_x$.

Restricting our attention to $p = O(\log(n)/\sqrt{n})$, note that for $k = k_x - O(k_x^2/n)$ we have

$$\frac{\binom{n}{k}(1-p)^{\binom{k}{2}}}{\binom{n}{k+1}(1-p)^{\binom{k+1}{2}}} = \frac{k+1}{n-k}(1-p)^{-k} = (1+o(1))\frac{ne^2}{k}, \tag{9}$$

and thus

$$\begin{aligned} E(n, k, 0) &= \binom{n}{k}(1-p)^{\binom{k}{2}} \exp\left\{-\left(1+o(1)\right)\frac{pk^3}{e^2n}\right\} \\ &\geq \left(\frac{ne^2}{2k}\right)^{\frac{ck^2}{n}} \exp\left\{-\left(1+o(1)\right)\frac{pk^3}{e^2n}\right\}. \end{aligned}$$

Therefore, $E(n, k, 0) > n^\epsilon$ if C is a sufficiently large constant. □

Note that we may henceforth assume that the estimates given in Equations 3–6 hold for $k_z \leq k \leq k_x$.

Next, for $k_z \leq k \leq k_x$ define

$$r_M(k) = r_M(n, k) := \arg \max_r \{E(n, k, r)\}.$$

We are now ready to define the random variable Z . Set $r_z = r_M(k_z)$, and let the variable Z count the number of augmented independent sets of order k_z with exactly r_z internal edges. Note that such sets have $k_z + r_z$ vertices and $\mathbb{E}(Z) = E(n, k_z, r_z) > n^\epsilon$.

We now move on to more accurate estimates for k_z and r_z , starting with an estimate for r_z that we obtain by estimating the ratio of consecutive terms of Equation (8). Note that we always have $r \leq k$, which we will use a couple of times below. First, we show that $F(n, k, r)^{n-k-r}$ will have an insignificant effect on the ratio. One can easily verify, using Equation (4), that $F(n, k, r) = 1 - o(1)$. Furthermore,

$$F(n, k, r + 1) - F(n, k, r) = (k + r)p^2(1 - p)^{k+r-1}$$

exactly. Therefore, we have

$$\begin{aligned} \frac{F(n, k, r + 1)^{n-k-r-1}}{F(n, k, r)^{n-k-r}} &= \left(\frac{F(n, k, r + 1)}{F(n, k, r)}\right)^{n-k-r} F(n, k, r + 1)^{-1} \\ &= (1 + o(1)) \left(1 + \frac{(k + r)p^2(1 - p)^{k+r-1}}{F(n, k, r)}\right)^{n-k-r} \\ &= (1 + o(1)) \left(1 + (1 + o(1))(k + r)p^2(1 - p)^{k+r-1}\right)^{n-k-r}. \end{aligned}$$

Since

$$(k + r)p^2(1 - p)^{k+r-1}(n - k - r) \leq 2kp^2n(1 - p)^{k-1} < \frac{k^3p^2}{n} = o(1),$$

it follows that

$$\frac{F(n, k, r + 1)^{n-k-r-1}}{F(n, k, r)^{n-k-r}} = 1 + o(1), \tag{10}$$

as desired.

With this bound on the ratio $F(n, k, r + 1)^{n-k-r-1}/F(n, k, r)^{n-k-r}$ in hand and applying Equation (4) as in Equation (9), we conclude that if $r = o(1/p)$, then we have

$$\begin{aligned} \frac{E(n, k, r + 1)}{E(n, k, r)} &= (1 + o(1)) \frac{\binom{n}{k+r+1} (1-p)^{\binom{k+r+1}{2}} \left(\frac{(k+r+1)(k-r)p}{2(r+1)} \right)}{\binom{n}{k+r} (1-p)^{\binom{k+r}{2}}} \\ &= (1 + o(1)) \left(\frac{k}{e^2 n} \right) \left(\frac{(k+r+1)(k-r)p}{2(r+1)} \right) \end{aligned} \tag{11}$$

$$= (1 + o(1)) \frac{k^3 p}{2e^2 n(r+1)} (1 - r^2/k^2). \tag{12}$$

This calculation provides the desired estimate for r_z .

Lemma 6.

$$r_z = \left\lceil (1 + o(1)) \frac{4 \log(np)^3}{e^2 n p^2} - 1 \right\rceil.$$

Proof. First, note that the form of Equation (12) suggests that $r_M(k)$ should be roughly $k^3 p/n$ as this is the regime in which the ratio is roughly one. Since $k^3 p/n = o(1/p)$, the estimate given by Equation (12) holds in the vicinity of $r = k^3 p/n$. Furthermore, for $r = \Omega(1/p)$ the arguments above can be easily adapted to show that $E(n, k, r + 1)/E(n, k, r)$ is bounded above by the expression in Equation (12). It follows that we have

$$r \geq k^3 p/(e^2 n) \implies \frac{E(n, k, r + 1)}{E(n, k, r)} < 1. \tag{13}$$

Hence, we can conclude that

$$r_M(k) \leq \frac{k^3 p}{e^2 n} = O\left(\frac{\log(np)k^2}{n}\right). \tag{14}$$

Note that if $r = O(k^3 p/n)$ then the $(1 - r^2/k^2)$ term in Equation (12) is equal to $1 + o(1)$. We conclude that we have

$$r_M(k) = \left\lceil \frac{k^3 p}{2e^2 n} (1 + o(1)) - 1 \right\rceil. \tag{15}$$

In light of Lemma 5, the proof is complete. □

We note in passing that Lemma 6 implies that r_z becomes relevant at $p = \Theta(\log(n)^{3/2}/n^{1/2})$, as $r_z = 0$ for p larger than this regime. Furthermore, for all p that we contemplate in this section (i.e., for $p > n^{-2/3+\epsilon}$) we have

$$r_z = o(k_x^{1/2}). \tag{16}$$

Finally, we turn to estimating k_z . Note that we may assume $p = O(\log(n)n^{-1/2})$ as we have $k_z = k_x$ for larger p by Lemma 5. Note further that Equation (15) implies that $r_M(k) = \Omega(\log(n))$ for $k_z \leq k \leq k_x$ for p in this regime. Recall that, appealing to Equation (6), we have

$$E(n, k_x, 0) = \mathbb{E}[X_{k_x}] e^{-(1+o(1)) \frac{pk_x^3}{e^2 n}}.$$

Next, observe that, since $F(n, k + 1, r) = F(n, k, r + 1)$, we can easily adapt the argument that proves Equation (10) to establish

$$\frac{F(n, k + 1, r)^{n-k-r-1}}{F(n, k, r)^{n-k-r}} = 1 + o(1)$$

for $k_z \leq k \leq k_x$ and $r = o(1/p)$. So we can calculate the ratio over a change in k in the same way that we established Equation (12). Thus, for $k_z \leq k \leq k_x$ and $r = o(1/p)$ we have

$$\frac{E(n, k + 1, r)}{E(n, k, r)} = \frac{k}{e^2 n} (1 + o(1)). \tag{17}$$

Using Equations (6), (17), (12) and (15), we write

$$\begin{aligned} E(n, k, r_M(k)) &= E(n, k_x, 0) \prod_{\ell=k}^{k_x-1} \frac{E(n, \ell, 0)}{E(n, \ell + 1, 0)} \prod_{r=0}^{r_M(k)-1} \frac{E(n, k, r + 1)}{E(n, k, r)} \\ &= \mathbb{E}[X_{k_x}] e^{-(1+o(1)) \frac{pk_x^3}{e^2 n}} \left(\frac{e^2 n}{k} (1 + o(1)) \right)^{k_x-k} \left((1 + o(1)) \frac{k^3 p}{2e^2 n} \right)^{r_M(k)} \frac{1}{r_M(k)!} \\ &= \mathbb{E}[X_{k_x}] e^{-(1+o(1)) \frac{pk_x^3}{2e^2 n}} \left(\frac{e^2 n}{k} (1 + o(1)) \right)^{k_x-k}. \end{aligned}$$

This estimate suffices to establish the following.

Lemma 7.

$$\begin{aligned} n^{-2/3+\epsilon} < p \ll \log(n)n^{-1/2} &\implies k_x - k_z = (1 + o(1)) \frac{k_x^2}{e^2 n} = (1 + o(1)) \frac{4 \log(np)^2}{e^2 p^2 n} \\ p = C \log(n)n^{-1/2} &\implies k_x - k_z = \left\lceil \frac{1}{e^2 C^2} - \frac{2 \log(\mathbb{E}[X_{k_x}])}{\log(n)} + 2\epsilon \pm o(1) \right\rceil. \end{aligned}$$

Proof. First, note that for p in these ranges we have

$$\log(n/k) = (1 + o(1)) \log(np) = (1 + o(1)) \frac{pk}{2}.$$

It follows that $k_x - k_z$ is the smallest integer κ such that

$$\log \mathbb{E}[X_k] - (1 + o(1)) \frac{k_x^3 p}{2e^2 n} + \kappa(1 + o(1)) \frac{pk_x}{2} > \epsilon \log(n).$$

If $p \ll \log(n)n^{-1/2}$, then $\log \mathbb{E}[X_k] = O(\log(n)) = o(k_x^3 p/n)$ and the first part of the Lemma follows. On the other hand, if $p = C \log(n)n^{-1/2}$, then $k_x^3 p/n = (1 + o(1)) \log(n)/C^2$ and $pk_x = \log(n)(1 + o(1))$ and we recover the second part of the lemma. □

3.2. First moment

Here, we show that, with high probability, no augmented independent set of order k appears for any $k_z + 2 \leq k \leq k_x + 1$. This is sufficient because we can simply consider the expected number of independent sets of size $k_x + 2$ to rule out the appearance of any larger independent set. We emphasize that we are restricting our attention to k that satisfy Equation (3). We begin with a simple observation.

Lemma 8. For all k that satisfy Equation (3), we have

$$\sum_{r=0}^k E(n, k, r) \leq 4 \sum_{r=0}^{2r_M(k)} E(n, k, r).$$

Proof. We begin by noting that Equation (12) can be adapted to show

$$\frac{E(n, k, r + 1)}{E(n, k, r)} \leq (1 + o(1)) \frac{k^3 p}{2e^2 n(r + 1)}.$$

Now, we consider cases depending on the value of $r_M = r_M(k)$. First, note that if $r_M = 0$ then we have $k^3 p / (2e^2 n) < 1 + o(1)$ and

$$\sum_{r=0}^k E(n, k, r) \leq (1 + o(1)) E(n, k, 0) \sum_{r=0}^{\infty} \frac{1}{r!} \leq 4E(n, k, 0).$$

Now, suppose $r_M \geq 1$. Recalling Equation (15), note that if $r > 2r_M$, then we have

$$E(n, k, r + 1) / E(n, k, r) \leq 2/3. \tag{18}$$

So we have

$$\begin{aligned} \sum_{r=0}^k E(n, k, r) &= \sum_{r=0}^{2r_M} E(n, k, r) + \sum_{r=2r_M+1}^k E(n, k, r) \\ &\leq \left(\sum_{r=0}^{2r_M} E(n, k, r) \right) + 3E(n, k, 2r_M + 1) \leq \left(\sum_{r=0}^{2r_M} E(n, k, r) \right) + 3E(n, k, 2r_M). \quad \square \end{aligned}$$

Note that it follows from the proof of Lemma 8 that we have

$$E(n, k + 1, k + 1) \leq \frac{2}{3} E(n, k + 1, k).$$

Applying this observation together with Lemma 8 and Equation (17), we have

$$\begin{aligned} \sum_{r=0}^{k+1} E(n, k + 1, r) &< \frac{5}{3} \sum_{r=0}^k E(n, k + 1, r) \\ &< \frac{20}{3} \sum_{r=0}^{2r_M(k+1)} E(n, k + 1, r) \\ &< \frac{k}{n} \sum_{r=0}^{2r_M(k+1)} E(n, k, r) \\ &< \frac{k}{n} \sum_{r=0}^k E(n, k, r). \end{aligned}$$

Therefore, by Lemma 8 and Equation (14), we have

$$\begin{aligned} \sum_{k=k_z+2}^{k_x+1} \left(\sum_{r=0}^k E(n, k, r) \right) &< \frac{2k_z}{n} \sum_{r=0}^{k_z+1} E(n, k_z + 1, r) \\ &\leq \frac{8k_z}{n} \sum_{r=0}^{2r_M(k_z+1)} E(n, k_z + 1, r) \\ &\leq \left(\frac{8k_z}{n} \right) \left(\frac{3k_z^3 p}{e^2 n} \right) E(n, k_z + 1, r_M(k_z + 1)) \\ &\leq \left(\frac{8k_z}{n} \right) \left(\frac{3k_z^3 p}{e^2 n} \right) n^\epsilon \\ &= O(\log(n)^4 n^{-2\epsilon}). \end{aligned}$$

Hence, whp no augmented independent set of order k for any $k \geq k_z + 2$ appears and by Lemma 4, we have $\alpha(G) \leq k_z + 1$ whp.

3.3. Second moment method applied to Z

Here, we prove by the second moment method that an augmented independent set of order k_z appears with high probability; more specifically, we prove that such a set with $k_z + r_z$ vertices and $r_z = r_M(k_z)$ internal edges appears whp. For convenience we let

$$k = k_z, \quad r = r_z = r_M(k_z), \quad \text{and} \quad \tilde{k} = k + r$$

throughout the rest of this subsection. Recall that the random variable Z counts the number of augmented independent sets of order k with r edges, and by the definition of r_M and Z we have

$$\mathbb{E}(Z) > n^\epsilon,$$

and our goal is to show that $Z > 0$ whp.

We now break up Z into a sum of indicator random variables. Let \mathbb{S} be the collection of all pairs (S, m_S) , where S is a set of \tilde{k} vertices and m_S is a matching consisting of r edges all of which are contained in S . Note that

$$|\mathbb{S}| = \binom{n}{\tilde{k}} \frac{\tilde{k}!}{(\tilde{k} - 2r)! 2^r r!}.$$

For each pair $(S, m_S) \in \mathbb{S}$, let Z_{S, m_S} be the indicator random variable for the event that S is an augmented independent set with matching m_S . We have

$$\begin{aligned} \text{Var}(Z) &\leq \sum_{(S, m_S) \in \mathbb{S}} \text{Var}(Z_{S, m_S}) + \sum_{S \neq T} \text{Cov}(Z_{S, m_S}, Z_{T, m_T}) \\ &\leq \mathbb{E}(Z) + \sum_{S \neq T} \mathbb{E}(Z_{S, m_S} Z_{T, m_T}) - \mathbb{E}(Z_{S, m_S}) \mathbb{E}(Z_{T, m_T}), \end{aligned}$$

where here and throughout this section such summations are over all $(S, m_S), (T, m_T) \in \mathbb{S}$ that have the specified S and/or T . Note that we are making use of the full covariance; we will see below that this is necessary to handle sets S, T such that $|S \cap T|$ is roughly k^2/n . Next, let

$$h_i := \frac{1}{\mathbb{E}(Z)^2} \sum_{|S \cap T|=i} \mathbb{E}(Z_{S, m_S} Z_{T, m_T}) - \mathbb{E}(Z_{S, m_S})\mathbb{E}(Z_{T, m_T}) \tag{19}$$

$$= \left(\frac{1}{|\mathbb{S}|}\right)^2 \sum_{|S \cap T|=i} \left(\frac{\mathbb{E}(Z_{T, m_T} | Z_{S, m_S} = 1)}{\mathbb{E}(Z_{S, m_S})} - 1\right). \tag{20}$$

Applying the second moment method, we have $\mathbb{P}(Z = 0) \leq \text{Var}[Z]/\mathbb{E}[Z]^2$. Thus, our goal is to show

$$\sum_{i=0}^{\bar{k}-1} h_i = o(1).$$

We emphasize that $\mathbb{E}(Z)$ can be significantly different from $\mathbb{E}(X)$ here. We consider three ranges of i to establish the desired bounds on $\sum h_i$.

3.3.1. Case 1: $i < \frac{1}{\log(n)p}$.

Fix a pair $(S, m_S) \in \mathbb{S}$, and define \mathcal{E} to be the event that $Z_{S, m_S} = 1$. In other words, \mathcal{E} is the event that S gives an augmented independent set with the prescribed matching m_S . Let \mathcal{E}_T be the event $Z_{T, m_T} = 1$. Recalling Equation (20), we seek to bound

$$h_i = \frac{1}{|\mathbb{S}|} \sum_{T: |S \cap T|=i} \left(\frac{\mathbb{P}(\mathcal{E}_T | \mathcal{E})}{\mathbb{P}(\mathcal{E})} - 1\right). \tag{21}$$

Now, for $\ell = 0, \dots, r$ let \mathbb{S}_ℓ be the collection of pairs $(T, m_T) \in \mathbb{S}$ with the property that m_S and m_T have exactly ℓ edges in common. Define

$$B = \left\lceil \frac{1}{\sqrt{p \ln(n)}} \right\rceil,$$

and, for $B < i < B^2$, let $\mathbb{S}_{i, \ell}$ be the collection of pairs $(T, m_T) \in \mathbb{S}_\ell$ such that $|T \cap S| = i$. Finally, let \mathbb{S}'_ℓ be the collection of pairs $(T, m_T) \in \mathbb{S}_\ell$ with the additional property that $|T \cap S| \leq B$. We make two claims.

Lemma 9. *If $(T, m_T) \in \mathbb{S}_{i, \ell}$ and $i < \frac{1}{\log(n)p}$ then*

$$\mathbb{P}(\mathcal{E}_T | \mathcal{E}) \leq e^{O(k^3 p^2 i/n) + o(1)} \frac{\mathbb{P}(\mathcal{E})}{p^\ell (1-p)^{\binom{\ell}{2}}}.$$

Proof. In order to discuss the conditioning on the event \mathcal{E} , we need to define an additional parameter. Let w be the number of edges in m_T that have exactly one vertex in $S \cap T$ and let W be the vertices in $T \setminus S$ that are included in such edges. To be precise, we define

$$W = \{v \in T \setminus S : \exists u \in S \cap T \text{ such that } uv \in m_T\}.$$

For each vertex $v \in W$, let vv' be the edge of m_T that contains v .

We condition as follows. First, we simply condition on all pairs within S appearing as edges and nonedges as required for the event \mathcal{E} . But we proceed more carefully for vertices v that are not in S . For such vertices, we must reveal information about the neighbors of v in S . We do this by observing

whether or not the potential edges of the form vu , where $u \in S$ actually are edges one at a time, starting with potential neighbors $u \in S \setminus T$, and stopping as soon as we observe at least two such neighbors. For vertices $v \in W$, we add the restriction that the edge $vv' \in m_T$ is the last edge observed in this process. So, for a vertex $v \in T \setminus S$, this process either observes two edges between v and $S \setminus T$ – and therefore observes no potential edges between v and $S \cap T$ – or the process observes at most one edge between v and $S \setminus T$ and therefore observes edges between v and $S \cap T$ and reveals that at least one such edge appears. In this latter case, the conditional probability of \mathcal{E}_T is zero *unless* this vertex v is in W and the observed edge is vv' .

For each set $U \subset T \setminus S$, we define \mathcal{F}_U to be the event that \mathcal{E} holds and U is the set of vertices $v \in T \setminus S$ with the property that our process reveals edges between v and $S \cap T$. Note that we have

$$\mathcal{E} = \bigvee_{U \subset T \setminus S} \mathcal{F}_U$$

and

$$U \not\subseteq W \implies \mathbb{P}(\mathcal{E}_T \wedge \mathcal{F}_U) = 0.$$

Next, let $I = S \cap T$, and define \mathcal{I} to be the event that the process detailed above resorts to revealing potential edges between v and I for more than $\log^2(n)k$ vertices $v \notin S$. Note that the probability that a particular vertex $v \notin S$ resorts to observing potential edges to I is

$$1 - \frac{1 - ((1 - p)^{\tilde{k}-i} + p(\tilde{k} - i)(1 - p)^{\tilde{k}-i-1})}{1 - (1 - p)^{\tilde{k}} - p\tilde{k}(1 - p)^{\tilde{k}-1}} = O\left(\frac{\log(n)k^2}{n^2}\right).$$

Thus, the expected number of vertices that observe edges to I is $O(\log(n)^2k^2/n) = o(k)$. As these events are independent, the Chernoff bound (see Corollary 21.9 in [10]) then implies

$$\mathbb{P}(\mathcal{I}) \leq \exp\{-\log^2(n)k\}. \tag{22}$$

Now, we are ready to put everything together. Applying the law of total probability, we have

$$\begin{aligned} \mathbb{P}(\mathcal{E}_T \mid \mathcal{E}) &= \sum_{U \subset T \setminus S} \left(\mathbb{P}(\mathcal{E}_T \mid \mathcal{F}_U \wedge \mathcal{I}) \cdot \frac{\mathbb{P}(\mathcal{F}_U \wedge \mathcal{I})}{\mathbb{P}(\mathcal{E})} + \mathbb{P}(\mathcal{E}_T \mid \mathcal{F}_U \wedge \bar{\mathcal{I}}) \cdot \frac{\mathbb{P}(\mathcal{F}_U \wedge \bar{\mathcal{I}})}{\mathbb{P}(\mathcal{E})} \right) \\ &\leq \frac{\mathbb{P}(\mathcal{I})}{\mathbb{P}(\mathcal{E})} + \sum_{U \subset T \setminus S} \mathbb{P}(\mathcal{E}_T \mid \mathcal{F}_U \wedge \bar{\mathcal{I}}) \cdot \frac{\mathbb{P}(\mathcal{F}_U)}{\mathbb{P}(\mathcal{E})} \\ &= \frac{\mathbb{P}(\mathcal{I})}{\mathbb{P}(\mathcal{E})} + \sum_{U \subseteq W} \mathbb{P}(\mathcal{E}_T \mid \mathcal{F}_U \wedge \bar{\mathcal{I}}) \cdot \frac{\mathbb{P}(\mathcal{F}_U)}{\mathbb{P}(\mathcal{E})}. \end{aligned}$$

Now, observe that we have

$$\frac{\mathbb{P}(\mathcal{F}_U)}{\mathbb{P}(\mathcal{E})} \leq \left(\frac{(1 - p)^{\tilde{k}-i} + p(\tilde{k} - i)(1 - p)^{\tilde{k}-i-1}}{1 - (1 - p)^{\tilde{k}} - p\tilde{k}(1 - p)^{\tilde{k}-1}} \right)^{|U|} = \left[(1 + o(1)) \frac{pk^3}{e^2n^2} \right]^{|U|},$$

and if $U \subseteq W$, then we have

$$\begin{aligned} \mathbb{P}(\mathcal{E}_T \mid \mathcal{F}_U \wedge \bar{\mathcal{I}}) &\leq \left(\frac{1}{i}\right)^{|U|} (1-p)^{\binom{\tilde{k}}{2} - \binom{i}{2} - i|U| - (r-|U|-\ell)} p^{r-|U|-\ell} \\ &\quad \cdot \left(1 - (1-p)^{\tilde{k}} - \tilde{k}p(1-p)^{\tilde{k}-1}\right)^{n-2\tilde{k}+i-\log(n)^2k} \\ &\leq \frac{\mathbb{P}(\mathcal{E})}{p^\ell(1-p)^{i^2/2}} \left(\frac{1}{ip}\right)^{|U|} \exp\{p \cdot 2ir\} \exp\left\{O\left(\frac{k^3p}{n^2}(\tilde{k} + \log(n)^2k)\right)\right\} \\ &\leq (1+o(1)) \frac{\mathbb{P}(\mathcal{E})}{p^\ell(1-p)^{i^2/2}} \left(\frac{1}{ip}\right)^{|U|} \exp\{p \cdot 2ir\}, \end{aligned}$$

where we use $|U| \leq \min\{i, r\}$ and $k^4p < n^{2-2\epsilon}$. Thus,

$$\begin{aligned} \sum_{U \subseteq W} \mathbb{P}(\mathcal{E}_T \mid \mathcal{F}_U \wedge \bar{\mathcal{I}}) \cdot \frac{\mathbb{P}(\mathcal{F}_U)}{\mathbb{P}(\mathcal{E})} \\ \leq \frac{(1+o(1))e^{O(k^3p^2i/n)}\mathbb{P}(\mathcal{E})}{p^\ell(1-p)^{i^2/2}} \sum_{u=0}^w \binom{w}{u} \left[(1+o(1)) \frac{k^3}{e^2in^2} \right]^u = \frac{(1+o(1))e^{O(k^3p^2i/n)}\mathbb{P}(\mathcal{E})}{p^\ell(1-p)^{i^2/2}}, \end{aligned}$$

where we use $w \leq i$ to bound the sum. We finally note that we have

$$\frac{\mathbb{P}(\mathcal{I})}{\mathbb{P}(\mathcal{E})^2} \leq \exp\{-\log^2(n)k + O(\log(n)k)\} = \exp\{-\Omega(\log^2(n)k)\} = o(1).$$

And the proof of the lemma is complete. □

Lemma 10. For $\ell = 1, \dots, r$, we have

$$|\mathbb{S}_\ell| \leq |\mathbb{S}| \left(\frac{3r^2}{n^2}\right)^\ell.$$

Furthermore, we have

$$|\mathbb{S}_{\ell,i}| \leq |\mathbb{S}| \left(\frac{3r^2i^2}{\tilde{k}^4}\right)^\ell \left(\frac{e\tilde{k}^2}{ni}\right)^i.$$

Proof. We have

$$|\mathbb{S}_\ell| \leq \binom{r}{\ell} \binom{n-2\ell}{\tilde{k}-2\ell} \frac{(\tilde{k}-2\ell)!}{(\tilde{k}-2r)!2^{r-\ell}(r-\ell)!}.$$

Thus,

$$\frac{|\mathbb{S}_\ell|}{|\mathbb{S}|} \leq \binom{r}{\ell} \frac{\tilde{k}(\tilde{k}-1)\cdots(\tilde{k}-2\ell+1)}{n(n-1)\cdots(n-2\ell+1)} \cdot \frac{2^\ell \cdot r(r-1)\cdots(r-\ell+1)}{\tilde{k}(\tilde{k}-1)\cdots(\tilde{k}-2\ell+1)} \leq \left(\frac{3r^2}{n^2}\right)^\ell.$$

The calculation for the second part of the lemma is similar.

$$|\mathbb{S}_{\ell,i}| \leq \binom{r}{\ell} \binom{\tilde{k}-2\ell}{i-2\ell} \binom{n-\tilde{k}}{\tilde{k}-i} \frac{(\tilde{k}-2\ell)!}{(\tilde{k}-2r)!2^{r-\ell}(r-\ell)!}.$$

Thus,

$$\begin{aligned} \frac{|\mathbb{S}_{\ell,i}|}{|\mathbb{S}|} &\leq \frac{\binom{r}{\ell} \binom{\tilde{k}-2\ell}{i-2\ell} \binom{n-\tilde{k}}{\tilde{k}-i} \binom{\tilde{k}}{i}}{\binom{n}{i} \binom{n-i}{\tilde{k}-i}} \cdot \frac{2^\ell \cdot r(r-1) \cdots (r-\ell+1)}{\tilde{k}(\tilde{k}-1) \cdots (\tilde{k}-2\ell+1)} \\ &\leq \binom{r}{\ell} \left(\frac{e\tilde{k}^2}{ni}\right)^i \left(\frac{i}{\tilde{k}}\right)^{2\ell} \cdot \frac{2^\ell \cdot r(r-1) \cdots (r-\ell+1)}{\tilde{k}(\tilde{k}-1) \cdots (\tilde{k}-2\ell+1)} \\ &\leq \left(\frac{3r^2i^2}{\tilde{k}^4}\right)^\ell \left(\frac{e\tilde{k}^2}{ni}\right)^i. \end{aligned} \quad \square$$

We are now ready to show that $\sum_{i=0}^{1/(p \log(n))} h_i = o(1)$ using Lemmas 9 and 10. We will split this into two subranges.

Subcase 1.1: $i \leq B$.

In this case, by Lemma 9 we have $\mathbb{P}(\mathcal{E}_T \mid \mathcal{E}) \leq (1 + o(1)) \frac{\mathbb{P}(\mathcal{E})}{p^\ell}$, hence

$$\begin{aligned} \sum_{i=0}^B h_i &= \frac{1}{|\mathbb{S}|} \sum_{\ell=0}^r \sum_{(T, m_T) \in \mathbb{S}'_\ell} \left(\frac{\mathbb{P}(\mathcal{E}_T \mid \mathcal{E})}{\mathbb{P}(\mathcal{E})} - 1 \right) \\ &= \frac{o(|\mathbb{S}'_0|)}{|\mathbb{S}|} + \sum_{\ell=1}^r \left(\frac{4r^2}{pn^2} \right)^\ell \\ &\leq o(1) + \sum_{\ell=1}^r \left(\frac{1}{n^{2/3}} \right)^\ell \\ &= o(1). \end{aligned}$$

Subcase 1.2: $B < i < B^2$.

Here, we have

$$\begin{aligned} h_i &= \frac{1}{|\mathbb{S}|} \sum_{\ell=0}^r \sum_{(T, m_T) \in \mathbb{S}_{\ell,i}} \left(\frac{\mathbb{P}(\mathcal{E}_T \mid \mathcal{E})}{\mathbb{P}(\mathcal{E})} - 1 \right) \\ &\leq \sum_{\ell=0}^r \left(\frac{|\mathbb{S}_{\ell,i}|}{|\mathbb{S}|} \right) \frac{e^{O(k^3 p^2 i/n) + o(1)}}{p^\ell (1-p)^{\binom{i}{2}}} \\ &\leq \sum_{\ell=0}^r \left(\frac{3r^2 i^2}{\tilde{k}^4} \right)^\ell \left(\frac{e\tilde{k}^2}{ni} \right)^i \left(\frac{2e^{O(k^3 p^2 i/n)}}{p^\ell (1-p)^{\binom{i}{2}}} \right) \\ &\leq \left(\frac{e^2 \tilde{k}^2}{ni(1-p)^{i/2}} \right)^i \sum_{\ell=0}^r \left(\frac{3r^2}{\tilde{k}^2 p} \right)^\ell \\ &\leq \left(\frac{e^3 \tilde{k}^2 p^{1/2} \ln(n)^{1/2}}{n} \right)^i \sum_{\ell=0}^r \left(\frac{3r^2}{\tilde{k}^2 p} \right)^\ell. \end{aligned}$$

Since $\frac{r^2}{\tilde{k}^2 p} = O\left(\frac{k^4 p}{n^2}\right) = o(1)$ and $\frac{\tilde{k}^2 p^{1/2} \log(n)^{1/2}}{n} = o(1)$, we have

$$\sum_{i=B}^{B^2} h_i = o(1).$$

3.3.2. Case 2: $\frac{1}{\log(n)p} \leq i < (1 - \epsilon)\tilde{k}$.

In this case, we can afford to be lenient with our bounds; in particular, we do not need to make use of the condition that vertices outside S have at least two neighbors in S . We begin by bounding $\frac{1}{\mathbb{E}(Z_{S,m_S})} \sum_{m_S, m_T} \mathbb{E}(Z_{S,m_S} Z_{T,m_T})$ for a fixed pair of sets S, T such that $|S \cap T| = i$. If there are ℓ edges in the intersection, a bound for the number of ways to choose m_S and m_T is $\tilde{k}^{4r-2\ell}$ (we have $2r - \ell$ edges total; each has less than \tilde{k}^2 choices.) Thus, we have

$$\begin{aligned} \sum_{m_S, m_T} \mathbb{E}(Z_{S,m_S} Z_{T,m_T}) &\leq \sum_{\ell=0}^r \tilde{k}^{4r-2\ell} p^{2r-\ell} (1-p)^{2\binom{\tilde{k}}{2} - \binom{i}{2} - 2r + \ell} \\ &= \tilde{k}^{4r} p^{2r} (1-p)^{2\binom{\tilde{k}}{2} - \binom{i}{2} - 2r} \sum_{\ell=0}^r \left(\frac{1-p}{\tilde{k}^2 p} \right)^\ell = (1 + o(1)) \tilde{k}^{4r} p^{2r} (1-p)^{2\binom{\tilde{k}}{2} - \binom{i}{2} - 2r}. \end{aligned}$$

Recalling Equation (6), we have

$$\begin{aligned} \sum_{m_S, m_T} \frac{\mathbb{E}(Z_{S,m_S} Z_{T,m_T})}{\mathbb{E}(Z_{S,m_S})^2} &\leq (1 + o(1)) \tilde{k}^{4r} (1-p)^{-\binom{i}{2}} e^{(2+o(1))\frac{pk^3}{e^2 n}} \\ &\leq (1-p)^{-\binom{i}{2}} e^{O(\log(n)r+1)}. \end{aligned}$$

It follows that

$$\begin{aligned} h_i &\leq \frac{\binom{\tilde{k}}{i} \binom{n-\tilde{k}}{\tilde{k}-i}}{\binom{n}{\tilde{k}}} (1-p)^{-\binom{i}{2}} e^{O(\log(n)r+1)} \\ &\leq \left(\frac{e\tilde{k}}{i} \right)^i \left(\frac{\tilde{k}}{n} \right)^i (1-p)^{-\binom{i}{2}} e^{O(\log(n)r+1)} \\ &\leq \left[\frac{e\tilde{k}^2}{in} \left((1-p)^{-\tilde{k}/2} \right)^{i/\tilde{k}} \right]^i e^{O(\log(n)r+1)} \\ &\leq \left[\frac{e^2 \log(n) p \tilde{k}^2}{n} \left(\frac{ne}{\tilde{k}} \right)^{1-\epsilon} \right]^i e^{O(\log(n)r+1)} \\ &\leq \left[\frac{e^{3-\epsilon} \log(n) p \tilde{k}^{1+\epsilon}}{n^\epsilon} \right]^i e^{O(\log(n)r+1)}. \end{aligned}$$

As $r = o(k^{1/2}) = o(i)$, we have

$$\sum_{i=1/(p \log(n))}^{(1-\epsilon)\tilde{k}} h_i = o(1).$$

3.3.3. Case 3: $i \geq (1 - \epsilon)\tilde{k}$

We calculate h_i as defined in Equation (19). For simplicity, let $j = \tilde{k} - i$. For some fixed i (hence j), there are $\binom{n}{\tilde{k}} \binom{n-\tilde{k}}{j} \binom{\tilde{k}}{i}$ ways to choose S and T . Now, fix sets S and T .

We now consider the number of ways to choose the matchings m_S and m_T that are compatible with the event $\{Z_{S,m_S} Z_{T,m_T} = 1\}$. Let R_I be the set of matching edges contained in $S \cap T$, and let $r_I := |R_I|$ (so r_I is the same as ℓ from Cases 1 and 2). Secondly, let R_O be the set of edges in m_S that are not in R_I , and let R'_O be the set of edges of m_T that are not in R_I , and define $r_O := |R_O| = |R'_O|$ (' I ' stands for 'inner', and ' O ' stands for 'outer'). Hence, $r = r_O + r_I$. Finally, let R_S be the matching edges in R_O which are contained completely in $S \setminus T$, and let $r_S = |R_S|$. See Figure 1.

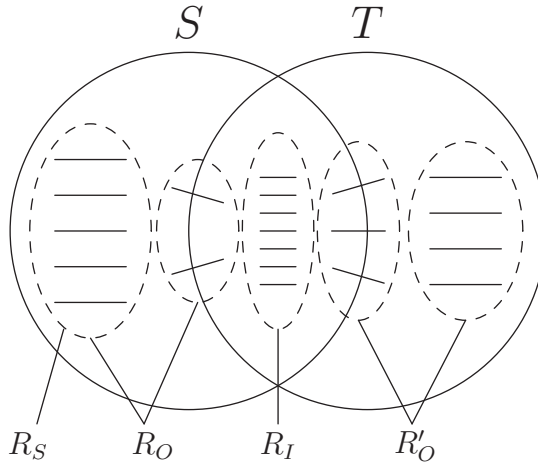


Figure 1. A depiction of the edges appearing in intersecting augmented independent sets.

We now count the number of ways to choose edges in R_O if r_O is fixed. To do this, we first note that if r_S is also fixed, then the number of ways to choose the edges that comprise R_S is

$$\frac{j!}{(j - 2r_S)!r_S!2^{r_S}}.$$

Then the number of ways to choose the rest of the rest of the edges in R_O is

$$\left(\frac{(j - 2r_S)!}{(j - r_S - r_O)!}\right)\left(\frac{(\tilde{k} - j)!}{(\tilde{k} - j - (r_O - r_S))!}\right)\left(\frac{1}{(r_O - r_S)!}\right);$$

therefore, the total number of ways to choose R_O is

$$\begin{aligned} &\sum_{r_S} \frac{1}{2^{r_S}} \left(\frac{j!}{r_S!(r_O - r_S)!}\right)\left(\frac{(\tilde{k} - j)!}{(\tilde{k} - j - (r_O - r_S))!(j - r_S - r_O)!}\right) \\ &\leq \sum_{r_S} \left(\frac{j!}{r_S!(r_O - r_S)!}\right)\left(\frac{(\tilde{k} - j)^{r_O - r_S}}{(j - r_S - r_O)!}\right) \\ &\leq (\tilde{k} - j)^{r_O} \sum_{r_S} \frac{j!}{r_S!(r_O - r_S)!(j - r_O)!} \\ &\leq 3^j \tilde{k}^{r_O}. \end{aligned} \tag{23}$$

By a symmetric argument, Equation (23) is an upper bound for the number of ways to choose the edges in R'_O as well.

Next, note that the number of ways to choose edges in R_I is bounded by $\tilde{k}^{2r_I} / (2^{r_I} r_I!)$. Putting these together, the number of ways to choose m_S and m_T is at most

$$\frac{9^j \tilde{k}^{2r}}{2^{r_I} r_I!} \leq \frac{9^j \tilde{k}^{2r} (2r)^{r_O}}{2^{r_I} r_I!} \tag{24}$$

(assuming that $r^{r_O} = 1$ if $r = r_O = 0$).

Now, we estimate the probability of the event $\{Z_{S,m_S}Z_{T,m_T} = 1\}$. The probability the edges within S and within T are chosen accordingly is

$$p^{r+r_O}(1-p)^{\binom{k}{2}+(\tilde{k}-j)j+\binom{j}{2}-r-r_O}.$$

Now, consider pairs of vertices from $S \setminus T$ to $T \setminus S$. Recall that a vertex $v \in S \setminus T$ has at least two neighbors in T . So, if v is already incident with an edge that goes into T among the edges already specified, then there is at least one edge from v to $T \setminus S$ (since each from $S \setminus T$ can send at most one edge to $S \cap T$). Such an edge appears with probability $1 - (1-p)^j \leq jp$. If a vertex $v \in S \setminus T$ is not in an edge of m_S that intersects $S \cap T$, then v is incident with at least two edges going to $T \setminus S$. The probability of these two edges appearing is at most j^2p^2 by the union bound. Since the number of vertices in the first category is at most r_O , the total probability of all edges between $S \setminus T$ and $T \setminus S$ being chosen accordingly is at most $(jp)^{2j-r_O}$. Hence, the probability that the edges within $S \cup T$ appear in accordance with the event $\{Z_{S,m_S}Z_{T,m_T} = 1\}$ is at most

$$p^{r+r_O}(1-p)^{\binom{k}{2}+(\tilde{k}-j)j+\binom{j}{2}-r-r_O}(jp)^{2j-r_O}.$$

Finally, the probability that all vertices outside $S \cup T$ have at least two neighbors in S is at most $(1 - (1-p)^{\tilde{k}} - \tilde{k}p(1-p)^{\tilde{k}-1})^{n-2\tilde{k}}$. Note that, since

$$\begin{aligned} (1 - (1-p)^{\tilde{k}} - \tilde{k}p(1-p)^{\tilde{k}-1})^{-\tilde{k}} &\leq \exp\{(1+o(1))k^2p(1-p)^k\} \\ &= \exp\{(1+o(1))k^4p/(e^2n^2)\} \\ &= 1 + o(1), \end{aligned}$$

we have

$$\mathbb{P}(Z_{S,m_S}Z_{T,m_T} = 1) \leq (1 + o(1))\mathbb{P}(Z_{S,m_S} = 1)p^{r_O}(1-p)^{(\tilde{k}-j)j+\binom{j}{2}-r-r_O}(jp)^{2j-r_O}.$$

Multiplying this probability estimate by the estimate for the number of choices for m_S and m_T given by Equation (24) and summing over r_O , we see that for a fixed S and T we have

$$\frac{1}{\mathbb{P}(Z_{S,m_S} = 1)} \sum_{m_S, m_T} \mathbb{P}(Z_{S,m_S}Z_{T,m_T} = 1)$$

is at most

$$\begin{aligned} (1 + o(1)) \sum_{r_O=0}^{\min\{r,j\}} \frac{9^j \tilde{k}^{2r} (2r)^{r_O}}{2^r r!} p^{r_O} (1-p)^{(\tilde{k}-j)j+\binom{j}{2}-r-r_O} (jp)^{2j-r_O} \\ = (1 + o(1)) \frac{\tilde{k}^{2r}}{2^r r!} \left(9p^2 j^2 (1-p)^{\tilde{k}-(j+1)/2}\right)^j \sum_{r_O=0}^{\min\{r,j\}} \left(\frac{2r}{j(1-p)}\right)^{r_O} \\ \leq (1 + o(1)) \frac{\tilde{k}^{2r}}{2^r r!} \left(9p^2 j^2 (1-p)^{\tilde{k}-(j+1)/2}\right)^j \max\{1, 2(3r)^j\} \\ \leq (2 + o(1)) \frac{\tilde{k}^{2r}}{2^r r!} \left(30p^2 j^2 (1-p)^{\tilde{k}-j/2} \max\{1, r\}\right)^j \\ \leq (3 + o(1)) \frac{\tilde{k}!}{(\tilde{k} - 2r)! 2^r r!} \left(30p^2 j^2 (1-p)^{\tilde{k}-j/2} \max\{1, r\}\right)^j, \end{aligned}$$

where we use $r = o(k^{1/2})$ in the last step (and assume $r^o = 1$ if $r = r_o = 0$). Recalling that $|\mathbb{S}| = \binom{n}{\tilde{k}} \frac{\tilde{k}!}{(\tilde{k}-2r)!2^r r!}$ and that the number of choices for the pair of sets S, T is $\binom{n}{\tilde{k}} \binom{n-\tilde{k}}{j} \binom{\tilde{k}}{j}$, we have

$$\begin{aligned} h_i \mathbb{E}(Z) &\leq \frac{1}{|\mathbb{S}| \mathbb{P}(Z_{S,m_S} = 1)} \sum_{|\mathbb{S} \cap T|=i} \sum_{m_S, m_T} \mathbb{P}(Z_{S,m_S} Z_{T,m_T} = 1) \\ &\leq (3 + o(1)) \binom{n - \tilde{k}}{j} \binom{\tilde{k}}{j} \left(30p^2 j^2 (1-p)^{\tilde{k}-j/2} \max\{1, r\}\right)^j \\ &\leq (3 + o(1)) \left(\frac{n\tilde{k}e^2}{j^2}\right)^j \left(30p^2 j^2 (1-p)^{\tilde{k}-j/2} \max\{1, r\}\right)^j \\ &= (3 + o(1)) \left(30e^2 n\tilde{k}p^2 (1-p)^{\tilde{k}-j/2} \max\{1, r\}\right)^j \\ &\leq (3 + o(1)) \left(31e^\epsilon \tilde{k}^{3-\epsilon} p^2 n^{-1+\epsilon} \max\{1, r\}\right)^j. \end{aligned}$$

As $r = o(k^{1/2})$ and $k = O(\log(n)n^{2/3-\epsilon})$, we have $\sum_{i=(1-\epsilon)\tilde{k}}^{\tilde{k}-1} h_i = o(1)$, as desired.

4. Weak anti-concentration of $\alpha(G_{n,p})$ for $n^{-1} < p < n^{-2/3}$

In this section, we present the argument of Sah and Sawhney that shows that two-point concentration of the independence number does not extend to $p = n^\gamma$ if $\gamma \leq -2/3$; that is, we prove Theorem 2. We note in passing that this proof is similar to an argument that appears in [2] (see the last item in the ‘concluding remarks and open problems’ section of [2]). Throughout this section, we assume $\omega(1/n) < p < (\log(n)/n)^{2/3}$. Recall Theorem 2:

Theorem 2 (Sah and Sawhney [18]). *Let $p = p(n)$ satisfy $\omega(1/n) < p < (\log(n)/(12n))^{2/3}$, and set*

$$\ell = n^{-1} p^{-3/2} \log(np)/2.$$

Then there exists $q = q(n)$ such that $p \leq q \leq 2p$ such that $\alpha(G_{n,q})$ is not concentrated on ℓ values.

For p in the specified range, we define

$$p' = p + n^{-1} \sqrt{p}.$$

We first observe that this choice of p' is close enough to p to ensure that there is no ‘separation’ of the intervals over which $G_{n,p}$ and $G_{n,p'}$ are respectively concentrated. To be precise, we establish the following:

Lemma 11. *If $\omega(n^{-1}) < p < o(1)$, then there is a sequence $k = k(n)$ such that*

$$\mathbb{P}[\alpha(G_{n,p}) \leq k] > \frac{1}{20} \quad \text{and} \quad \mathbb{P}[\alpha(G_{n,p'}) \geq k] > \frac{1}{20}$$

for n sufficiently large.

Proof. First, observe that, since the distributions of $e(G_{n,p})$ and $e(G_{n,p'})$ are approximately Gaussian with equal variances and with means that are $1/\sqrt{2}$ standard deviations apart, there is some value m such that

$$\mathbb{P}[e(G_{n,p}) > m] > \frac{1}{10} \quad \text{and} \quad \mathbb{P}[e(G_{n,p'}) < m] > \frac{1}{10}.$$

Now let k be the median value of $\alpha(G_{n,m})$. As the addition of edges does not increase the independence number, the lemma follows. □

Now, we prove Theorem 2. Assume for the sake of contradiction that if n is sufficiently large and $p \leq q \leq 2p$, then there is an interval I such that $|I| \leq \ell$ and

$$\mathbb{P}[\alpha(G_{n,q}) \in I] > \frac{49}{50}.$$

Consider the sequence p_0, p_1, p_2, \dots , where $p_0 = p$ and $p_{i+1} = p'_i$ for $i \geq 0$, and let z be the lowest integer such that $p_z \geq 2p$; it is easy to see that $z \leq n\sqrt{p}$. Let $I_0 = [a_0, b_0], I_1 = [a_1, b_1], \dots, I_{z-1} = [a_{z-1}, b_{z-1}]$ be intervals of ℓ values such that $\alpha(G_{n,p_i})$ lies in I_i with probability at least $49/50$ for each i . By Lemma 11, we have $a_i \leq b_{i+1}$ for all $i < z$. This implies $b_{i+1} \geq b_i - \ell$. Iterating this observation gives

$$b_z \geq b_0 - z\ell \geq b_0 - \frac{\log np}{2p}.$$

On the other hand, by Equation (1), we have

$$b_0 - b_z = (1 + o(1))\frac{2 \log(np)}{p} - (1 + o(1))\frac{2 \log(2np)}{2p} = (1 + o(1))\frac{\log(np)}{p}.$$

This is a contradiction.

5. Anti-concentration of $\alpha(G_{n,p})$ for $p = c/n$

In this section, we establish anti-concentration of the independence number of $G_{n,p}$ for $p = c/n$, where c is a constant. We note in passing that for very small $p(n)$ it is easy to show that $\alpha(G)$ is not narrowly concentrated. Indeed, if $p = n^\gamma$ for some $\gamma \in (-2, -3/2)$, then whp no two edges intersect, and hence the independence number is determined by the number of edges that appear. Since the distribution of the number of edges is roughly Poisson with mean $\mu \approx n^{2+\gamma}/2$, then we will not have concentration of $\alpha(G)$ on fewer than $\Theta(n^{1+\gamma/2})$ values.

We now state our anti-concentration result.

Theorem 12. *Let $p = c/n$, where $c > 0$ is a constant, and let $k = k(n)$ be an arbitrary sequence of integers. There exists a constant $L > 0$ such that $\mathbb{P}(\alpha(G_{n,p}) = k) \leq Ln^{-1/2}$ for n sufficiently large.*

Theorem 12 implies that $\alpha(G_{n,c/n})$ is not concentrated on fewer than $\sqrt{n}/(2L)$ values. Note that this is the best possible anti-concentration result in this regime (up to the constant) as standard martingale methods (such as Hoeffding–Azuma or Talagrand’s inequality) show that, for any function $k(n) = \omega(\sqrt{n})$, the independence number is concentrated on some interval of length $k(n)$ (this assertion holds for any value of p).

For a fixed graph G , let $T(G)$ be the subgraph of G given by all connected components that are trees on 4 vertices. Let $t(G)$ be the number of components in $T(G)$. Finally, let $\bar{T}(G)$ be the remainder of the graph G (so $G = T(G) + \bar{T}(G)$). Let $p = c/n$ and let $G = G_{n,p}$. We set $\mu = \mu(c) = \frac{2c^3 e^{-4c}}{3}$. Note that we have

$$E[t(G)] = \binom{n}{4} \cdot 16 \cdot p^3 (1 - p)^{3+4(n-4)} = (1 + o(1))\frac{16c^3 e^{-4c} n}{24} = (1 + o(1))\mu n.$$

We note that $t(G)$ is a well concentrated random variable.

Lemma 13.

$$\mathbb{P}(t(G) < \mu n/2) = O\left(\frac{1}{n}\right).$$

Proof. Here, we use standard second moment method once again. For vertex set S with $|S| = 4$, let W_S be the random variable that indicates if the graph induced on S is an isolated tree. Set $W = t(G)$, and note that $t(G)$ is the sum of all $\binom{n}{4}$ indicators. By Chebychev, we have

$$\mathbb{P}\left(W < \frac{\mu n}{2}\right) \leq \mathbb{P}\left(|W - \mathbb{E}(W)| \geq \frac{\mathbb{E}(W)}{3}\right) \leq \frac{9\text{Var}(W)}{\mathbb{E}(W)^2}.$$

Hence

$$\mathbb{P}\left(W < \frac{\mu n}{2}\right) \leq \frac{9\text{Var}(W)}{\mathbb{E}(W)^2} < 9\left(\frac{1}{\mathbb{E}(W)} + \max_{S,T} \left\{ \frac{\mathbb{E}(W_S W_T) - \mathbb{E}(W_S)\mathbb{E}(W_T)}{\mathbb{E}(W_S)^2} \right\}\right).$$

Note that the only case where we get a positive covariance is where $S \cap T = \emptyset$, and in this case, we have

$$\frac{\mathbb{E}(W_S W_T) - \mathbb{E}(W_S)\mathbb{E}(W_T)}{\mathbb{E}(W_S)^2} = (1 - p)^{-16} - 1 = \frac{16c + o(1)}{n}.$$

Therefore,

$$\mathbb{P}\left(t(G) < \frac{\mu n}{2}\right) < \frac{9 + o(1)}{\mu n} + \frac{144c + o(1)}{n} = O\left(\frac{1}{n}\right). \quad \square$$

With this observation in hand, we are ready to proceed to the proof of Theorem 12. Let k be a fixed integer. Let \mathcal{T} be the collection of all graphs H on $v_H \leq n$ vertices such that $4 \mid n - v_H$ and H has no connected component that is a tree on four vertices. Let \mathcal{T}' be the collection of graphs in $H \in \mathcal{T}$ such that $v_H \leq n - 2\mu n$ vertices. Applying the law of total probability, we have

$$\begin{aligned} \mathbb{P}(\alpha(G) = k) &= \sum_{H \in \mathcal{T}} \mathbb{P}(\alpha(G) = k \mid \bar{T}(G) = H) \mathbb{P}(\bar{T}(G) = H) \\ &\leq \mathbb{P}(t(G) \leq \mu n/2) + \sum_{H \in \mathcal{T}'} \mathbb{P}(\alpha(G) = k \mid \bar{T}(G) = H) \mathbb{P}(\bar{T}(G) = H). \end{aligned}$$

So, it suffices to bound the conditional probability

$$\mathbb{P}(\alpha(G) = k \mid \bar{T}(G) = H),$$

where H is a graph in \mathcal{T}' . Under this conditioning, the graph G is H together with a forest of $(n - v_H)/4$ trees on four vertices. Furthermore, the number of trees in this forest which are stars on four vertices is a binomial random variable with $(n - v_H)/4$ trials with probability of success $1/4$ on each trial. Let B be this random variable, and note that if G is drawn from the conditional probability space on the collection of graphs with $\bar{T}(G) = H$, then we have

$$\alpha(G) = \alpha(H) + (n - v_H)/2 + B.$$

Therefore, we have

$$\mathbb{P}(\alpha(G) = k \mid \bar{T}(G) = H) = \mathbb{P}(B = k - \alpha(H) - (n - v_H)/2 \mid \bar{T}(G) = H).$$

But B is a binomial random variable with a linear number of trials, and therefore the probability that B is any particular value is at most L/\sqrt{n} for some constant L .

6. Conclusion

In this section, we present some remarks and open questions related to the extent of concentration of $\alpha(G_{n,p})$ for $p < n^{-2/3}$. We write $f = \tilde{O}(g)$ if f/g is bounded above by a polylogarithmic function and $f = \tilde{\Omega}(g)$ if f/g is bounded below by a polylogarithmic function.

- The precise extent of concentration of $\alpha(G_{n,p})$ for $p = n^\gamma$ with $-1 < \gamma < -2/3$ remains an interesting open question. Note that Theorem 2 gives a lower bound of $\tilde{\Omega}(n^{-1-3\gamma/2})$ on the extent of concentration in this range. Furthermore, Theorems 1 and 12 show that this lower bound is of the correct order at either end of the interval. So it is natural to ask if this lower bound is actually tight throughout this range.

Question 14. Suppose $p = n^\gamma$ where γ is a constant such that $-1 < \gamma < -2/3$. Is $\alpha(G_{n,p})$ concentrated on $\tilde{O}(n^{-1-3\gamma/2})$ values?

- We do not see a way to adapt the argument given in Section 4 to prove anti-concentration of $\alpha(G_{n,m})$ for $m \leq n^{4/3}$. We conjecture that we have such anti-concentration.

Conjecture 15. Suppose $m = n^\eta$, where η is a constant such that $1 < \eta < 4/3$. Then $\alpha(G_{n,m})$ is not concentrated on $n^{2-3\eta/2-\epsilon}$ values.

At the moment, we have no lower bound on the extent of concentration $\alpha(G_{n,m})$ in this regime. It seems natural to suspect that Conjecture 15 is closely related to the questions of proving a stronger, uniform form of anti-concentration of $\alpha(G_{n,p})$ for $p = n^\gamma$ with $-1 < \gamma < -2/3$. We believe that the following strengthening of Theorem 2 holds.

Conjecture 16. If $p = n^\gamma$ where γ is a constant such that $-1 < \gamma < -2/3$, then we have

$$\mathbb{P}(\alpha(G_{n,p}) = k) = \tilde{O}(n^{1+3\gamma/2})$$

for any sequence $k = k(n)$.

Note that we proved Conjecture 16 for $\gamma = -1$ (i.e., $p = 1/n$) in Section 5.

- We conclude by noting that the first moment argument using augmented independent sets given in Section 3.2 can be adapted in straightforward way to give the following upper bound on the independence number of $G_{n,p}$ for $p \leq n^{-2/3}$.

Theorem 17. If $\omega(1/n) < p \leq n^{-2/3}$, then we have

$$\alpha(G_{n,p}) \leq k_x - \Omega\left(\frac{\log(np)^2}{p^2 n}\right)$$

with high probability, where k_x is the largest integer such that $\binom{n}{k_x} (1-p)^{\binom{k_x}{2}} > 1$.

This improvement can be extended to give a small linear improvement on the best known upper bound on $\alpha(G_{n,p})$, where $p = c/n$ and c is a constant. In fact, a first moment argument using maximal independent sets is sufficient to give a linear improvement on the upper bound in this context. As the precise magnitude of the optimal improvement requires a lengthy calculation, we exclude this in the interest of brevity.

7. Appendix: two-point concentration of $\alpha(G_{n,p})$ for $n^{-1/2+\epsilon} < p < 1/\log(n)^2$

Here, we show that the second moment method applied to the random variable Y_k , which is the number of maximal independent sets of size k , suffices to establish two point concentration of $\alpha(G_{n,p})$ for $n^{-1/2+\epsilon} < p < 1/\log(n)^2$.

Recall that X_k is the random variable which counts the number of independent sets of size k in $G_{n,p}$, and we defined k_x to be the largest integer such that

$$\mathbb{E}(X_{k_x}) > n^{2\epsilon}. \tag{25}$$

Set $X = X_{k_x}$ and $Y = Y_{k_x}$.

Theorem 18. Consider $p = p(n)$ such that $n^{-1/2+\epsilon} < p < 1/\log(n)^2$. Then $\alpha(G_{n,p}) \in \{k_x, k_x + 1\}$ whp.

Proof. Set $k = k_x$. First, we show by the standard first moment method that no independent set of size $k + 2$ appear whp:

$$\begin{aligned} \mathbb{P}(\alpha(G_{n,p}) \geq k + 2) &\leq \binom{n}{k + 2} (1 - p)^{\binom{k+2}{2}} \\ &= \binom{n}{k + 1} (1 - p)^{\binom{k+1}{2}} \left(\frac{n - k - 1}{k + 2} (1 - p)^{k+1} \right) \\ &= (1 + o(1)) \binom{n}{k + 1} (1 - p)^{\binom{k+1}{2}} \frac{n}{k} \left(\frac{k}{ne} \right)^2 \\ &= o\left(\frac{k}{n^{1/2}}\right) \\ &= o(1), \end{aligned}$$

where we use the fact that $k = o(n^{1/2})$ when $p > n^{-1/2+\epsilon}$ in the last line.

Next, we show that an independent set of size k exists whp. Here, we use the second moment method, but the variance of X itself is too large. So we work with the random variable Y which counts the number of *maximal* independent sets of size k . The first moment calculation is straightforward, using Equation (5):

$$\begin{aligned} \mathbb{E}(Y) &= \binom{n}{k} (1 - p)^{\binom{k}{2}} (1 - (1 - p)^k)^{n-k} \\ &= \binom{n}{k} (1 - p)^{\binom{k}{2}} e^{-O(k^2/n)} \end{aligned} \tag{26}$$

$$\begin{aligned} &= \mathbb{E}(X)(1 - o(1)) \\ &= \omega(1). \end{aligned} \tag{27}$$

By Chebyshev’s inequality, all that is left to show is that $\frac{\text{Var}(Y)}{\mathbb{E}(Y)^2} = o(1)$. We write Y as the sum of indicator variables Y_S , over all sets S such that $|K| = k$, of the random variable Y_S which is the indicator for the event that S is a maximal independent vertex set. We define indicator variables X_S for not necessarily maximal independent sets analogously. Then we have

$$\text{Var}(Y) = \sum_S \text{Var}(Y_S) + \sum_{S \neq T} \text{Cov}(Y_S, Y_T) \leq \mathbb{E}(Y) + \sum_{S \neq T} \mathbb{E}(Y_S Y_T).$$

Next, define, for all relevant i :

$$\begin{aligned} f_i &= \frac{1}{\mathbb{E}(X)^2} \sum_{|S \cap T|=i} \mathbb{E}(X_S X_T) = \frac{\binom{k}{i} \binom{n-k}{k-i}}{\binom{n}{k}} (1 - p)^{-\binom{i}{2}}, \\ g_i &= \frac{1}{\mathbb{E}(Y)^2} \sum_{|S \cap T|=i} \mathbb{E}(Y_S Y_T), \text{ and} \\ \kappa_i &= \frac{f_{i+1}}{f_i}. \end{aligned}$$

Since $\mathbb{E}(Y) = \omega(1)$, our goal is to show that

$$\sum_{i=0}^{k-1} g_i = o(1).$$

We consider three ranges of i : $i < \frac{\epsilon}{2}k_x$, $\frac{\epsilon}{2}k_x \leq i \leq (1 - \epsilon)k_x$, and $i > (1 - \epsilon)k_x$ (where ϵ is the constant in the statement of the Theorem). For each range, we show that the sum of variables g_i within that range is $o(1)$. For the first two ranges, we will actually work with variables f_i instead, noting that, since $Y_S \leq X_S$ and by similar reasoning to Equation (27), $g_i \leq f_i(1 + o(1))$.

First, consider $i < \frac{\epsilon}{2}k$. Using the explicit formula for f_i we have

$$\kappa_i = \frac{(k - i)^2(1 - p)^{-i}}{(i + 1)(n - 2k + i + 1)} < \frac{k^2(1 - p)^{-\epsilon k/2}}{n} = (1 + o(1)) \frac{k^2}{n} \left(\frac{ne}{k}\right)^\epsilon = O\left(\frac{\log(n)^2}{n^\epsilon}\right) = o(1).$$

As $f_2 = O(k^4/n^2) = o(1)$ then $\sum_{i=2}^{\epsilon k/2} f_i$ is a geometric sum with leading term and ratio in $o(1)$.

Next, consider the range $\frac{\epsilon}{2}k \leq i \leq (1 - \epsilon)k$. We have

$$\begin{aligned} f_i &\leq 2^k \left(\frac{k}{n}\right)^i (1 - p)^{-\binom{i}{2}} \\ &\leq \left[2^{k/i} \frac{k}{n} \left((1 - p)^{-k/2}\right)^{i/k}\right]^i \\ &\leq \left[(1 + o(1))2^{2/\epsilon} \frac{k}{n} \left(\frac{ne}{k}\right)^{1-\epsilon}\right]^i \\ &\leq \left[(1 + o(1))2^{2/\epsilon} e^{1-\epsilon} \left(\frac{k}{n}\right)^\epsilon\right]^i. \end{aligned} \tag{28}$$

Hence, the sum of variables f_i over $\frac{\epsilon}{2}k_x \leq i \leq (1 - \epsilon)k_x$ is $o(1)$.

The range where $i > (1 - \epsilon)k$ is where we use the variables g_i instead of f_i . Given two vertex sets S and T with intersection size i , and given that they are both independent, consider the probability that both are maximal independent sets. This probability is bounded above by the probability that each vertex in $S \setminus T$ is adjacent to at least one vertex in $T \setminus S$, which is equal to $(1 - (1 - p)^{k-i})^{k-i} \leq ((k - i)p)^{k-i}$. Therefore, we have

$$\begin{aligned} g_i &\leq (1 + o(1)) \frac{\binom{k}{i} \binom{n-k}{k-i}}{\binom{n}{k}} (1 - p)^{-\binom{i}{2}} ((k - i)p)^{k-i} \\ &= \frac{1 + o(1)}{\mathbb{E}[X]} \left(\binom{k}{k-i} (k - i)^{k-i} \right) \left(\binom{n-k}{k-i} (1 - p)^{\binom{k}{2} - \binom{i}{2}} p^{k-i} \right) \\ &< \frac{1 + o(1)}{\mathbb{E}[X]} (ke)^{k-i} n^{k-i} (1 - p)^{(1-\epsilon)k(k-i)} p^{k-i} \\ &< \frac{1}{\mathbb{E}[X]} \left((1 + o(1))eknp \left(\frac{k}{ne}\right)^{2(1-\epsilon)} \right)^{k-i} \\ &< \frac{1}{\mathbb{E}[X]} \left(22 \log(n)^3 n^{-\epsilon+2\epsilon^2} \right)^{k-i}, \end{aligned}$$

so $\sum_{i > (1-\epsilon)k} g_i = o(1)$, as desired. □

Acknowledgments. We thank the anonymous referee for useful comments and suggestions.

Competing interest. The authors have no competing interest to declare.

Funding statement. This work was supported in part by a grant from the Simons Foundation (587088, TB).

References

- [1] D. Achlioptas and A. Naor, ‘The two possible values of the chromatic number of a random graph’, *Annals of Mathematics* **162** (2005), 1335–1351.
- [2] N. Alon and M. Krivelevich, ‘The concentration of the chromatic number of random graphs’, *Combinatorica* **17** (1997), 303–313.
- [3] T. Bohman and J. Hofstad, ‘A critical probability for biclique partition of $G_{n,p}$ ’, *Journal of Combinatorial Theory B* **166** (2024), 50–79.
- [4] B. Bollobás, *Random Graphs*, second edn. (Cambridge University Press, 2001).
- [5] B. Bollobás and P. Erdős, ‘Cliques in random graphs’, in *Mathematical Proceedings of the Cambridge Philosophical Society*, vol. 80 (Cambridge University Press, 1976), 419–427.
- [6] A. Coja-Oghlan and C. Efthymios, ‘On independent sets in random graphs’, *Random Structures & Algorithms* **47** (2015), 436–486.
- [7] A. Coja-Oghlan, K. Panagiotou and A. Steger, ‘On the chromatic number of random graphs’, *Journal of Combinatorial Theory, Series B* **98**(5) (2008), 980–993.
- [8] V. Dani and C. Moore, ‘Independent sets in random graphs from the weighted second moment method’, in *Lecture Notes in Computer Science: Approximation, Randomization and Combinatorial Optimization*, vol. 6845 (Springer, 2011), 472–482.
- [9] A. Frieze, ‘On the independence number of random graphs’, *Discrete Mathematics* **81**(2) (1990), 171–175.
- [10] A. Frieze and M. Karoński, *Introduction to Random Graphs* (Cambridge University Press, 2016).
- [11] R. Glebov, A. Liebenau and T. Szabó, ‘On the concentration of the domination number of the random graph’, *SIAM Journal on Discrete Mathematics* **29** (2015), 1186–1206.
- [12] A. Heckel, ‘Non-concentration of the chromatic number of a random graph’, *Journal of the American Mathematical Society* **34**(1) (2021), 245–260.
- [13] A. Heckel and O. Riordan, ‘How does the chromatic number of a random graph vary?’ *Journal of the London Mathematical Society* **108**(5) (2023), 1769–1815.
- [14] S. Janson, T. Łuczak and A. Ruciński, *Random Graphs* (John Wiley & Sons, 2000).
- [15] R. Karp and M. Sipser, ‘Maximum matchings in sparse random graphs’, in *Proceedings of the 22nd Symposium on the Foundations of Computing* (IEEE, 1981), 364–375.
- [16] T. Łuczak, ‘A note on the sharp concentration of the chromatic number of random graphs’, *Combinatorica* **11** (1991), 295–297.
- [17] D. Matula, ‘On complete subgraphs of a random graph’, in *Proceedings of the Second Chapel Hill Conference on Combinatory Mathematics and its Applications* (University of North Carolina at Chapel Hill, 1970), 356–369.
- [18] A. Sah and M. Sawhney. Personal communication.
- [19] E. Shamir and J. Spencer, ‘Sharp concentration of the chromatic number on random graphs $G_{n,p}$ ’, *Combinatorica* **7**(1) (1987), 121–129.
- [20] E. Surya and L. Warnke, ‘On the concentration of the chromatic number of random graphs’, (2022), Preprint, [arXiv:2201.00906](https://arxiv.org/abs/2201.00906).