Regular induced subgraphs of a random graph

Michael Krivelevich* Benny Sudakov† Nicholas Wormald ‡

Abstract

An old problem of Erdős, Fajtlowicz and Staton asks for the order of a largest induced regular subgraph that can be found in every graph on n vertices. Motivated by this problem, we consider the order of such a subgraph in a typical graph on n vertices, i.e., in a binomial random graph $G(n, 1/2)$. We prove that with high probability a largest induced regular subgraph of $G(n, 1/2)$ has about $n^{2/3}$ vertices.

1 Introduction

A rather old and apparently quite difficult problem of Erdős, Fajtlowicz and Staton (see [4] or [3], page 85) asks for the order of a largest induced regular subgraph that can be found in every graph on n vertices. By the known estimates for graph Ramsey numbers (c.f., e.g., [5]), every graph on n vertices contains a clique or an independent set of size $c \ln n$, for some positive constant $c > 0$, providing a trivial lower bound of $c \ln n$ for the problem. Erdős, Fajtlowicz and Staton conjectured that the quantity in question is $\omega(\log n)$. So far this conjecture has not been settled. Some progress has been achieved in upper bounding this function of n: Bollobás in an unpublished argument showed (as stated in [3]) the existence of a graph on n vertices without an induced regular subgraph on at least $n^{1/2+\epsilon}$ vertices, for any fixed $\epsilon > 0$ and sufficiently large n. A slight improvement has recently been obtained by Alon and the first two authors [1], who took the upper bound down to $cn^{1/2} \log^{3/4} n$.

Given the simplicity of the problem's statement, its appealing character and apparent notorious difficulty, it is quite natural to try and analyze the behavior of this graph theoretic parameter for a typical graph on n vertices, i.e. a graph drawn from the probability space $G(n, 1/2)$ of graphs. (Recall that the ground set of the probability space $G(n, p)$ is composed of all graphs on n labeled vertices, where each unordered pair $\{i, j\}$ appears as an edge in G, drawn from $G(n, p)$, independently and with probability p. In the case $p = 1/2$ all labeled graphs G on n vertices are equiprobable: $Pr[G] = 2^{-\binom{n}{2}}$.) This is the subject of the present paper.

*School of Mathematical Sciences, Tel Aviv University, Tel Aviv 69978, Israel. E-mail: krivelev@post.tau.ac.il.
Research supported in part by a USA-Israel BSF Grant, by a grant from the Israel Science Foundation, and by a Pazy Memorial Award.

†Department of Mathematics, UCLA 90095, Los Angeles, USA. E-mail: bsudakov@math.ucla.edu.
Research supported in part by NSF CAREER award DMS-0812005 and a USA-Israeli BSF grant.

‡Department of Combinatorics and Optimization, University of Waterloo, Waterloo ON, Canada. E-mail: nwormald@uwaterloo.ca.
Supported by the Canada Research Chairs Program and NSERC.

We say that a graph property $\mathcal{P}$ holds with high probability, or *whp* for brevity, if the probability of a random graph to have $\mathcal{P}$ tends to 1 as n tends to infinity. It was shown by Cheng and Fang [2] that the random graph $G(n, 1/2)$ *whp* contains no induced regular subgraph on $cn/\log n$ vertices. We improve the upper bound, and give a nearly matching lower bound, as follows.

Theorem 1.1 *Let G be a random graph $G(n, 1/2)$. Then with high probability every induced regular subgraph of G has at most $2n^{2/3}$ vertices. On the other hand, for $k = o(n^{2/3})$, with high probability G contains a set of k vertices that span a $(k-1)/2$-regular graph.*

It is instructive to compare this result with the above mentioned result of [1]. Alon et al. also used a certain probability space of graphs to derive their upper bound of $O(n^{1/2}\log^{3/4} n)$. Yet, their model of random graphs is much more heterogeneous in nature (the expected degrees of vertices vary significantly there, see [1] for full details). As expected, the rather homogeneous model $G(n, 1/2)$ produces a sizably weaker upper bound for the Erdős-Fajtlowicz-Staton problem.

The difficult part of our proof is the lower bound, for which we use the second moment method. The main difficulty lies in getting an accurate bound on the variance of the number of d-regular graphs on k vertices, where $d = (k-1)/2$. Our main tool for achieving this goal is an estimate on the number $N(k, H)$ of d-regular graphs on k vertices which contain a given subgraph H, when H is not too large. Provided H has $o(\sqrt{k})$ vertices, and its degree sequence satisfies some conditions which are quite typical for random graphs, we obtain an asymptotic formula for $N(k, H)$ which is of independent interest; see Theorem 5.1.

In Section 2 we introduce some notation and technical tools utilized in our arguments, and then prove a rather straightforward upper bound of Theorem 1.1. A much more delicate lower bound is then proven in Section 3. The technical lemma used in this proof relies on the above-mentioned estimate of $N(k, H)$. The proof of this estimate is relegated to Section 4. The final section of the paper contains some concluding remarks.

2 Notation, tools and the upper bound

In this short section we describe some notation and basic tools to be used later in our proofs. Then we establish the upper bound part of Theorem 1.1.

We will utilize the following (standard) asymptotic notation. For two functions $f(n)$, $g(n)$ of a natural number n, we write $f(n) = o(g(n))$, whenever $\lim_{n\to\infty} f(n)/g(n) = 0$; $f(n) = \omega(g(n))$ if $g(n) = o(f(n))$. Also, $f(n) = O(g(n))$ if there exists a constant $C > 0$ such that $f(n) \leq Cg(n)$ for all n; $f(n) = \Omega(g(n))$ if $g(n) = O(f(n))$, and $f(n) = \Theta(g(n))$ if both $f(n) = O(g(n))$ and $f(n) = \Omega(g(n))$. We write $f \sim g$ if the ratio f/g tends to 1 when the underlying parameter tends to infinity. For a real x and positive integer a, define $[x]_a = x(x-1)\cdots(x-a+1)$. All logarithms in this paper have the natural base. We will use the bound $\binom{n}{k} \leq (en/k)^k$, valid for all positive n and k.

For a positive integer k, let $m(k)$ be the largest even integer not exceeding $(k-1)/2$.

Let $G(\mathbf{d})$ denote the number of labeled simple graphs on k vertices with degree sequence $\mathbf{d} =$

$(d_1, d_2, \ldots, d_k)$, where the degree of vertex i is d_i. Also, we denote

$$p_k = \mathbb{P}[\text{a random graph } G(k, 0.5) \text{ is } m(k)\text{-regular}] .$$

Clearly, $p_k = G(\mathbf{d})2^{-\binom{k}{2}}$, with all d_i being set equal to $m(k)$.

We will repeatedly cite the following corollary of a result of McKay and the third author (see Theorems 2 and 3 of [6]).

Theorem 2.1 *Let $d_j = d_j(k)$, $1 \le j \le k$ be integers such that $\sum_{j=1}^k d_j = \lambda k(k-1)$ is an even integer where $1/3 < \lambda < 2/3$, and $|\lambda k - d_j| = O(k^{1/2+\epsilon})$ uniformly over j, for some sufficiently small fixed $\epsilon > 0$. Then*

$$G(\mathbf{d}) = f(\mathbf{d})\big(\lambda^\lambda(1-\lambda)^{1-\lambda}\big)^{\binom{k}{2}} \prod_{j=1}^{k} \binom{k-1}{d_j} \tag{1}$$

where

- *$f(\mathbf{d}) = O(1)$, and*
- *if $\max\{|\lambda k - d_j|\} = o(\sqrt{k})$, then $f(\mathbf{d}) \sim \sqrt{2}e^{1/4}$, uniformly over the choice of such a degree sequence $\mathbf{d}$.*

It is a routine matter to check that $\big(\lambda^\lambda(1-\lambda)^{1-\lambda}\big)^{\binom{k}{2}} \prod_{j=1}^{k} \binom{k-1}{d_j} = O\left(2^{-\binom{k}{2}}\binom{k-1}{m(k)}^k\right)$. (One way to see this is as follows. We can show that the expression increases if the smallest d_i is increased, unless they are all at least $(k-1)/2+O(1)$. By a symmetrical argument, we may assume they have an upper bound of the same form. Then, for such a sequence, adding bounded numbers to all d_i changes the expression by a bounded factor.) Hence $G(\mathbf{d})2^{-\binom{k}{2}}$ is $O(p_k)$ for every degree sequence $\mathbf{d}$ covered by Theorem 2.1. Also, using Stirling's formula it is straightforward to verify that $p_k = \big((1+o(1))\sqrt{\pi k/2}\big)^{-k}$ and that $p_{k-1}/p_k = \Theta(\sqrt{k})$.

In order to prove the upper bound in Theorem 1.1, we show that, for a given k and r, the probability that a random graph on k vertices is r-regular is $O(p_k)$. (For future use we prove here a somewhat more general statement.) We then use the above-mentioned estimate for p_k and apply the union bound over all possible values of r.

Lemma 2.2 *For every degree sequence $\mathbf{d} = (d_1, \ldots, d_k)$,*

$$\mathbb{P}[G(k, 0.5) \text{ has degree sequence } \mathbf{d}] = O(p_k) .$$

Proof. Following the remark after Theorem 2.1, assume that $G(\mathbf{d})2^{-\binom{k}{2}} \le Cp_k$ for every degree sequence $\mathbf{d}$ covered by the theorem, where $C > 0$ is an absolute constant. Let $\mathbf{d}$ be a degree sequence of length k for which $G(\mathbf{d})$ is maximal (which is obviously equivalent to choosing $\mathbf{d}$ to be a most probable degree sequence in $G(k, 1/2)$). Write $\mathbb{P}[G(k, 0.5)$ has degree sequence $\mathbf{d}] = a_k p_k$. If all degrees in $\mathbf{d}$ satisfy $|d_i - k/2| \le k^{1/2+\epsilon}$, then we can apply Theorem 2.1. Otherwise, there is d_i, say, d_k, deviating from $k/2$ by at least $k^{1/2+\epsilon}$, for some fixed $\epsilon > 0$. To bound the probability that $G(k, 1/2)$ has degree sequence $\mathbf{d}$, we first expose the edges from vertex k to the rest of the graph. By standard estimates on the tails of the binomial distribution, the probability that k has the required degree is $\exp\{-\Omega(k^{2\epsilon})\}$. The edges exposed induce a new degree sequence on vertices $1, \ldots, k-1$. The probability that the random graph $G(k-1, 0.5)$ has this degree sequence is at most $a_{k-1}p_{k-1}$ in our notation. We thus obtain: $a_k p_k \leq \max(Cp_k, a_{k-1}p_{k-1} \cdot O(\exp\{-k^{2\epsilon}\}))$. Recalling that $p_{k-1}/p_k = \Theta(\sqrt{k})$, we obtain that $a_k \leq \max(C, a_{k-1} \cdot O(\exp\{-k^{\epsilon}\}))$ for large enough k. The result follows by induction starting from k_0 for which the above factor $O(\exp\{-k^{\epsilon}\})$ is less than 1; the final bound is $a_k \leq \max(C, a_{k_0})$. ∎

In order to complete the proof of the upper bound of Theorem 1.1, note that by Lemma 2.2 the probability that a fixed set V_0 of k vertices spans a regular subgraph in $G(n, 1/2)$ is $O(kp_k)$. Summing over all $k \geq k_0 = 2n^{2/3}$ and all vertex subsets of size k, we conclude that the probability that $G(n, 1/2)$ contains an induced regular subgraph on at least k_0 vertices is

$$\sum_{k \geq k_0} \binom{n}{k} \cdot O(kp_k) \leq \sum_{k \geq k_0} \left(\frac{en}{k}\right)^k k\big((1+o(1))\sqrt{\pi k/2}\big)^{-k} \leq n^2 \cdot \left(\frac{(1+o(1))\sqrt{2}en}{\sqrt{\pi}k_0^{3/2}}\right)^{k_0} = o(1)\,.$$

3 A lower bound

In this section we give a proof of the lower bound in our main result, Theorem 1.1. (To be more accurate, we give here most of the proof, deferring the proof of a key technical lemma to the next section.) The proof uses the so-called second moment method and proceeds by estimating carefully the first two moments of the random variable $X = X(k)$, counting the number of $(k-1)/2$-regular induced subgraphs on k vertices in $G(n, 1/2)$. For convenience we assume throughout the proof that $k \equiv 1 \pmod 4$. (Since this estimate is used for proving the lower bound of Theorem 1.1, we can allow ourselves to choose k in such a way without losing essentially anything in the lower bound.) It is somewhat surprising to be able to apply successfully the second moment method to sets of such a large size, however two earlier instances of similar application can be found in [7] and [9].

So let X be the random variable counting the number of $(k-1)/2$-regular induced subgraphs on k vertices in $G(n, 0.5)$. We write $X = \sum_{|A|=k} X_A$, where X_A is the indicator random variable for the event that a vertex subset A spans a $(k-1)/2$-regular subgraph. Then

$$\mathbb{E}[X] = \sum_{|A|=k} \mathbb{E}[X_A] = \binom{n}{k} p_k\,.$$

Plugging in the estimate for p_k cited after the statement of Theorem 2.1, it is straightforward to verify that $\mathbb{E}[X]$ tends to infinity for $k = o(n^{2/3})$; in fact, $\mathbb{E}[X] = (\omega(1))^k$ in this regime. A corollary of Chebyshev's inequality is that $\mathbb{P}[X > 0] \geq 1 - (\mathbb{V}ar[X]/\mathbb{E}^2[X])$, and therefore in order to prove that *whp* $G(n, 1/2)$ contains an induced regular subgraph on k vertices, it is enough to establish that $\mathbb{V}ar[X] = o(\mathbb{E}^2[X])$.

In order to estimate the variance of X we need to estimate the correlation between the following events: "A spans a $(k-1)/2$-regular subgraph" and "B spans a $(k-1)/2$-regular subgraph", where A, B are k-element vertex subsets whose intersection is of size $i \geq 2$. To this end, define

$$p_{k,i} = \max_{|V(H)|=i} \mathbb{P}[G(k, 0.5) \text{ is } (k-1)/2\text{-regular} \mid G[i] = H],$$

where the maximum in the expression above is taken over all graphs H on i vertices, and $G[i]$ stands for the subgraph of $G(k,1/2)$ spanned by the first i vertices. Since $X = \sum_{|A|=k} X_A$, we have:

$$\begin{aligned}
\mathbb{V}ar[X] &= \mathbb{E}[X^2] - \mathbb{E}^2[X] = \sum_{|A|=k} \mathbb{V}ar[X_A] + \sum_{i=2}^{k-1} \sum_{\substack{|A|=|B|=k \\ |A\cap B|=i}} \Big(\mathbb{E}[X_A X_B] - \mathbb{E}[X_A]\mathbb{E}[X_B]\Big) \\
&\le \sum_{|A|=k} \mathbb{E}[X_A] + \sum_{i=2}^{k-1} \sum_{\substack{|A|=|B|=k \\ |A\cap B|=i}} \Big(\mathbb{P}[X_A=1]\mathbb{P}[X_B=1|X_A=1] - \mathbb{P}[X_A=1]\mathbb{P}[X_B=1]\Big) \\
&\le \mathbb{E}[X] + \binom{n}{k} p_k \cdot \sum_{i=2}^{k-1} \binom{k}{i}\binom{n-k}{k-i}(p_{k,i} - p_k) \ . \qquad (2)
\end{aligned}$$

As a warm-up, we first show that a rather crude estimate for (2) suffices to prove that $\mathbb{V}ar[X] = o(\mathbb{E}^2[X])$ for $k = o(\sqrt{n})$. We start with the following bound for $p_{k,i}$.

Lemma 3.1 *For* $2 \le i \le k-1$,

$$p_{k,i} = O\left(\binom{k-i}{\lfloor \frac{k-i}{2} \rfloor}^i 2^{-(k-i)i} p_{k-i} \right) \ .$$

Also, $\frac{p_{k,i}}{p_k} \le C e^{k \log \frac{k}{k-i}}$, *for a sufficiently large constant* $C > 0$.

Proof. First, given H, expose the edges from H to the remaining $k-i$ vertices (denote the latter set by X). For every $v \in H$, we require $d(v,X) = (k-1)/2 - d_H(v)$. This happens with probability

$$\binom{k-i}{(k-1)/2 - d_H(v)} 2^{-k+i} \le \binom{k-i}{\lfloor \frac{k-i}{2} \rfloor} 2^{-k+i}$$

(the middle binomial coefficient is the largest one). Hence the probability that all i vertices from $V(H)$ have the required degree of $(k-1)/2$ in G is at most the i-th power of the right hand side of the above expression.

Now, conditioned on the edges from H to X, we ask what is the probability that the subgraph spanned by X has the required degree sequence (each $v \in X$ should have exactly $(k-1)/2 - d(v,H)$ neighbors in X). Observe that by Lemma 2.2 the probability that $G[X]$ has the required degree sequence is at most $C_0 p_{k-i}$ for some absolute constant $C_0 > 0$, providing the first claimed estimate for $p_{k,i}$.

From Theorem 2.1, $p_t = \Theta\left(2^{-2\binom{t}{2}} \binom{t-1}{\lfloor (t-1)/2 \rfloor}^t\right)$. Therefore, the ratio $p_{k,i}/p_k$ can be estimated as follows:

$$\begin{aligned}
\frac{p_{k,i}}{p_k} &\leq C_0\binom{k-i}{\lfloor\frac{k-i}{2}\rfloor}^i 2^{-(k-i)i}\frac{p_{k-i}}{p_k} \leq C_1\binom{k-i}{\lfloor\frac{k-i}{2}\rfloor}^i 2^{-(k-i)i}\frac{2^{-2\binom{k-i}{2}}\binom{k-i-1}{\lfloor\frac{k-i-1}{2}\rfloor}^{k-i}}{2^{-2\binom{k}{2}}\binom{k-1}{\frac{k-1}{2}}^k} \\
&\leq C_1\left(\frac{\binom{k-i}{\lfloor\frac{k-i}{2}\rfloor}2^{-(k-i)}}{\binom{k-1}{\frac{k-1}{2}}2^{-(k-1)}}\right)^k \leq C\left[\frac{(\frac{k-1}{2})^2(\frac{k-1}{2}-1)^2\cdots(\frac{k-i}{2}+1)^2}{(k-1)(k-2)\cdots(k-i+1)}2^{i-1}\right]^k \\
&= C\left[\frac{k-1}{k-2}\frac{k-3}{k-4}\cdots\frac{k-i+2}{k-i+1}\right]^k \\
&\leq C\exp\left\{\Big(\frac{1}{k-2}+\frac{1}{k-4}+\cdots+\frac{1}{k-i+1}\Big)k\right\}\ .
\end{aligned}$$

In the third inequality above we used that $\binom{k-i-1}{\lfloor\frac{k-i-1}{2}\rfloor} \leq \frac{1}{2}(1+\frac{1}{k-i})\binom{k-i}{\lfloor\frac{k-i}{2}\rfloor}$. Observe that $\sum_{j=k-i+1}^{k-2}\frac{1}{j} < \int_{k-i}^{k}\frac{dx}{x} = \log\frac{k}{k-i}$. This completes the proof of the second part of the lemma. ∎

Now we complete a proof of a weaker version of the lower bound of Theorem 1.1, by showing that *whp* $G(n, 1/2)$ contains an induced $(k-1)/2$-regular subgraph on $k = o(\sqrt{n})$ vertices. Omitting the term $-p_k$ in the sum in (2) and using $\mathbb{E}[X] = \binom{n}{k}p_k$, we obtain:

$$\frac{\mathbb{V}ar[X]}{\mathbb{E}^2[X]} \leq \frac{\sum_{i=2}^{k-1}\binom{k}{i}\binom{n-k}{k-i}p_{k,i}}{\binom{n}{k}p_k} + \frac{1}{\mathbb{E}[X]} = \sum_{i=2}^{k-1}\frac{\binom{k}{i}\binom{n-k}{k-i}}{\binom{n}{k}}\frac{p_{k,i}}{p_k} + o(1). \tag{3}$$

Denote

$$g(i) = \frac{\binom{k}{i}\binom{n-k}{k-i}}{\binom{n}{k}}\frac{p_{k,i}}{p_k}\,.$$

Let us first estimate the ratio of the binomial coefficients involved in the definition of $g(i)$.

$$\begin{aligned}
\frac{\binom{k}{i}\binom{n-k}{k-i}}{\binom{n}{k}} &\leq \frac{\left(\frac{ek}{i}\right)^i\binom{n}{k-i}}{\binom{n}{k}} = \left(\frac{ek}{i}\right)^i\frac{k(k-1)\cdots(k-i+1)}{(n-k+i)(n-k+i-1)\cdots(n-k+1)} \\
&\leq \left(\frac{ek}{i}\right)^i\left(\frac{k}{n-k+i}\right)^i \leq \left(\frac{3k^2}{in}\right)^i\,.
\end{aligned}$$

To analyze the asymptotic behavior of $g(i)$, we consider three cases.

Case 1. $i \leq k/2$**.** In this case, by Lemma 3.1 and the inequality $\log(1+x) \leq x$ for $x \geq 0$ we have:

$$\frac{p_{k,i}}{p_k} \leq Ce^{k\log\frac{k}{k-i}} = Ce^{k\log(1+\frac{i}{k-i})} \leq Ce^{k\frac{i}{k-i}} \leq Ce^{2i}\,.$$

We thus get the following estimate for $g(i)$:

$$g(i) = \frac{\binom{k}{i}\binom{n-k}{k-i}}{\binom{n}{k}}\frac{p_{k,i}}{p_k} \leq \left(\frac{3k^2}{in}\right)^i \cdot Ce^{2i} \leq C\left(\frac{3e^2k^2}{in}\right)^i\,. \tag{4}$$

The above inequality is valid for all values of k. When $k = o(\sqrt{n})$ it gives that $g(i) = (o(1))^i$.

Case 2. $k/2 \le i \le k - \frac{k}{\log k}$**.** Recalling Lemma 3.1 again, we have $p_{k,i}/p_k \le Ce^{k \log \frac{k}{k-i}} \le Ce^{k \log\log k}$. Hence in this case

$$\begin{aligned} g(i) &\le \left(\frac{3k^2}{in}\right)^i Ce^{k\log\log k} \le \left(\frac{6k}{n}\right)^{k/2} Ce^{k\log\log k} \\ &\le \left(\frac{1}{\sqrt{k}}\right)^{k/2} Ce^{k\log\log k} = Ce^{-\frac{k\log k}{4}+k\log\log k} \le e^{-k}. \end{aligned}$$

For future reference, it is important to note here that in the calculation above we used $6k/n \le k^{-1/2}$. This inequality stays valid as long as $k \le (n/6)^{2/3}$.

Case 3. $i \ge k - \frac{k}{\log k}$**.** In this case it suffices to use the trivial estimate $p_{k,i} \le 1$. We also need that $\mathbb{E}[X] = \binom{n}{k} p_k = (\omega(1))^k$. Therefore,

$$g(i) = \frac{\binom{k}{i}\binom{n-k}{k-i} p_{k,i}}{\binom{n}{k} p_k} \le \frac{\binom{k}{i}\binom{n-k}{k-i}}{(\omega(1))^k} \le \frac{2^k n^{k-i}}{(\omega(1))^k} \le \frac{2^k n^{k/\log k}}{(\omega(1))^k} = \frac{e^{O(k)}}{(\omega(1))^k} \le e^{-k}.$$

In the above calculation we used the assumption $\log n = O(\log k)$. In the complementary case $k = n^{o(1)}$ the expression $\binom{n}{k} p_k$ behaves like $\left(cn/k^{3/2}\right)^k \ge n^{k/2}$, while the numerator in the expression for $g(i)$ is at most $2^k n^{k/\log k} = n^{o(k)}$, and the estimate works as well. Note that, as in Case 2, the inequality here remains valid even for k as large as $n^{2/3}$.

It thus follows that $\sum_{i=2}^{k-1} g(i)$ is negligible, implying in turn that $\mathbb{V}ar[X] = o(\mathbb{E}^2[X])$, and thus X is with high probability positive by the Chebyshev inequality.

Now we proceed to the proof of the "real" lower bound of Theorem 1.1, i.e. assume that k satisfies $k = o(n^{2/3})$. In this case estimating the variance of the random variable X, defined as the number of induced $(k-1)/2$-regular subgraphs on k vertices, becomes much more delicate. We can no longer ignore the term $-p_k$ in the sum in (2). Instead, we show that for small values of i in this sum $p_{k,i}$ is asymptotically equal to p_k. In words, this means that knowing the edges spanned by the first i vertices of a random graph $G = G(k, 1/2)$ does not affect by much the probability of G being $(k-1)/2$-regular. We claim this formally for $i = o(\sqrt{k})$ in the following key lemma.

Lemma 3.2 *For* $i = o(\sqrt{k})$,

$$p_{k,i} = (1 + o(1)) p_k.$$

The proof of this lemma is rather involved technically. We thus postpone it to the next section. We now show how to complete the proof assuming its correctness. We first repeat estimate (2):

$$\begin{aligned} \frac{\mathbb{V}ar[X]}{\mathbb{E}^2[X]} &\le \frac{1}{\mathbb{E}[X]} + \sum_{i=2}^{k-1} \frac{\binom{k}{i}\binom{n-k}{k-i}(p_{k,i} - p_k)}{\binom{n}{k} p_k} \\ &\le o(1) + \sum_{i=2}^{t} \frac{\binom{k}{i}\binom{n-k}{k-i}(p_{k,i} - p_k)}{\binom{n}{k} p_k} + \sum_{i=t}^{k-1} \frac{\binom{k}{i}\binom{n-k}{k-i}}{\binom{n}{k}} \frac{p_{k,i}}{p_k}, \end{aligned} \tag{5}$$

where $t = t(k,n)$ is chosen so that $t = \omega(k^2/n)$ but $t = o(\sqrt{k})$. Since $k = o(n^{2/3})$ such a function is easily seen to exist. Due to our choice of t we can apply Lemma 3.2 to the first sum above. It thus follows that

$$\sum_{i=2}^{t} \frac{\binom{k}{i}\binom{n-k}{k-i}(p_{k,i} - p_k)}{\binom{n}{k} p_k} = \sum_{i=2}^{t} \frac{\binom{k}{i}\binom{n-k}{k-i} \cdot o(p_k)}{\binom{n}{k} p_k} \le o(1) \cdot \frac{\sum_{i=0}^{k} \binom{k}{i}\binom{n-k}{k-i}}{\binom{n}{k}} = o(1)\,.$$

As for the second sum in (5) we can utilize the same case analysis as done before for $k = o(\sqrt{n})$. The only difference is in Case 1, that now covers all i from t till $k/2$. Therefore, for every i in this new interval we can use inequality (4) to conclude

$$\frac{\binom{k}{i}\binom{n-k}{k-i}}{\binom{n}{k}} \frac{p_{k,i}}{p_k} \le C\left(\frac{3e^2k^2}{in}\right)^i \le C\left(\frac{3e^2k^2}{tn}\right)^i = (o(1))^i\,.$$

This completes the proof of Theorem 1.1. ∎

4 Proof of key lemma

The proof of Lemma 3.2 is overall along the lines of the proof of Lemma 3.1, though requiring a much more detailed examination of the probabilities involved. Throughout this section, let k be odd and, for simplicity, denote $(k-1)/2$ by d. Let $\mathcal{D}_i$ be the set of integer vectors $\mathbf{d} = (d_1, \ldots, d_i)$ such that $0 \le d_j \le i-1$ for $1 \le j \le i$, and $kd - \sum_j d_j$ is even. Given $\mathbf{d} \in \mathcal{D}_i$, let $N(\mathbf{d})$ denote the number of graphs G on vertex set $[k]$ for which $G[i]$ has no edges, and $d_G(j) = d - d_j$ for $j \in [i]$, whilst $d_G(j) = d$ for $i < j \le k$. Note that if $\mathbf{d}$ is the degree sequence of a graph H on vertex set $[i]$, then $N(\mathbf{d})$ is the number of d-regular graphs G on vertex set $[k]$ for which $G[i] = H$. Of course, in this case $N(\mathbf{d})$ is nonzero only if kd is even, and hence k is congruent to 1 mod 4.

Proposition 4.1 *Assume $i = o(\sqrt{k})$. Given $\mathbf{d} \in \mathcal{D}_i$ and a nonnegative vector $\mathbf{s} = (s_1, \ldots, s_i)$ with $\sum_j s_j$ even, put $\mathbf{d}' = \mathbf{d} - \mathbf{s}$. Then, uniformly over such $\mathbf{d}$ and $\mathbf{s}$ with the additional properties that $\mathbf{d}' \in \mathcal{D}_i$ and $\sum_{j=1}^{i} s_j \le k^{3/4}$,*

$$\frac{N(\mathbf{d})}{N(\mathbf{d}')} \sim \frac{\displaystyle\prod_{j=1}^{i} \binom{d - d_j + s_j}{s_j}}{\displaystyle\prod_{j=1}^{i} \binom{d + 1 - (i - d_j)}{s_j}}.$$

Proof. We use a comparison type argument. Since it is quite complicated, we give the idea of the proof first. For any vector $\mathbf{c} = (c_1, \ldots, c_j)$, write $\mathbf{c}^*$ for the vector $(d - c_1, \ldots, d - c_j)$. Let $V_1 = \{1, \ldots, i\}$ and $V_2 = \{i+1, \ldots, k\}$. For simplicity, suppose that $s_1 = s_2 = 1$, and $s_j = 0$ for $j \ge 3$. We can compute $N(\mathbf{d})$ as the number of possible outcomes of two steps. The first step is to choose a bipartite graph B with bipartition (V_1, V_2) and degree sequence $\mathbf{d}^*$ in V_1. The second step is to add the remaining edges between vertices in V_2 such that those vertices will have degree d. By comparison, to count $N(\mathbf{d}')$ we choose in the first step B' with degree sequence $\mathbf{d}'^*$ in V_1, and then do the second step for each such

B'. The proof hinges around the fact that there is a correspondence between the set of possible B and B' such that the number of ways of performing the second step is roughly the same, at least for most of the corresponding pairs (B, B').

The correspondence is many-to-many. For a graph B we may add two edges, incident with vertices 1 and 2, to obtain a graph B'. The number of ways this can be done, without creating multiple edges, is $\prod_{j=1}^{2}\big(k-i-(d-d_j)\big) = \prod_{j=1}^{2}\big(d+1-(i-d_j)\big)$. Conversely, each B' comes from $\prod_{j=1}^{2}(d-d_j+1)\big)$ different B. The ratio of these quantities gives the asymptotic ratio between $N(\mathbf{d})$ and $N(\mathbf{d}')$ claimed in the theorem, ignoring the number of ways of performing the second step. Our actual argument gets more complicated because not only some bipartite graphs must be excluded, but also some sets of edges to be added to them. So we will present equations relating to the above argument in a slightly different form to make exclusion of various terms easier.

Let $\mathcal{B}$ denote the set of bipartite graphs with bipartition (V_1, V_2). For $B \in \mathcal{B}$, write $\mathbf{D}_j(B)$ for the degree sequence of B on the vertices in V_j in the natural order, so $\mathbf{D}_1(B) = \big(d_B(1), \ldots, d_B(i)\big)$ and $\mathbf{D}_2(B) = \big(d_B(i+1), \ldots, d_B(k)\big)$. For $\mathbf{d} = (d_1, \ldots, d_i)$ with $kd - \sum_j d_j$ even, let $\mathcal{B}(\mathbf{d}^*)$ denote $\{B \in \mathcal{B} : \mathbf{D}_1(B) = \mathbf{d}^*\}$. Let $G\big(\mathbf{D}_2(B)^*\big)$ denote the number of graphs with degree sequence $\mathbf{D}_2(B)^*$. Clearly,

$$N(\mathbf{d}) = \sum_{B\in\mathcal{B}(\mathbf{d}^*)} G\big(\mathbf{D}_2(B)^*\big). \tag{6}$$

Suppose that we wish to add to B a set S of edges joining V_1 and V_2, without creating any multiple edges, such that the degree of $j \in V_1$ $(1 \le j \le i)$ in the graph induced by S is s_j (as given in the statement of the proposition). The family of all such sets S will be denoted by $\mathcal{S}(B, \mathbf{s})$. Note that necessarily $|S| \le k^{3/4}$ for $S \in \mathcal{S}(B, \mathbf{s})$. The cardinality of $\mathcal{S}(B, \mathbf{s})$ is $\prod_{j=1}^{i}\binom{d+1-(i-d_j)}{s_j}$, because $d_B(j) = d_j^* = d - d_j$, so (as in the sketch above) j has $d + 1 - (i - d_j)$ spare vertices in V_2 to which it may be joined. Hence we can somewhat artificially rewrite (6) as

$$N(\mathbf{d}) = \frac{1}{\displaystyle\prod_{j=1}^{i}\binom{d+1-(i-d_j)}{s_j}} \sum_{B\in\mathcal{B}(\mathbf{d}^*)}\ \sum_{S\in\mathcal{S}(B,\mathbf{s})} G\big(\mathbf{D}_2(B)^*\big). \tag{7}$$

Also for $B' \in \mathcal{B}(\mathbf{d}'^*)$ define $\mathcal{S}'(B', \mathbf{s})$ to be the family of sets $S \subseteq E(B')$ such that the degree of $j \in V_1$ in the graph induced by S is s_j $(1 \le j \le i)$. Since $d_{B'}(j) = d - d'_j$ and $d'_j = d_j - s_j$, a similar argument gives

$$N(\mathbf{d}') = \frac{1}{\displaystyle\prod_{j=1}^{i}\binom{d-d_j+s_j}{s_j}} \sum_{B'\in\mathcal{B}(\mathbf{d}'^*)}\ \sum_{S\in\mathcal{S}'(B',\mathbf{s})} G\big(\mathbf{D}_2(B')^*\big). \tag{8}$$

The rest of the proof consists of showing that the significant terms in the last two equations can be put into 1-1 correspondence such that corresponding terms are asymptotically equal.

We first need to show that for a typical $B \in \mathcal{B}(\mathbf{d}^*)$, the variance of the elements of $\mathbf{D}_2(B)$ (as a sequence) is small. Given $\mathbf{d}$, define $\bar{d} = (k-i)^{-1}\sum_{j\in V_1}(d - d_j)$, and note that this is equal to $(k-i)^{-1}\sum_{j\in V_2} d_B(j)$ for every $B \in \mathcal{B}(\mathbf{d}^*)$. Since $\bar{d}$ is determined uniquely by $\mathbf{d}$, the value of $\bar{d}$ is the same for all $B \in \mathcal{B}(\mathbf{d}^*)$.

Lemma 4.2 *Let $\mathbf{d} \in \mathcal{D}_i$, and select B uniformly at random from $\mathcal{B}(\mathbf{d}^*)$. Then*

$$\mathbb{E}\left(\sum_{j\in V_2} \big(\bar{d} - d_B(j)\big)^2\right) \le i(k-i).$$

Proof. First observe that in B, the neighbours of any vertex $t \in V_1$ form a random subset of V_2 of size d_t^*, and these subsets are independent for different t. So for fixed $j \in V_2$, $d_B(j)$ is distributed as a sum of i independent 0-1 variables with mean $\sum_{t\in V_1} d_t^*/(k-i) = \bar{d}$. It follows that the variance of $d_B(j)$ is less than i. Hence $\mathbb{E}\big(\bar{d} - d_B(j)\big)^2 < i$, and the lemma follows by linearity of expectation. ∎

Returning to the proof of the proposition, we will apply Theorem 2.1 to estimate $G\big(\mathbf{D}_2(B)^*\big)$. This graph has $k-i$ vertices, degree sequence $\{d - d_B(j), j \in V_2\}$, and its degree sum is $e^* := (k-i)d - e(B)$, where $e(B) = \sum_{j\in V_2} d_B(j) = (k-i)\bar{d}$ is the number of edges in the bipartite graph B. Consider λ from Theorem 2.1. We see that

$$\lambda = \lambda(\mathbf{d}) := \frac{e^*}{(k-i)(k-i-1)} = \frac{d-\bar{d}}{k-i-1}. \tag{9}$$

The product of binomials in (1) is in this case

$$\prod_{j\in V_2} \binom{k-i-1}{d-d_B(j)}. \tag{10}$$

For every $B \in \mathcal{B}(\mathbf{d}^*)$, all components of the vector $\mathbf{D}_2(B)$ are at most $|V_1| = i$. Thus

$$x_j := \frac{k-i-1}{2} - (d - d_B(j)) = d_B(j) - i/2 = O(i) = o(\sqrt{k}).$$

We have

$$\binom{a}{a/2+x} = \binom{a}{\lfloor a/2 \rfloor} \exp\big(-2x^2/a + O(x^3/a^2)\big) \tag{11}$$

for $x = o(\sqrt{a})$, which may be established for instance by analyzing the ratio of the binomial coefficients. Hence

$$\prod_{j=i+1}^{k} \binom{k-i-1}{d-d_B(j)} = \binom{k-i-1}{\lfloor \frac{k-i-1}{2} \rfloor}^{k-i} \exp\left(\frac{-2\sum x_j^2}{k-i} + o(i)\right). \tag{12}$$

(Note that here and in the rest of the proof, the asymptotic relations hold uniformly over $\mathbf{d} \in \mathcal{D}_i$.) Since $i = o(\sqrt{k})$ we can choose a function ω of n such that $\omega \to \infty$ and $\omega^2 i = o(\sqrt{k})$. Define $\hat{\mathcal{B}}_\omega(\mathbf{d}^*)$ to be the subset of $\mathcal{B}(\mathbf{d}^*)$ that contains those B for which

$$\sum_{j\in V_2} \big(\bar{d} - d_B(j)\big)^2 \le \omega i(k-i). \tag{13}$$

Since $\sum_{j\in V_2} d_B(j) = e(B)$ is the same for all bipartite graphs $B \in \mathcal{B}(\mathbf{d}^*)$, by definition of x_j we have that $\sum_j x_j^2 - \sum_j d_B^2(j)$ also does not depend on B. Similarly, the sum in (13) differs from $\sum_j d_B^2(j)$ by a constant independent of B. Therefore $\sum_j x_j^2$ for all $B \in \hat{\mathcal{B}}_{\omega/2}(\mathbf{d}^*)$ is smaller than the corresponding sum for $B \in \mathcal{B}(\mathbf{d}^*) \setminus \hat{\mathcal{B}}_\omega(\mathbf{d}^*)$ by an additive term of at least $\omega i(k-i)/2$. This implies that the product of binomials in (12) is larger, for all $B \in \hat{\mathcal{B}}_{\omega/2}(\mathbf{d}^*)$, than for any $B \in \mathcal{B}(\mathbf{d}^*) \setminus \hat{\mathcal{B}}_{\omega}(\mathbf{d}^*)$. Also, from Lemma 4.2 and Markov's inequality, almost all members of $\mathcal{B}(\mathbf{d}^*)$ are in $\hat{\mathcal{B}}_{\omega/2}(\mathbf{d}^*)$. Moreover, since all degrees in degree sequence $\mathbf{D}_2(B)^*$ deviate from $(k-i)/2$ by at most $O(i) = o(\sqrt{k})$, the function $f(\mathbf{D}_2(B)^*)$ from Theorem 2.1 is $\sim \sqrt{2}e^{1/4}$ for all $B \in \mathcal{B}(\mathbf{d}^*)$. Combining these observations, we conclude that the contribution to (6) from $B \notin \hat{\mathcal{B}}_{\omega}(\mathbf{d}^*)$ is $o\big(N(\mathbf{d})\big)$. Thus, the same observation holds for (7). That is,

$$N(\mathbf{d}) \sim \frac{1}{\displaystyle\prod_{j=1}^{i} \binom{d+1-(i-d_j)}{s_j}} \sum_{B \in \hat{\mathcal{B}}_{\omega}(\mathbf{d}^*)} \sum_{S \in \mathcal{S}(B,\mathbf{s})} G\big(\mathbf{D}_2(B)^*\big). \tag{14}$$

We also note for later use, that by (13) and Cauchy's inequality, for all $B \in \hat{\mathcal{B}}_{\omega}(\mathbf{d}^*)$

$$\sum_{j \in V_2} \big|\bar{d} - d_B(j)\big| \le (k-i)\sqrt{\omega i}. \tag{15}$$

Fix $B \in \hat{\mathcal{B}}_{\omega}(\mathbf{d}^*)$. Consider S chosen uniformly at random from $\mathcal{S}(B,\mathbf{s})$, and let $r_m(S)$ denote the number of edges of S incident with a vertex $m \in V_2$. Fixing m and using that $|S| \le k^{3/4}$, we can bound the probability that $r_m(S) \ge 5$ by

$$\sum_{\{j_1,\dots,j_5\} \subseteq V_1} \prod_{t=1}^{5} \frac{s_{j_t}}{k-i-(d-d_{j_t})} \le \left(\sum_{j \in V_1} \frac{s_j}{d-i}\right)^5 = \left(\frac{|S|}{d-i}\right)^5 = O(k^{-5/4}).$$

Hence by Markov's inequality, with probability $1 - O(k^{-1/4})$, $S \in \mathcal{S}(B,\mathbf{s})$ satisfies

(i) $\max_{j \in V_2} r_j(S) \le 4$.

We would next like to bound $\sum_{j \in V_2} \big|\bar{d} - d_B(j)\big| r_j(S)$. To do this, note that we may choose the edges in S incident with any given vertex sequentially, each time selecting a random neighbour from those vertices of V_2 still eligible to be joined to. For each such edge joining to such a random vertex $j \in V_2$, by (13) there are, as a crude bound, at least $(k-i)/3$ vertices of V_2 to choose from (for k sufficiently large). Amongst the eligible vertices, the average value of $\big|\bar{d} - d_B(j)\big|$ must be at most $3\sqrt{\omega i}$ by (15). Hence, if X_h denotes the value of $\big|\bar{d} - d_B(j)\big|$ for the h-th edge added, we have $\mathbb{E}X_h \le 3\sqrt{\omega i}$. Thus $\mathbb{E}\sum_h X_h \le 3\sqrt{\omega i}|S| \le 3\sqrt{\omega i}k^{3/4}$. Noting that $\sum_h X_h = \sum_{j \in V_2} \big|\bar{d} - d_B(j)\big| r_j(S)$ and using Markov's inequality, we deduce that almost all $S \in \mathcal{S}(B,\mathbf{s})$ (more precisely all except the fraction $3/\sqrt{\omega} = o(1)$ of them, at most) satisfy

(ii) $\displaystyle\sum_{j \in V_2} \big|\bar{d} - d_B(j)\big| r_j(S) \le \omega\sqrt{i}k^{3/4}$.

Define $\hat{\mathcal{S}}(B,\mathbf{s})$ to be the set of $S \in \mathcal{S}(B,\mathbf{s})$ satisfying both the properties (i) and (ii). Then, since each $S \in \mathcal{S}(B,\mathbf{s})$ contributes equally to (14),

$$N(\mathbf{d}) \sim \frac{1}{\displaystyle\prod_{j=1}^{i} \binom{d+1-(i-d_j)}{s_j}} \sum_{B \in \hat{\mathcal{B}}_{\omega}(\mathbf{d}^*)} \sum_{S \in \hat{\mathcal{S}}(B,\mathbf{s})} G\big(\mathbf{D}_2(B)^*\big). \tag{16}$$

Let $B \in \hat{\mathcal{B}}_\omega(\mathbf{d}^*)$ and $S \in \hat{\mathcal{S}}(B, \mathbf{s})$. Then, using (13) together with (i) and (ii), we get

$$\begin{aligned} \sum_{j \in V_2} \big(\bar{d} - d_B(j) - r_j(S)\big)^2 &\le \omega i(k-i) + 2 \sum_{j \in V_2} \big|\bar{d} - d_B(j)\big| r_j(S) + \sum_{j \in V_2} r_j(S)^2 \\ &\le \omega i(k-i) + 2\omega\sqrt{i}k^{3/4} + 16(k-i) \sim \omega ik. \end{aligned}$$

Hence, for k sufficiently large, those S appearing in the range of the summation in (16) satisfy $B + S \in \hat{\mathcal{B}}_{2\omega}(\mathbf{d}'^*)$, where $B + S$ is the graph obtained by adding the edges in S to B (and noting that $\mathbf{d}'^* = \mathbf{d}^* + \mathbf{s}$). Since, as we saw, the contribution to (6) from $B \notin \hat{\mathcal{B}}_\omega(\mathbf{d}^*)$ is $o\big(N(\mathbf{d})\big)$, we may also relax the constraint on B in the summation in (16), to become $B \in \hat{\mathcal{B}}_{2\omega}(\mathbf{d}^*)$. Now redefining 2ω as ω, we obtain

$$N(\mathbf{d}) \sim \frac{1}{\displaystyle\prod_{j=1}^{i} \binom{d+1-(i-d_j)}{s_j}} \sum_{(B,S) \in \mathcal{W}} G\big(\mathbf{D}_2(B)^*\big), \tag{17}$$

where $\mathcal{W}$ denotes the set of all (B, S) such that $B \in \hat{\mathcal{B}}_\omega(\mathbf{d}^*)$, $S \in \hat{\mathcal{S}}(B, \mathbf{s})$ and $B + S \in \hat{\mathcal{B}}_\omega(\mathbf{d}'^*)$.

Define $\hat{\mathcal{S}}'(B', \mathbf{s})$, analogous to $\hat{\mathcal{S}}(B, \mathbf{s})$, to be the set of $S \in \mathcal{S}'(B', \mathbf{s})$ with maximum degree in V_2 at most 5 and also obeying property (ii) above, where $B = B' - S$. Since B has density close to 1/2, B and its complement are more or less equivalent, and adding or deleting edges should have a similar effect. Indeed, the above argument applied to (8), with suitable small modification, gives

$$N(\mathbf{d}') \sim \frac{1}{\displaystyle\prod_{j=1}^{i} \binom{d-d_j+s_j}{s_j}} \sum_{(B',S) \in \mathcal{W}'} G\big(\mathbf{D}_2(B')^*\big) \tag{18}$$

where $\mathcal{W}'$ denotes the set of all (B', S) such that $B' \in \hat{\mathcal{B}}_\omega(\mathbf{d}'^*)$, $S \in \hat{\mathcal{S}}(B', \mathbf{s})$ and $B' - S \in \hat{\mathcal{B}}_\omega(\mathbf{d}^*)$.

Observe that $(B, S) \in \mathcal{W}$ if and only if $(B + S, S) \in \mathcal{W}'$. So the summation in (18) is equal to

$$\sum_{(B,S) \in \mathcal{W}} G\big(\mathbf{D}_2(B+S)^*\big).$$

Hence, comparing with (17), the proposition follows if we show that

$$G\big(\mathbf{D}_2(B)^*\big) \sim G\big(\mathbf{D}_2(B+S)^*\big) \tag{19}$$

uniformly for all $(B, S) \in \mathcal{W}$.

We may apply (1) to both sides of (19). Write $g(\lambda, n) = \big(\lambda^\lambda (1-\lambda)^{1-\lambda}\big)^{\binom{n}{2}}$. Notice that $\lambda(\mathbf{d})$, as defined in (9), satisfies:

$$\lambda(\mathbf{d}) = \frac{\sum_{j \in V_2}(d - d_B(j))}{(k-i)(k-i-1)}$$

— which is exactly λ for the degree sequence $\mathbf{D}_2(B)^*$ as defined in Theorem 2.1. The same applies to $\lambda(\mathbf{d}')$ and the degree sequence $\mathbf{D}_2(B + S)^*$. Using

$$\sum_{j \in V_2}(d - d_B(j)) = (k-i)d - \sum_{t \in V_1} d_B(t) = (k-i)d - id + \sum_{t=1}^{i} d_t = \frac{(k-2i)(k-1)}{2} + \Theta(i^2),$$

it is easy to derive from (9) that both $\lambda(\mathbf{d})$ and $\lambda(\mathbf{d}')$ are $1/2 + O(i^2/k^2) = 1/2 + o(k^{-1})$ for all $\mathbf{d}, \mathbf{d}' \in \mathcal{D}_i$. For such λ the derivative of $\log g(\lambda, k-i)$ is $\binom{k-i}{2}\log(\lambda/(1-\lambda)) = O(k^2 i^2/k^2) = o(k)$. Moreover, $|\lambda(\mathbf{d}) - \lambda(\mathbf{d}')| = O(k^{-5/4})$ because the values of $\bar{d}$ for B and $B+S$ differ by $O(|S|/k)$. Hence $g(\lambda(\mathbf{d}), k-i) \sim g(\lambda(\mathbf{d}'), k-i)$. It is now also easy to see that $f(\mathbf{D}_2(B)^*), f(\mathbf{D}_2(B+S)^*)$ from Theorem 2.1 satisfy: $f(\mathbf{D}_2(B)^*) \sim f(\mathbf{D}_2(B+S)^*) \sim \sqrt{2}e^{1/4}$, since all degrees in these two degree sequences deviate from $(k-i)/2$ by $O(i) = o(\sqrt{k-i})$.

It only remains to consider the product of binomials in the two sides of (19) after the application of Theorem 2.1. Recalling the expression (10), the ratio of these two products is

$$\prod_{j\in V_2} \binom{k-i-1}{d-d_B(j)} \Big/ \binom{k-i-1}{d-d_B(j)-r_j(S)}.$$

Since $B \in \hat{\mathcal{B}}_\omega(\mathbf{d}^*)$, all $r_j(S) \le 4$ and $\sum_j r_j(S) \le k^{3/4}$, so using $d = (k-1)/2$ and $d_B(j) \le i$, this expression is, up to a multiplicative factor of $1 + O(k^{-1})\sum_j r_j^2(S) = 1 + o(1)$, equal to

$$\begin{aligned}\prod_{j\in V_2}\left(\frac{k-i-1-d+d_B(j)}{d-d_B(j)}\right)^{r_j(S)} &= \prod_{j\in V_2}\left(\frac{d-i/2-\big(i/2-d_B(j)\big)}{d-i/2+\big(i/2-d_B(j)\big)}\right)^{r_j(S)} \\ &= \prod_{j\in V_2}\left(1+O\left(\frac{i/2-d_B(j)}{k}\right)\right)^{r_j(S)}.\end{aligned}$$

By its definition, $\bar{d} = (k-i)^{-1}\sum_{j\in V_1}(d-d_j) = i/2 + O(i^2/k) = i/2 + o(1)$, and so using condition (ii) (the right hand side of which is $\omega\sqrt{i}k^{3/4} = o(k)$), and not forgetting $\sum r_j(S) \le k^{3/4}$, we get

$$\sum_{j\in V_2} r_j(S)\frac{|i/2-d_B(j)|}{k} \le \sum_{j\in V_2}\frac{|\bar{d}-d_B(j)|r_j(S)}{k} + \sum_{j\in V_2}\frac{|\bar{d}-i/2|r_j(S)}{k} = o(1).$$

Hence, the expression above is asymptotic to 1. This argument shows that (19) holds with the required uniformity. ∎

For a slightly simpler version of the formula in Proposition 4.1, put

$$\hat{d} = d - \frac{1}{2}(i-1) = \frac{1}{2}(k-i)$$

(which is in some sense the average degree of vertices of side V_1 in the bipartite graph B) and

$$\delta_j = d_j - \frac{1}{2}(i-1).$$

Then the proposition gives

$$\frac{N(\mathbf{d})}{N(\mathbf{d}')} \sim \prod_{j=1}^{i}\frac{[\hat{d}-\delta_j+s_j]_{s_j}}{[\hat{d}+\delta_j]_{s_j}}.$$

Recalling that δ_j and s_j are at most $i = o(\sqrt{k})$ and using $\log(1+x) = x - x^2/2 + O(x^3)$, we have

$$\begin{aligned}[][\hat{d}-\delta_j+s_j]_{s_j} &= \hat{d}^{s_j}\exp\big(-\delta_j s_j/\hat{d} + s_j^2/2\hat{d} + o(1/\sqrt{k})\big),\\ [\hat{d}+\delta_j]_{s_j} &= \hat{d}^{s_j}\exp\big(\delta_j s_j/\hat{d} - s_j^2/2\hat{d} + o(1/\sqrt{k})\big).\end{aligned}$$

Thus, we may rewrite the assertion of Proposition 4.1 as

$$\frac{N(\mathbf{d})}{N(\mathbf{d}')} \sim \exp\left\{\sum_{j=1}^{i}\big(-2\delta_j s_j/\hat{d} + s_j^2/\hat{d}\big)\right\} \sim \frac{\exp\left\{\sum_{j=1}^{i}(\delta_j - s_j)^2/\hat{d}\right\}}{\exp\left\{\sum_{j=1}^{i}\delta_j^2/\hat{d}\right\}}. \tag{20}$$

To proceed, we extend this formula so that $\mathbf{s}$ is permitted to have negative entries.

Corollary 4.3 *Assume $i = o(\sqrt{k})$. Given $\mathbf{d} \in \mathcal{D}_i$ and an integer vector $\mathbf{s} = (s_1, \dots, s_i)$ with $\sum_j s_j$ even, put $\mathbf{d}' = \mathbf{d} - \mathbf{s}$. Then, uniformly over such $\mathbf{d}$ and $\mathbf{s}$ with the additional properties that $\mathbf{d}' \in \mathcal{D}_i$ and $\sum_{j=1}^{i} |s_j| \le k^{3/4}$,*

$$\frac{N(\mathbf{d})}{N(\mathbf{d}')} \sim \exp\left\{\frac{1}{\hat{d}}\sum_{j=1}^{i}\big(-2\delta_j s_j + s_j^2\big)\right\}.$$

Proof. Define the vector $\mathbf{s}'$ by turning the negative entries of $\mathbf{s}$ into 0; that is, the jth entry of $\mathbf{s}'$ is s_j if $s_j \ge 0$, and 0 otherwise. (If this causes the sum of entries to change parity, leave one of these entries as -1. It is easy to modify the following proof accordingly.) Let $\mathbf{s}'' = \mathbf{s}' - \mathbf{s}$. The jth entry of $\mathbf{s}''$ is $-s_j$ if $s_j < 0$, and 0 otherwise. We can now estimate the product

$$\frac{N(\mathbf{d})}{N(\mathbf{d} - \mathbf{s}')} \cdot \frac{N(\mathbf{d} - \mathbf{s}')}{N(\mathbf{d}')} = \frac{N(\mathbf{d})}{N(\mathbf{d} - \mathbf{s}')} \Big/ \frac{N(\mathbf{d}')}{N(\mathbf{d}' - \mathbf{s}'')}$$

using two applications of (20), noting that all entries of $\mathbf{s}'' = \mathbf{s}' - \mathbf{s}$ are nonnegative and that the δ_j' defined for degree sequence $\mathbf{d}'$ equals $\delta_j - s_j$. ∎

Define $\mathbf{d}_0$ to be the constant sequence of length i, all of whose entries are $\lfloor (i-1)/2 \rfloor$. If $kd - i\lfloor (i-1)/2 \rfloor$ is odd, adjust the first entry to $\lfloor (i-1)/2 \rfloor + 1$ to ensure that $\mathbf{d}_0 \in \mathcal{D}_i$. We can use the following result to compare the number of graphs with an arbitrary degree sequence $\mathbf{d}$ on $G[i]$ to the number with $\mathbf{d}_0$. Recall, however, that $N(\mathbf{d})$ is defined even if $\mathbf{d}$ is not the degree sequence of any graph on $[i]$.

Corollary 4.4 *(i) If $\mathbf{d} \in \mathcal{D}_i$ then $N(\mathbf{d}) \le N(\mathbf{d}_0)\big(1 + o(1)\big)$.*
(ii) If, in addition, $\sum_{j=1}^{i} \delta_j^2 = o(k)$, then $N(\mathbf{d}) \sim N(\mathbf{d}_0)$.

Proof. We treat the case that the first entry of $\mathbf{d}_0$ was not adjusted for the parity reason above; in the other case, only trivial modifications are required, for which we omit the details. Part (ii) of the corollary follows immediately from Corollary 4.3 by putting $s_j = \lceil \delta_j \rceil$ for each j, since if $\sum_{j=1}^{i} \delta_j^2 = o(k)$ then by Cauchy's inequality $\sum |\delta_j| = o(\sqrt{ik}) = o(k^{3/4})$. (Note also that $\hat{d} \sim k/2$.)

For the first part, let $\mathbf{d}$ maximise $N(\mathbf{d})$. If $\sum_{j=1}^{i} \delta_j^2 < k$ say, the above argument shows that $N(\mathbf{d}) \le N(\mathbf{d}_0)\big(1 + o(1)\big)$. So assume that $\sum_{j=1}^{i} \delta_j^2 > k$. Putting $s_j = \lceil \delta_j \rceil$ and applying Corollary 4.3 shows the result, provided $\sum |\lceil \delta_j \rceil| \le k^{3/4}$. If the latter condition fails, we can simply define $s_j = \alpha_j \delta_j$ for some $0 \le \alpha_j \le 1$ such that $\sum_j |s_j|$ is just below $k^{3/4}$ and is even. Since $|s_j| \le |\delta_j|$ and they both have the same sign, we can conclude that $\sum_j (-2\delta_j s_j + s_j^2) \le -\sum_j s_j^2$. By Cauchy's inequality, the sum of the squares of s_j grows asymptotically faster than k. Let $\mathbf{s} = (s_1, \dots, s_i)$ and let $\mathbf{d}' = \mathbf{d} - \mathbf{s}$.

Then, by Corollary 4.3 we obtain $N(\mathbf{d}) = o(N(\mathbf{d}'))$, which contradicts the maximality assumption and proves the result. ∎

Define $\mathbf{d}_1$ to be the constant sequence of length k, all of whose entries are $d = (k-1)/2$. We can now determine the asymptotic value of $N(\mathbf{d}_0)$. Recall that k is odd; we will now assume that $k \equiv 1 \pmod 4$ to ensure that $N(\mathbf{d}_0)$ is not 0. (If $k \equiv 3 \pmod 4$, we could prove a similar result by subtracting 1 from the first entry of $\mathbf{d}_1$.

Corollary 4.5 *For $k \equiv 1$ (mod 4) we have $N(\mathbf{d}_0) \sim G(\mathbf{d}_1)/2^{\binom{i}{2}}$.*

Proof. Let H be one of the $2^{\binom{i}{2}}$ graphs on vertex set $[i]$ chosen at random, and let $\mathbf{d}_H = \{d_1, \dots, d_i\}$ be its degree sequence. Then d_j is a binomially distributed random variable with expectation $(i-1)/2$ and variance $(i-1)/4$. Hence for $\delta_j = d_j - (i-1)/2$ we have $\mathbb{E}[\delta_j^2] = \mathbb{V}ar[d_j] = (i-1)/4$. Then

$$\mathbb{E} \sum_{j \in [i]} \delta_j^2 \leq i^2 ,$$

and, by Markov's inequality, *whp* $\sum_{j \in [i]} \delta_j^2 = o(k)$. Thus, from Corollary 4.4(ii) it follows that for almost all graphs H, $N(\mathbf{d}_H) \sim N(\mathbf{d}_0)$. Part (i) of the same corollary shows that for all other graphs, $N(\mathbf{d}_H) \leq (1 + o(1))N(\mathbf{d}_0)$. Since $N(\mathbf{d}_H)$ is the number of d-regular graphs G on vertex set $[k]$ for which $G[i] = H$, we have that $G(\mathbf{d}_1) = \sum_H N(\mathbf{d}_H) = (1 + o(1))N(\mathbf{d}_0)2^{\binom{i}{2}}$, and the corollary follows. ∎

Proof of Lemma 3.2. From Corollary 4.4,

$$p_{k,i} \sim \mathbb{P}[G(k, 0.5) \text{ is } (k-1)/2\text{-regular} \mid G[i] = H] ,$$

where H is a chosen to be a graph with degree sequence $\mathbf{d}_0$. Note that the number of random edges outside H to be exposed is $\binom{k}{2} - \binom{i}{2}$, and each of them appears independently and with probability 1/2. Therefore, the above probability equals to $N(\mathbf{d}_0)/2^{\binom{k}{2}-\binom{i}{2}}$. By Corollary 4.5, this is asymptotic to $G(\mathbf{d}_1)/2^{\binom{k}{2}} = p_k$. ∎

5 Concluding remarks

Our technique for proving Proposition 4.1 is a rather complicated comparison argument somewhat related to the method of switchings used for graphs of similar densities in [8]. One might be tempted to try proving the result for $|S| = \sum_{j=1}^{i} s_j = 2$, as sketched in the first part of the proof, and then applying this repeatedly, as in the proof of Corollary 4.4, to go from one degree sequence to another. However, this seems to provide insufficient accuracy. Similarly, attempts to use switchings directly were not successful.

Of independent interest is the following estimate for the probability that a regular graph with k vertices and degree $(k-1)/2$ contains a given induced subgraph with degree sequence $(d_1, \dots, d_i)$ on its first i vertices. This gives an asymptotic formula provided the sum of the absolute values of $\delta_j = d_j - (i-1)/2$ is a bounded multiple of $k^{3/4}$, and otherwise gives less accurate bounds.

Theorem 5.1 *Assume $i = o(\sqrt{k})$, with $k \equiv 1 \pmod 4$. Let H be a graph on vertex set $[i]$ with degree sequence $\mathbf{d} = (d_1, \dots, d_i)$. Then the probability that a random $\frac{1}{2}(k-1)$-regular graph G on vertex set $[k]$ has the induced subgraph $G[i]$ equal to H is*

$$2^{-\binom{i}{2}} \exp\left(\frac{2}{k-i}\sum_{j=1}^{i} -\delta_j^2\right) \exp\left(o(k^{-3/4})\sum_{j=1}^{i}|\delta_i|\right)$$

where $\delta_j = d_j - (i-1)/2$.

Proof. We may use the argument in the proof of Corollary 4.4 to jump from $\mathbf{d}$ to $\mathbf{d}_0$, using at most $k^{-3/4}\sum_{j=1}^{i}|\delta_i|$ applications of Corollary 4.3, and then apply Corollary 4.5. ∎

Acknowledgement. We would like to thank the referees of the paper for their careful reading and many helpful remarks.

References

[1] N. Alon, M. Krivelevich and B. Sudakov, Large nearly regular induced subgraphs, *SIAM J. Discrete Math.* **22** (2008), 1325–1337.

[2] Q. Cheng and F. Fang, Kolmogorov random graphs only have trivial stable colorings, *Inform. Process. Lett.* **81** (2002), 133–136.

[3] F. R. K. Chung and R. L. Graham, **Erdős on Graphs: His Legacy of Unsolved Problems**, A. K. Peters, Ltd., Wellesley, MA, 1998.

[4] P. Erdős, On some of my favourite problems in various branches of combinatorics, *Fourth Czechoslovakian Symposium on Combinatorics, Graphs and Complexity (Prachatice, 1990)*, 69–79, Ann. Discrete Math., 51, North-Holland, Amsterdam, 1992.

[5] R. L. Graham, B. L. Rothschild and J. H. Spencer, **Ramsey Theory**, 2nd Edition, Wiley, New York, 1990.

[6] B. D. McKay and N. C. Wormald, Asymptotic enumeration by degree sequence of graphs of high degree, *Europ. J. Combinatorics* **11** (1990), 565–580.

[7] S. Janson, The numbers of spanning trees, Hamilton cycles and perfect matchings in a random graph, *Combin. Probab. Comput.* **3** (1994), 97–126.

[8] M. Krivelevich, B. Sudakov, V. Vu and N. C. Wormald, Random regular graphs of high degree, *Random Structures and Algorithms* **18** (2001), 346–363.

[9] O. Riordan, Spanning subgraphs of random graphs, *Combin. Probab. Comput.* **9** (2000), 125–148.