UPPER TAILS VIA HIGH MOMENTS AND ENTROPIC STABILITY

MATAN HAREL, FRANK MOUSSET, AND WOJCIECH SAMOTIJ

ABSTRACT. Suppose that X is a bounded-degree polynomial with nonnegative coefficients on the p-biased discrete hypercube. Our main result gives sharp estimates on the logarithmic upper tail probability of X whenever an associated extremal problem satisfies a certain entropic stability property. We apply this result to solve two long-standing open problems in probabilistic combinatorics: the upper tail problem for the number of arithmetic progressions of a fixed length in the p-random subset of the integers and the upper tail problem for the number of cliques of a fixed size in the random graph $G_{n,p}$. We also make significant progress on the upper tail problem for the number of copies of a fixed regular graph H in $G_{n,p}$. To accommodate readers who are interested in learning the basic method, we include a short, self-contained solution to the upper tail problem for the number of triangles in $G_{n,p}$ for all $p = p(n)$ satisfying $n^{-1} \log n \ll p \ll 1$.

CONTENTS

1. INTRODUCTION

Suppose that $Y = (Y_1, \ldots, Y_N)$ is a sequence of independent Bernoulli random variables with mean p and that $X = X(Y)$ is an N-variate polynomial with nonnegative real coefficients. Perhaps the simplest question that can be asked about the typical behaviour of X is whether it satisfies a law of large numbers, i.e., whether $X \to \mathbb{E}[X]$ in probability as $N \to \infty$. Once this is established, it is natural to ask for quantitative estimates of the probability of the event that X differs from its mean by a significant amount. In the special case where $Y \mapsto X(Y)$ is a linear function, this problem is addressed by the classical theory of large deviations, see [13, 25]. This theory shows that, under mild conditions on the coefficients of the linear function X and the assumptions $p \to 0$ and $Np \to \infty$,

$$\mathbb{P}\big(|X - \mathbb{E}[X]| \geqslant \delta\, \mathbb{E}[X]\big) = e^{-(I(\delta)+o(1))Np}$$

Date: April 13, 2019.

This research is supported in part by the Israel Science Foundation grants 1147/14 (FM and WS), 1028/16 (FM), and 1145/18 (FM and WS), the ERC Starting Grant 633509 (FM) and 678520 (MH), and the Zuckerman Postdoctoral Fellowship Program (MH).

for an explicitly computable function $I\colon (0,\infty) \to (0,\infty]$. However, there are many natural situations where one would like to consider nonlinear polynomials $X(Y)$, as in the following two examples. We use the notation $[\![N]\!] = \{1,\dots,N\}$.

Example 1.1 (Arithmetic progressions in random sets of integers)**.** A *k-term arithmetic progression* is a sequence of k integers of the form $\bigl(a, a+b, a+2b, \dots, a+(k-1)b\bigr)$, where $b > 0$. We write $[\![N]\!]_p$ for the random subset of $[\![N]\!]$ obtained by including every element of $[\![N]\!]$ independently with probability p. Let $X_{N,p}^{k\text{-AP}}$ denote the number of k-term arithmetic progressions in $[\![N]\!]_p$. Then $X_{N,p}^{k\text{-AP}}$ can be considered as a polynomial with nonnegative coefficients and total degree k in the independent $\mathrm{Ber}(p)$ random variables $Y_1,\dots,Y_N$, where Y_i is the indicator variable of the event that $i \in [\![N]\!]_p$; explicitly,

$$X_{N,p}^{k\text{-AP}} = \sum_{b>0} \sum_{\substack{a \geqslant 1 \\ a+(k-1)b \leqslant N}} \prod_{i=0}^{k-1} Y_{a+ib}.$$

We remark that, unlike [9] and several other works, we count only genuine arithmetic progressions (i.e., we do not consider degenerate progressions of the form $(a,\dots,a)$) and we count every progression only once (as opposed to counting $(a_1, a_2, \dots, a_k)$ and $(a_k, a_{k-1}, \dots, a_1)$ as two different progressions).

Example 1.2 (Subgraph counts in random graphs)**.** Fix a nonempty graph H and let $X_{n,p}^H$ be the number of copies of H in the random graph $G_{n,p}$. Then $X_{n,p}^H$ can be written as a polynomial with nonnegative coefficients and total degree e_H in the $N = \binom{n}{2}$ indicator random variables of the possible edges of $G_{n,p}$. More precisely, fix an arbitrary bijection $\sigma_n \colon \binom{[\![n]\!]}{2} \to [\![N]\!]$ (the precise choice is irrelevant) and, for every $i \in [\![N]\!]$, let Y_i be the indicator variable of the event that $\sigma_n^{-1}(i)$ is an edge in $G_{n,p}$. Then $Y_1,\dots,Y_N$ are independent $\mathrm{Ber}(p)$ random variables and

$$X_{n,p}^H = \sum_{\substack{H' \subseteq K_n \\ H' \cong H}} \prod_{e \in E(H')} Y_{\sigma_n(e)},$$

where K_n denotes the complete graph on the vertex set $[\![n]\!]$.

In this paper, we will always assume that δ is fixed and $p \to 0$ as $N \to \infty$.

The large deviation problem for the variables described above is significantly more involved than the linear case; in particular, the lower and upper tail probabilities—that is, $\mathbb{P}\bigl(X \leqslant (1-\delta)\,\mathbb{E}[X]\bigr)$ and $\mathbb{P}\bigl(X \geqslant (1+\delta)\,\mathbb{E}[X]\bigr)$, respectively—exhibit dramatically different behaviours. On the one hand, using a combination of Harris's inequality [35] and Janson's inequality [37], one can show that $X = X_{N,p}^{k\text{-AP}}$ satisfies

$$e^{-C_1(\delta)\min\{\mathbb{E}[X],Np\}} \leqslant \mathbb{P}\bigl(X \leqslant (1-\delta)\,\mathbb{E}[X]\bigr) \leqslant e^{-C_2(\delta)\min\{\mathbb{E}[X],Np\}} \tag{1}$$

for some positive $C_1(\delta)$ and $C_2(\delta)$.[1] Similar bounds are available for $X = X_{n,p}^H$. On the other hand, there are no comparably simple tools that allow one to easily obtain similar estimates on the logarithm of the upper tail probability. The standard concentration inequalities due to Azuma–Hoeffding [36], Talagrand [58], or Kim–Vu [46, 60] yield bounds that are far from tight in Examples 1.1 and 1.2. For a survey discussing these and other classical approaches to the 'infamous upper tail' problem, see [39].

Unlike the lower tail, the upper tail is susceptible to the influence of small structures whose appearance increases the value X atypically, a phenomenon that we refer to as *localisation*. For example, in the case of $X = X_{N,p}^{k\text{-AP}}$ where $k \geqslant 3$, a typical subset of size $m = o(N)$ contains $\Theta\bigl(N^2(m/N)^k\bigr) = o(m^2)$ k-term arithmetic progressions, whereas some very rare subsets (notably an interval of length m) can contain as many as $\Theta(m^2)$ such progressions. The event that $[\![N]\!]_p$

[1] For more precise results, we refer the interested reader to [42, 48, 52, 62].

contains an interval of length $\Theta(\sqrt{\mathbb{E}[X]})$ thus provides a lower bound on the upper tail probability. More precisely, $\mathbb{P}\big(X \geqslant (1+\delta)\,\mathbb{E}[X]\big) \geqslant \exp\big(-O(\sqrt{\mathbb{E}[X]}\log(1/p))\big)$, which is significantly larger than the lower tail probability (1) for most p. In order to properly analyse the upper tail event, one must account for these local effects, which frequently requires understanding the peculiar combinatorial nature of the random variable X.

The last decade has seen the development of an increasingly powerful theory of 'nonlinear large deviations', which began with the work of Chatterjee–Dembo [17] and was further developed by Eldan [27], Cook–Dembo [22], and Augeri [2, 3]. Whenever a general function f of i.i.d. random variables satisfies certain smoothness and complexity conditions, these results can be used to express the upper tail probability $\mathbb{P}\big(f \geqslant (1+\delta)\,\mathbb{E}[f]\big)$ in terms of an associated variational problem. In the case where f is a polynomial with nonnegative coefficients on the hypercube, this variational problem is able to capture the presence of localisation, if it occurs. In the two examples mentioned above, nonlinear large deviation theory gives tight control of the upper tail probabilities whenever $p \geqslant N^{-\alpha}$ for some constant $\alpha > 0$. However, the best-known constant α is not optimal in both examples.

Our main contribution is a general method for proving sharp bounds on the upper tail probability of the polynomial $X = X(Y)$ in the presence of localisation. In many cases where localisation occurs, our approach can also give a coarse description of the tail event. At the heart of our method lies an adaptation of the classical moment argument of Janson, Oleszkiewicz, and Ruciński [38], which we use to formalise the intution that the upper tail event is dominated by the appearance of near-minimisers of the combinatorial optimisation problem

$$\Phi_X(\delta) = \min\big\{|I|\log(1/p) : I \subseteq [\![N]\!] \text{ and } \mathbb{E}[X \mid Y_i = 1 \text{ for all } i \in I] \geqslant (1+\delta)\,\mathbb{E}[X]\big\}. \qquad (2)$$

Roughly speaking, we say that $I \subseteq [\![N]\!]$ is a *core* if it is a feasible set for the above optimisation problem, its size is $O\big(\Phi_X(\delta)\big)$, and it satisfies a certain natural rigidity condition (we will give a more precise definition in Section 3). We show that the upper tail probability is approximately equal to the probability of the appearance of a core. In particular, when the number of cores of size m is $(1/p)^{o(m)}$, a property we term *entropic stability*, then a union bound implies that $-\log\mathbb{P}\big(X \geqslant (1+\delta)\,\mathbb{E}[X]\big)$ is well-approximated by $\Phi_X(\delta)$. We will verify that the random variables $X_{N,p}^{k\text{-AP}}$ and $X_{n,p}^{H}$ (for a large class of graphs H) satisfy the entropic stability condition under optimal, or nearly optimal, assumptions on p.

One important caveat that we have ignored so far is that the upper tail exhibits localisation only when the expectation of X tends to infinity sufficiently quickly. In fact, if $\mathbb{E}[X]$ is of constant order, then, under relatively mild conditions, X converges in distribution to a Poisson random variable and no localisation occurs. We show that, for the two examples discussed above, the upper tail continues to have Poisson behaviour even when $\mathbb{E}[X]$ goes to infinity sufficiently slowly. In the cases of k-term arithmetic progressions in $[\![N]\!]_p$ and cliques in $G_{n,p}$, our results for the Poisson and localised regimes cover almost the whole range of densities $p \to 0$ with $\mathbb{E}[X] \to \infty$, leaving the upper tail probability undetermined only at densities for which it is believed that the two behaviours coexist.

1.1. **Arithmetic progressions in random sets of integers.**

Let $X = X_{N,p}^{k\text{-AP}}$ denote the number of k-term arithmetic progressions in $[\![N]\!]_p$. It is not hard to see that $\mathbb{E}[X] = \Theta(N^2p^k)$. Whenever this expectation vanishes, the upper tail event is commensurate with the probability of $X \geqslant 1$, which can be controlled using Markov's inequality. More generally, if $\mathbb{E}[X]$ is bounded, then it follows from standard techniques that X is asymptotically Poisson [7]; in this case, the large deviations of X are those of a Poisson random variable with mean $\mathbb{E}[X]$. For the remainder of this section, we shall thus assume that $\mathbb{E}[X] \to \infty$, i.e., that $p^{k/2} \gg N^{-1}$.

Improving an earlier estimate due to Janson and Ruciński [41], Warnke [61] proved that under fairly general assumptions (in particular, for constant $\delta > 0$ and all p bounded away from 1),

$$-\log \mathbb{P}(X \geqslant (1+\delta)\,\mathbb{E}[X]) = \Theta\left(\min\left\{\big((1+\delta)\log(1+\delta)-\delta\big)\,\mathbb{E}[X], \sqrt{\delta\,\mathbb{E}[X]}\log(1/p)\right\}\right), \quad (3)$$

where the constants implicit in the Θ-notation are independent of δ. Note that the two terms of the minimum correspond to the dominance of the Poisson and the localised regimes, respectively.

Since then, it has been an open problem to determine the missing constant factor in (3). Using the above-mentioned framework of Eldan [27], Bhattacharya, Ganguly, Shao, and Zhao [9] were able to do so in the range $N^{-\frac{1}{12(k-1)}}(\log N)^{O(1)} \ll p^{k/2} \ll 1$. This was subsequently improved by Briët–Gopi [14] to the slightly wider range $N^{-\frac{1}{12\lceil (k-1)/2\rceil}}(\log N)^{O(1)} \ll p^{k/2} \ll 1$, also using Eldan's result. The two theorems below improve significantly on these results and determine the precise rate of the upper tail for all $N^{-1} \ll p^{k/2} \ll 1$, excepting the case $p^{k/2} = \Theta(N^{-1}\log N)$. The first result concerns the range where the minimum in (3) is $\sqrt{\delta\,\mathbb{E}[X]}\log(1/p)$.

Theorem 1.3. *Let $k \geqslant 3$ be an integer and let $X = X_{N,p}^{k\text{-}AP}$ denote the number of k-term arithmetic progressions in $[\![N]\!]_p$. Then, for every fixed positive constant δ and all $p = p(N)$ satisfying $N^{-1}\log N \ll p^{k/2} \ll 1$,*

$$\lim_{N\to\infty} \frac{-\log \mathbb{P}\big(X \geqslant (1+\delta)\,\mathbb{E}[X]\big)}{Np^{k/2}\log(1/p)} = \sqrt{\delta}.$$

Observe that Theorem 1.3 shows that the upper tail probability is well-approximated by the probability of appearance of an interval (or, more generally, an arithmetic progression) of length $\sqrt{\delta N^2 p^k}$ in $[\![N]\!]_p$. Since each such interval contains approximately $\delta\,\mathbb{E}[X]$ arithmetic progressions of length k, it is not hard to see that conditioning $[\![N]\!]_p$ on the appearance of such a set will cause the upper tail event to occur with sizable probability. Conversely, our methods may be used to prove that the upper tail event is dominated by the appearance of some set of size $(1+o(1))\sqrt{\delta N^2 p^k}$ that contains nearly $\delta\,\mathbb{E}[X]$ arithmetic progressions of length k. It seems natural to guess that each such set is, in some sense, close to an arithmetic progression. However, this is not the case, as was shown by Green–Sisask [33]. We currently do not know a structural characterisation of the sets described above, which prevents us from proving a qualitative description of the upper tail event. For further discussion, we refer to Section 10.

The second result treats the complementary range $N^{-1} \ll p^{k/2} \ll N^{-1}\log N$, where the upper tail has Poisson behaviour.

Theorem 1.4. *Let $k \geqslant 3$ be an integer and let $X = X_{N,p}^{k\text{-}AP}$ denote the number of k-term arithmetic progressions in $[\![N]\!]_p$. Then, for every fixed positive constant δ and all $p = p(N)$ satisfying $N^{-1} \ll p^{k/2} \ll N^{-1}\log N$,*

$$\lim_{N\to\infty} \frac{-\log \mathbb{P}\big(X \geqslant (1+\delta)\,\mathbb{E}[X]\big)}{\mathbb{E}[X]} = (1+\delta)\log(1+\delta) - \delta.$$

1.2. **Subgraph counts in random graphs.** Let $X = X_{N,p}^{H}$ be the number of copies of a fixed graph H in $G_{n,p}$. Note that $\mathbb{E}[X] = \Theta(n^{v_H}p^{e_H})$. Since controlling the distribution of X for completely general graphs involves many technical difficulties (see for example [11, 56]), we will restrict our attention to connected, Δ-regular graphs H. If the expected value of X is bounded, then X converges to a Poisson random variable, as was shown independently by Bollobás [10] and by Karoński–Ruciński [43]. In view of this, for the remainder of this section, we shall assume that $\mathbb{E}[X] \to \infty$, i.e., that $p^{\Delta/2} \gg n^{-1}$. As mentioned before, we will also assume that $p \to 0$; the case where $p \in (0,1)$ is fixed, which is fundamentally different, was addressed in [20, 50].

The problem of controlling the upper tail of X has a long history. A sequence of papers [40, 47, 60], culminating in the work of Janson, Oleszkiewicz, and Ruciński [38], resulted in upper and lower bounds on the logarithmic upper tail probability that differed by a multiplicative

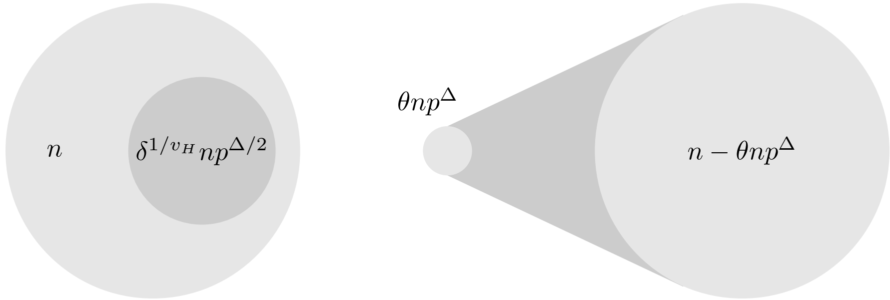

FIGURE 1. The two constructions giving lower bounds for the upper tail probability of $X = X_{n,p}^H$.

factor of $\log(1/p)$. In the case where H is a clique, Chatterjee [15] and DeMarco–Kahn [23] independently added the missing logarithmic factor to the upper bound, thus establishing the order of magnitude of the logarithmic upper tail probability. The theory of nonlinear large deviations (discussed above) provides a variational description of the dependence of the upper tail probability $\mathbb{P}(X \geqslant (1+\delta)\,\mathbb{E}[X])$ on δ for a certain range of $p \to 0$, as established in [2, 17, 22, 27]; the strongest of these results require $p^{\Delta/2} \gg n^{-\Delta/(6\Delta-4)}$ for general graphs [22] and $p^{\Delta/2} \gg n^{-1/2}$ for the case where H is a cycle [2] (disregarding polylogarithmic factors). The associated variational problem was solved by Lubetzky–Zhao [51] when H is a clique and by Bhattacharya, Ganguly, Lubetzky, and Zhao [8] for general H. For a more detailed overview of these techniques, we refer the reader to the book of Chatterjee [16].

The solution to the variational problem is expressed in terms of the independence polynomial of a graph. For any H, define $P_H(x) = \sum_k i_k(H)x^k$, where $i_k(H)$ is the number of independent sets of H of size k, and let $\theta = \theta(\delta)$ be the unique positive solution to $P_H(\theta) = 1 + \delta$.[2] There are two constructions that yield lower bounds for the upper tail probability (see Figure 1). In both cases, one plants a 'small' subgraph whose presence ensures that $G_{n,p}$ contains $(1+\delta)\,\mathbb{E}[X]$ copies of H with good probability. The first of these subgraphs is a clique on $\delta^{1/v_H} np^{\Delta/2}$ vertices (as in the left side of the figure), which contains the extra $\delta\,\mathbb{E}[X]$ copies of H required by the upper tail event (up to lower-order corrections). The second subgraph is a complete bipartite graph with parts of size θnp^{Δ} and $n - \theta np^{\Delta}$, respectively (as in the right side of the figure); since we are implicitly assuming that θnp^{Δ} is an integer, rounding errors play a significant role here unless $np^{\Delta} \gg 1$. A short calculation shows that the expected number of copies of H which intersect this graph is approximately $\delta\,\mathbb{E}[X]$ and thus the actual number of such copies is almost $\delta\,\mathbb{E}[X]$ with good probability. In both cases, the complement of the planted subgraph typically contains approximately $\mathbb{E}[X]$ copies of H. Formalising this argument, one obtains the two lower bounds

$$\mathbb{P}\big(X \geqslant (1+\delta)\,\mathbb{E}[X]\big) \geqslant p^{(\delta^{2/v_H}+o(1))n^2p^{\Delta}} \quad \text{and} \quad \mathbb{P}\big(X \geqslant (1+\delta)\,\mathbb{E}[X]\big) \geqslant p^{(\theta+o(1))n^2p^{\Delta}},$$

which correspond to planting the clique and the complete bipartite graph, respectively. (Recall that the latter bound is valid only when $np^{\Delta} \gg 1$.)

Our main result is that, when H is not bipartite, one of the above bounds is tight in nearly the whole range of densities. When H is bipartite, we prove tight bounds on the logarithmic upper tail probability only when $p^{\Delta/2} \geqslant n^{-1/2-o(1)}$. While we believe that our general approach can be used to solve the upper tail problem for nearly all densities for which localisation occurs, as in the non-bipartite case, the arguments presented in this paper are not sufficiently strong to achieve this goal. The reason is that the aforementioned entropic stability property ceases to hold in the bipartite case as soon as $p^{\Delta/2} \leqslant n^{-1/2-\Theta(1)}$, see Section 10.

[2]We note that $i_0(H) = 1$ for every graph H, so that, for example, $P_{K_r}(x) = 1 + rx$.

Theorem 1.5. *Let $\Delta \geqslant 2$ be an integer, let H be a connected, nonbipartite, Δ-regular graph, and let $X = X^H_{N,p}$ denote the number of copies of H in $G_{n,p}$. Then for every fixed positive constant δ and all $p = p(n)$ satisfying $n^{-1}(\log n)^{\Delta v_H^2} \ll p^{\Delta/2} \ll 1$,*

$$\lim_{n\to\infty} \frac{-\log \mathbb{P}\big(X \geqslant (1+\delta)\,\mathbb{E}[X]\big)}{n^2 p^\Delta \log(1/p)} = \begin{cases} \delta^{2/v_H}/2 & \text{if } np^\Delta \to 0, \\ \min\{\delta^{2/v_H}/2, \theta\} & \text{if } np^\Delta \to \infty, \end{cases}$$

where θ is the unique positive solution to $P_H(\theta) = 1+\delta$. Additionally, if $p^{\Delta/2} \geqslant n^{-1/2-o(1)}$, then the assumption that H is nonbipartite is not necessary.

We note that the theorem leaves open the case where $np^\Delta \to c \in (0,\infty)$. In this regime, the explicit dependence of the upper tail probability on δ involves various integrality conditions and is therefore quite complicated. In the next subsection, we explicitly treat this regime when H is a clique.

Our next result concerns the Poisson regime of the upper tail.

Theorem 1.6. *Let $\Delta \geqslant 2$ be an integer, let H be a connected, Δ-regular graph, and let $X = X^H_{n,p}$ denote the number of copies of H in $G_{n,p}$. Then, for every fixed positive constant δ and all $p = p(n)$ satisfying $n^{-1} \ll p^{\Delta/2} \ll n^{-1}(\log n)^{\frac{1}{v_H-2}}$,*

$$\lim_{n\to\infty} \frac{-\log \mathbb{P}\big(X \geqslant (1+\delta)\,\mathbb{E}[X]\big)}{\mathbb{E}[X]} = (1+\delta)\log(1+\delta) - \delta.$$

We remark that it has been known for some time that the logarithmic upper tail probability in this regime is of order $\mathbb{E}[X]$; this was proved by DeMarco and Kahn [23] in the case when H is a clique and by Šileikis [55] in the more general case of strictly balanced H (which includes the connected, Δ-regular graphs).

Finally, we point out that, regrettably, the powers of the logarithms in the assumptions of Theorems 1.5 and 1.6 do not match. Since the exponent $1/(v_H-2)$ in the assumptions of Theorem 1.6 is best possible (see the discussion in Section 8), we conjecture that the conclusion of Theorem 1.5 remains true under the weaker assumption that $n^{-1}(\log n)^{\frac{1}{v_H-2}} \ll p^{\Delta/2} \ll 1$.

1.3. **Clique counts in random graphs.** We now consider the case of $X = X^H_{n,p}$ where H is a clique on $r \geqslant 3$ vertices. Thanks to the simpler structure of these graphs, we are able to prove significantly stronger results in this setting. First, we are able to determine the explicit dependence of the logarithmic upper tail probability on δ even when $np^{r-1} \to c \in (0,\infty)$. Moreover, we resolve the upper tail problem for the optimal range of densities $n^{-1}(\log n)^{\frac{1}{r-2}} \ll p^{\frac{r-1}{2}} \ll 1$, complementing the range covered by Theorem 1.6. Finally, we give a structural description of $G_{n,p}$ conditioned on the upper tail event.

In order to formally state the theorem, it is convenient to define

$$\psi_r(\delta, c, x) = \frac{\big(\delta(1-x)\big)^{2/r}}{2} + \frac{\lfloor x\delta c/r\rfloor + \{x\delta c/r\}^{\frac{1}{r-1}}}{c}, \tag{4}$$

where δ and c are nonnegative reals, $x \in [0,1]$, and $\{a\}$ denotes the fractional part of a, and

$$\varphi_r(\delta, c) = \min_{x\in[0,1]} \psi_r(\delta, c, x). \tag{5}$$

For an intuitive explanation of the combinatorial meaning of these definitions, we refer to the discussion at the beginning of Section 6. An easy convexity argument shows that the minimum in the definition of φ_r is attained when $x \in \{0, r\lfloor \delta c/r\rfloor/(\delta c), 1\}$, see Lemma 6.1. This leads to the explicit formula

$$\varphi_r(\delta, c) = \min\left\{\frac{\delta^{2/r}}{2}, \frac{\lfloor \delta c/r\rfloor + \{\delta c/r\}^{1/(r-1)}}{c}, \frac{\lfloor \delta c/r\rfloor}{c} + \frac{(r\{\delta c/r\}/c)^{2/r}}{2}\right\}.$$

Theorem 1.7. *Let $r \geqslant 3$ be an integer and let $X = X_{n,p}^{K_r}$ denote the number of r-vertex cliques in the random graph $G_{n,p}$. Then, for every fixed positive constant δ and all $p = p(n)$ satisfying $n^{-1}(\log n)^{\frac{1}{r-2}} \ll p^{\frac{r-1}{2}} \ll 1$,*

$$\lim_{n\to\infty} \frac{-\log \mathbb{P}\big(X \geqslant (1+\delta)\,\mathbb{E}[X]\big)}{n^2 p^{r-1}\log(1/p)} = \begin{cases} \delta^{2/r}/2 & \text{if } np^{r-1} \to 0,\\ \varphi_r(\delta, c) & \text{if } np^{r-1} \to c \in (0,\infty),\\ \min\left\{\delta^{2/r}/2, \delta/r\right\} & \text{if } np^{r-1} \to \infty.\end{cases}$$

Our next result describes the typical structure of the random graph $G_{n,p}$ conditioned upon the upper tail event. Define the following three events:

(i) Let $\mathrm{UT}(\delta)$ be the upper tail event $\{X \geqslant (1+\delta)\,\mathbb{E}[X]\}$.
(ii) Let $\mathrm{Clique}_\varepsilon(x)$ be the event that $G_{n,p}$ contains a set $U \subseteq V(G)$ of size at least $(1-\varepsilon)x^{1/r}np^{(r-1)/2}$ such that $G[U]$ has minimum degree at least $(1-\varepsilon)|U|$.
(iii) Let $\mathrm{Hub}_\varepsilon(x)$ be the event that $G_{n,p}$ contains a set $U \subseteq V(G)$ such that at least $\lfloor(1-\varepsilon)|U|\rfloor$ vertices in U have degree at least $(1-\varepsilon)n$ and

$$e(U, V(G)\setminus U) \geqslant (1-\varepsilon)n\big(\lfloor xnp^{r-1}/r\rfloor + \{xnp^{r-1}/r\}^{\frac{1}{r-1}}\big).$$

Observe that $\mathrm{Clique}_\varepsilon(0)$ and $\mathrm{Hub}_\varepsilon(0)$ hold vacuously.

Theorem 1.8. *Let $r \geqslant 3$ be an integer and let δ, ε, and c be fixed positive constants. The following holds for all $p = p(n)$ satisfying $n^{-1}(\log n)^{\frac{1}{r-1}} \ll p^{\frac{r-1}{2}} \ll 1$.*

(i) If $np^{r-1} \to 0$, then

$$\mathbb{P}\big(\,\mathrm{Clique}_\varepsilon(\delta) \mid \mathrm{UT}(\delta)\big) \to 1.$$

(ii) If $np^{r-1} \to c$, then, letting $x^* = r\lfloor \delta c/r\rfloor/(\delta c)$,

$$\mathbb{P}\left(\bigcup_{x\in\{0,x^*,1\}} \Big(\mathrm{Clique}_\varepsilon\,\big(\delta(1-x)\big) \cap \mathrm{Hub}_\varepsilon(\delta x)\Big) \mid \mathrm{UT}(\delta)\right) \to 1,$$

Moreover, if $x \mapsto \psi_r(\delta, c, x)$ has a unique minimiser $x \in \{0, x^*, 1\}$, then

$$\mathbb{P}\Big(\mathrm{Clique}_\varepsilon\,\big(\delta(1-x)\big) \cap \mathrm{Hub}_\varepsilon(\delta x) \mid \mathrm{UT}(\delta)\Big) \to 1.$$

(iii) If $np^{r-1} \to \infty$, then

$$\mathbb{P}\big(\,\mathrm{Clique}_\varepsilon(\delta) \cup \mathrm{Hub}_\varepsilon(\delta) \mid \mathrm{UT}(\delta)\big) \to 1.$$

Moreover,

$$\mathbb{P}\big(\,\mathrm{Clique}_\varepsilon(\delta) \mid \mathrm{UT}(\delta)\big) \to \begin{cases} 1 & \text{if } \delta^{2/r}/2 < \delta/r,\\ 0 & \text{if } \delta^{2/r}/2 > \delta/r.\end{cases}$$

Note that Theorem 1.8 remains agnostic about the exact structure of the conditional model in the case where there are multiple minimisers to $x \mapsto \psi_r(\delta, np^{r-1}, x)$. However, it is not too difficult to show that for every r, the set of $(\delta, c) \in (0,\infty)^2$ for which $x \mapsto \psi_r(\delta, c, x)$ has multiple minimisers has Lebesgue measure zero. Figure 2 gives a graphical representation of the assertion of Theorem 1.8 in the case where $r = 3$ and $np^2 \to c$. As the figure illustrates, the conditional model undergoes infinitely many phase transitions if $\delta^{2/3}/2 > \delta/3$ (that is, $\delta < 3.375$) and no phase transition at all if $\delta^{2/3}/2 < \delta/3$.

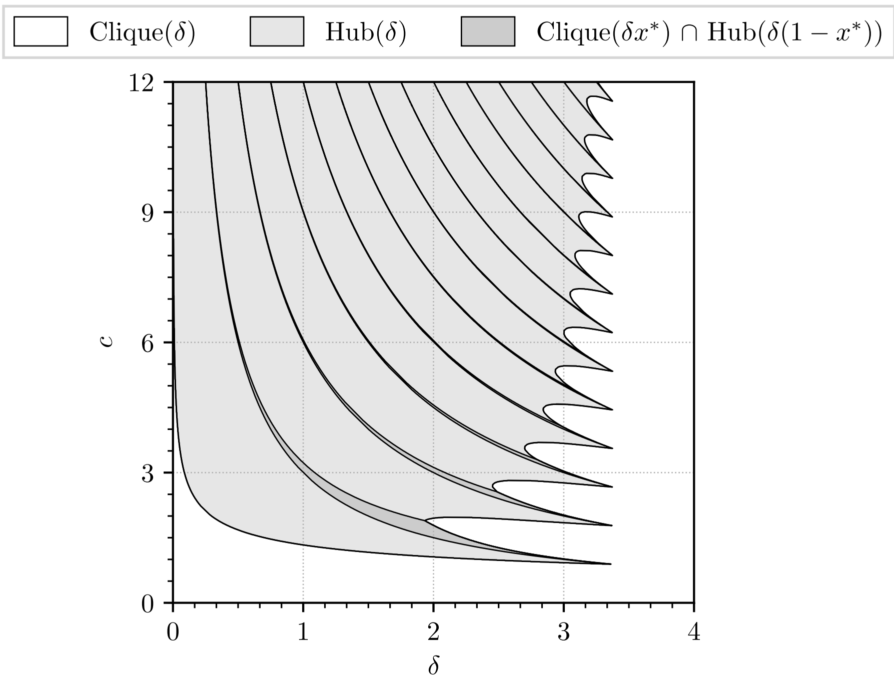

FIGURE 2. Asymptotic structure of $G_{n,p}$ conditioned upon the upper tail event $\mathrm{UT}(\delta) = \{X_{n,p}^{K_3} \geqslant (1+\delta)\,\mathbb{E}[X_{n,p}^{K_3}]\}$ when $np^2 \to c$ as $n \to \infty$. In this conditional model, it is highly probable that we observe either $\mathrm{Clique}_\varepsilon(\delta)$, $\mathrm{Hub}_\varepsilon(\delta)$, or $\mathrm{Clique}_\varepsilon(\delta x^*) \cap \mathrm{Hub}_\varepsilon((1-\delta)x^*)$ with $x^* = 3\lfloor \delta c/3 \rfloor/(\delta c) \in (0,1)$, depending on the values of δ and c (all regions are open).

1.4. **Organisation of the paper.** In Section 2, we present a short and self-contained solution to the upper tail problem for triangle counts in $G_{n,p}$. This section is somewhat redundant, since its content is just a special case of the more general Proposition 6.4. We include it in order to demonstrate our method in a simple setting that conveniently avoids many technical complications that arise in the general case.

Section 3 introduces a concentration inequality that gives a general condition under which the logarithmic upper tail probability can be approximated by $\Phi_X(\delta)$, the solution to the optimisation problem (2).

In Section 4, we use the inequality developed in Section 3 to determine the asymptotics of the logarithmic upper tail probability of $X_{N,p}^{k\text{-AP}}$ in the complete range of densities where localisation occurs. After collecting some graph-theoretic tools in Section 5, we study the localised regime of the upper tails of $X_{n,p}^{K_r}$ and $X_{n,p}^{H}$ for connected, Δ-regular graphs H in Sections 6 and 7, respectively. We note that the three Sections 4, 6, and 7 are logically independent and may be read in any order; however, both Sections 6 and 7 rely on the tools of Section 5.

In Section 8, we prove various results related to the Poisson regime; in particular, we give the proofs of Theorems 1.4 and 1.6. The arguments we use there do not rely on the methods developed in Section 3, but rather on explicit calculations of high factorial moments. Section 9 contains a brief discussion on extending the result from Section 3 to the more general case of nonnegative random variables on the hypercube. Finally, in Section 10 we make some concluding remarks and discuss open problems.

1.5. **Notation.** Before ending the introduction, we collect some notation which will be used throughout the paper. For any graph G, let $V(G)$ and $E(G)$ denote the vertex and edge sets of

G, respectively, and set $v_G = |V(G)|$ and $e_G = |E(G)|$. For two graphs J and G, we let $N(J,G)$ be the number of copies of J in G, and $\mathrm{Emb}(J,G)$ be the set of embeddings of J into G—i.e., injective maps from $V(J)$ to $V(G)$ that map edges of J to edges of G. For an edge $uv \in E(G)$, we also let $N(J,G;uv)$ and $\mathrm{Emb}(J,G;uv)$ be the number of copies of J that contain the edge uv, and the set of embeddings that map an edge of J to uv, respectively. Finally, for a subset I of $[\![N]\!]$, we let $\mathbb{E}_I[X] = \mathbb{E}[X \mid Y_I = 1]$. If subsets $I \subseteq [\![N]\!]$ can be identified with subgraphs $G \subseteq K_n$, as in Example 1.2, we will write $\mathbb{E}_G[X]$ instead of $\mathbb{E}_I[X]$.

2. Triangles in random graphs

Assume that $n^{-1}\log n \ll p \ll 1$ and let X denote the number of triangles in $G_{n,p}$. Using the shorthand notation $\mathbb{E}_G[X] = \mathbb{E}[X \mid G \subseteq G_{n,p}]$, we define, for each positive δ,

$$\Phi_X(\delta) = \min\big\{e_G \log(1/p) : G \subseteq K_n \text{ and } \mathbb{E}_G[X] \geqslant (1+\delta)\,\mathbb{E}[X]\big\}. \tag{6}$$

Note that this agrees with the definition (2). The goal of this section is to prove that, for every fixed positive ε and all large enough n,

$$(1-\varepsilon)\Phi_X(\delta-\varepsilon) \leqslant -\log \mathbb{P}\big(X \geqslant (1+\delta)\,\mathbb{E}[X]\big) \leqslant (1+\varepsilon)\Phi_X(\delta+\varepsilon). \tag{7}$$

At this point, we do not address the issue of evaluating $\Phi_X(\delta)$. For the sake of completeness, let us mention that a special case of a more general result of Lubetzky–Zhao [51] is that, when $n^{-1} \ll p \ll 1$,

$$\lim_{n\to\infty} \frac{\Phi_X(\delta)}{n^2p^2\log(1/p)} = \begin{cases} \delta^{2/3}/2 & \text{if } np^2 \to 0 \\ \min\big\{\delta/3, \delta^{2/3}/2\big\} & \text{if } np^2 \to \infty. \end{cases}$$

In Section 6, we shall fill in the gap at $p = \Theta(n^{-1/2})$ to obtain an asymptotic formula for $\Phi_X(\delta)$ in the full range of interest.

We now give a proof of (7), where we may assume without loss of generality that $\varepsilon \leqslant \delta/10$. All statements that we make in this section should be understood to hold only for sufficiently large n. We start with some easy observations. First, for every graph G with $O(1)$ edges, we have $\mathbb{E}_G[X] - \mathbb{E}[X] = O(1+np^2) \ll (\delta-2\varepsilon)\,\mathbb{E}[X]$, and so $\Phi_X(\delta-2\varepsilon) \gg \log(1/p) \gg 1$. Second, the condition in (6) is satisfied when G is a clique on $(1+\delta)^{1/3}np$ vertices, and therefore $\Phi_X(\delta-\varepsilon) \leqslant (1+\delta)^{2/3}n^2p^2\log(1/p)/2$.

The easier of the two inequalities in (7) is the upper bound. To prove it, let G be a graph attaining the minimum in the definition of $\Phi_X(\delta+\varepsilon)$ and let $\mathbb{P}_G(\cdot) = \mathbb{P}(\cdot \mid G \subseteq G_{n,p})$. Since X never exceeds $\binom{n}{3}$, then

$$(1+\delta+\varepsilon)\,\mathbb{E}[X] \leqslant \mathbb{E}_G[X] \leqslant \binom{n}{3}\mathbb{P}_G\big(X \geqslant (1+\delta)\,\mathbb{E}[X]\big) + (1+\delta)\,\mathbb{E}[X],$$

and so

$$\mathbb{P}_G\big(X \geqslant (1+\delta)\,\mathbb{E}[X]\big) \geqslant \varepsilon\,\mathbb{E}[X]/\binom{n}{3} = \varepsilon p^3.$$

Hence,

$$\begin{aligned} -\log \mathbb{P}\big(X \geqslant (1+\delta)\,\mathbb{E}[X]\big) &\leqslant -\log\Big(\mathbb{P}(G \subseteq G_{n,p}) \cdot \mathbb{P}_G\big(X \geqslant (1+\delta)\,\mathbb{E}[X]\big)\Big) \\ &\leqslant e_G\log(1/p) + \log(1/\varepsilon p^3). \end{aligned}$$

Since $e_G\log(1/p) = \Phi_X(\delta+\varepsilon) \geqslant \Phi_X(\delta-2\varepsilon) \gg \log(1/p)$, this establishes the lower bound in (7).

We now turn to proving the lower bound. Let $C = C(\varepsilon,\delta)$ denote a sufficiently large positive constant. We call a graph $G \subseteq K_n$ a *seed* if

(S1) $\mathbb{E}_G[X] \geqslant (1+\delta-\varepsilon)\,\mathbb{E}[X]$ and
(S2) $e_G \leqslant Cn^2p^2\log(1/p)$.

Similarly, we call a graph $G^* \subseteq K_n$ a *core* if

(C1) $\mathbb{E}_{G^*}[X] \geqslant (1+\delta-2\varepsilon)\,\mathbb{E}[X]$,
(C2) $e_{G^*} \leqslant Cn^2p^2\log(1/p)$, and
(C3) $\min_{e\in E(G^*)}\big(\mathbb{E}_{G^*}[X]-\mathbb{E}_{G^*\setminus e}[X]\big) \geqslant \varepsilon\,\mathbb{E}[X]/\big(Cn^2p^2\log(1/p)\big)$.

We make three claims.

Claim 2.1. *$\mathbb{P}\big(X \geqslant (1+\delta)\,\mathbb{E}[X]\big) \leqslant (1+\varepsilon)\mathbb{P}\big(G_{n,p}$ contains a seed$\big)$.*

Claim 2.2. *Every seed contains a core.*

Claim 2.3. *For every m, there are at most $(1/p)^{\varepsilon m}$ cores with exactly m edges.*

Let us first show how these three claims imply the lower bound in (7):

$$\begin{aligned}\mathbb{P}\big(X \geqslant (1+\delta)\,\mathbb{E}[X]\big) &\overset{\text{Claim 2.1}}{\leqslant} (1+\varepsilon)\mathbb{P}\big(G_{n,p}\text{ contains a seed}\big)\\ &\overset{\text{Claim 2.2}}{\leqslant} (1+\varepsilon)\mathbb{P}\big(G_{n,p}\text{ contains a core}\big)\\ &\leqslant (1+\varepsilon)\sum_{m=0}^{\infty} p^m\cdot|\{G^*\subseteq K_n : G^*\text{ is a core with } m \text{ edges}\}|\\ &\overset{\text{Claim 2.3}}{\leqslant} (1+\varepsilon)\sum_{m=m_{\min}}^{\infty} p^{(1-\varepsilon)m},\end{aligned}$$

where $m_{\min}$ is the minimal number of edges in a core. Since (C1) implies that $\Phi_X(\delta-2\varepsilon) \leqslant m_{\min}\cdot\log(1/p)$, the assumption $p \ll 1$ yields

$$\mathbb{P}\big(X \geqslant (1+\delta)\,\mathbb{E}[X]\big) \leqslant (1+2\varepsilon)\exp\big(-(1-\varepsilon)\Phi_X(\delta-2\varepsilon)\big).$$

Finally, as $\Phi_X(\delta-2\varepsilon) \gg 1$, we obtain

$$-\log\mathbb{P}\big(X \geqslant (1+\delta)\,\mathbb{E}[X]\big) \geqslant (1-2\varepsilon)\Phi_X(\delta-2\varepsilon),$$

thus proving (7) with 2ε instead of ε. It remains to prove Claims 2.1, 2.2, and 2.3.

Proof of Claim 2.1. We refine a classical argument due to Janson, Oleszkiewicz, and Ruciński [38]. Let Z be the indicator random variable of the event that $G_{n,p}$ does not contain a seed and let $\ell = \lceil (C/3)n^2p^2\log(1/p)\rceil$. Since $XZ \geqslant 0$ and $Z^\ell = Z$, Markov's inequality gives

$$\mathbb{P}\big(X \geqslant (1+\delta)\,\mathbb{E}[X]\text{ and } G_{n,p}\text{ contains no seed}\big) = \mathbb{P}\big(XZ \geqslant (1+\delta)\,\mathbb{E}[X]\big) \leqslant \frac{\mathbb{E}[X^\ell Z]}{(1+\delta)^\ell\,\mathbb{E}[X]^\ell}. \quad (8)$$

We write $X = \sum_T Y_T$, where the sum ranges over all triangles T in K_n and Y_T is the indicator random variable of the event that T is contained in $G_{n,p}$. For every subgraph $G \subseteq K_n$, let Z_G be the indicator random variable of the event that $G \cap G_{n,p}$ does not contain a seed. Observe that $Z_{G'} \leqslant Z_G$ whenever $G \subseteq G'$. In particular, since $Z = Z_{K_n}$, then, for every $k \in [\![\ell]\!]$,

$$\begin{aligned}\mathbb{E}[X^kZ] &= \sum_{T_1,\dots,T_k}\mathbb{E}[Y_{T_1}\cdots Y_{T_k}\cdot Z]\\ &\leqslant \sum_{T_1,\dots,T_k}\mathbb{E}[Y_{T_1}\cdots Y_{T_k}\cdot Z_{T_1\cup\dots\cup T_k}]\\ &\leqslant \sum_{T_1,\dots,T_{k-1}}\mathbb{E}[Y_{T_1}\cdots Y_{T_{k-1}}\cdot Z_{T_1\cup\dots\cup T_{k-1}}]\cdot\sum_{T_k}\mathbb{E}[Y_{T_k}\mid Y_{T_1}\dots Y_{T_{k-1}}\cdot Z_{T_1\cup\dots\cup T_{k-1}}=1],\end{aligned}$$

where we can let the first sum in the last line range only over sequences $T_1,\dots,T_{k-1}$ for which the event $\big\{Y_{T_1}\cdots Y_{T_{k-1}}\cdot Z_{T_1\cup\dots\cup T_{k-1}}=1\big\}$ has positive probability. This is equivalent to saying that the graph $T_1\cup\dots\cup T_{k-1}$ does not contain a seed and thus $Y_{T_1}\cdots Y_{T_{k-1}}\cdot Z_{T_1\cup\dots\cup T_{k-1}} = Y_{T_1}\cdots Y_{T_{k-1}}$. Moreover, since $e_{T_1\cup\dots\cup T_{k-1}} \leqslant 3(k-1) \leqslant 3(\ell-1) \leqslant Cn^2p^2\log(1/p)$, then

$$\sum_{T_k}\mathbb{E}[Y_{T_k}\mid Y_{T_1}\dots Y_{T_{k-1}}=1] = \mathbb{E}_{T_1\cup\dots\cup T_{k-1}}[X] < (1+\delta-\varepsilon)\,\mathbb{E}[X],$$

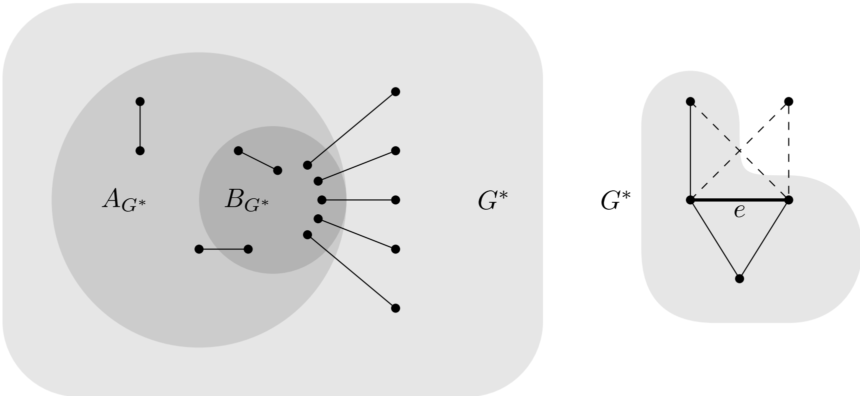

FIGURE 3. Left: The sets A_{G^*} and B_{G^*} of high-degree vertices capture the edges of the core. Right: Three different types of triangles containing an edge e in the core G^*.

as otherwise $T_1 \cup \cdots \cup T_{k-1}$ would be a seed, see (S1) and (S2). Therefore,

$$\sum_{T_1,\ldots,T_k} \mathbb{E}[Y_{T_1} \cdots Y_{T_k} \cdot Z_{T_1\cup\cdots\cup T_k}] < (1+\delta-\varepsilon)\,\mathbb{E}[X] \cdot \sum_{T_1,\ldots,T_{k-1}} \mathbb{E}[Y_{T_1} \cdots Y_{T_{k-1}} \cdot Z_{T_1\cup\cdots\cup T_{k-1}}]$$

and it follows by induction that $\mathbb{E}[X^\ell Z] < (1+\delta-\varepsilon)^\ell\,\mathbb{E}[X]^\ell$. Substituting this inequality into (8) gives

$$\mathbb{P}\big(X \geqslant (1+\delta)\,\mathbb{E}[X] \text{ and } G_{n,p} \text{ contains no seed}\big) \leqslant \left(\frac{1+\delta-\varepsilon}{1+\delta}\right)^\ell.$$

Since the probability that $G_{n,p}$ contains a seed is at least $e^{-\Phi_X(\delta-\varepsilon)}$, the probability that $G_{n,p}$ contains a given seed of smallest size, the bounds $1 \ll \Phi_X(\delta-\varepsilon) \leqslant (1+\delta)^{2/3}n^2p^2\log(1/p)/2$ imply

$$\left(\frac{1+\delta-\varepsilon}{1+\delta}\right)^\ell \leqslant \left(\frac{1+\delta-\varepsilon}{1+\delta}\right)^{(C/3)n^2p^2\log(1/p)} \leqslant \varepsilon\mathbb{P}\big(G_{n,p} \text{ contains a seed}\big)$$

whenever the constant C is sufficiently large. This implies the assertion of the claim. □

Proof of Claim 2.2. Let G be a seed. Define a sequence $G = G_0 \supseteq G_1 \supseteq \cdots \supseteq G_s = G^*$ of subgraphs of G by repeatedly setting $G_{i+1} = G_i \setminus e$ for some edge $e \in G_i$ such that $\mathbb{E}_{G_i} - \mathbb{E}_{G_i\setminus e}[X] < \varepsilon\,\mathbb{E}[X]/\big(Cn^2p^2\log(1/p)\big)$, as long as such an edge e exists. By construction, G^* clearly satisfies (C3). Since $e_{G^*} \leqslant e_G \leqslant Cn^2p^2\log(1/p)$, we see that (C2) holds as well. Finally, as $s \leqslant e_G \leqslant Cn^2p^2\log(1/p)$, we have

$$\mathbb{E}_G[X] - \mathbb{E}_{G^*}[X] = \sum_{i=0}^{s-1}\big(\mathbb{E}_{G_i}[X] - \mathbb{E}_{G_{i+1}}[X]\big) < \varepsilon\,\mathbb{E}[X].$$

Since G is a seed, then $\mathbb{E}_G[X] \geqslant (1+\delta-\varepsilon)\,\mathbb{E}[X]$, and we obtain (C1). □

Proof of Claim 2.3. We bound the number of cores with m edges from above. This number is zero whenever $m > Cn^2p^2\log(1/p)$, by (C2). We may thus assume that $m \leqslant Cn^2p^2\log(1/p)$. Given a core G^*, we denote by A_{G^*} the set of vertices of G^* with degree at least $\varepsilon np/\big(30C\log(1/p)\big)$ and by $B_{G^*} \subseteq A_{G^*}$ the set of vertices of G^* with degree at least $\varepsilon n/\big(30C\log(1/p)\big)$. Since G^* has m edges, then

$$|A_{G^*}| \leqslant a := \frac{60Cm\log(1/p)}{\varepsilon np} \quad \text{and} \quad |B_{G^*}| \leqslant b := \frac{60Cm\log(1/p)}{\varepsilon n}.$$

The key observation, which we will now verify, is that every edge of G^* is either contained in A_{G^*} or has an endpoint in B_{G^*}, see Figure 3 for an illustration. To see this, consider some edge

$e \in E(G^*)$. For every nonempty graph $F \subseteq K_3$, let $N(F, G^*; e)$ denote the number of copies of F in G^* that contain e. By considering how the $n-2$ triangles in K_n that contain e intersect G^* (see Figure 3), one can see that

$$\mathbb{E}_{G^*}[X] - \mathbb{E}_{G^* \setminus e}[X] \leqslant \big(N(K_3, G^*; e) + N(K_{1,2}, G^*; e) \cdot p + np^2\big) \cdot (1-p).$$

Using $\mathbb{E}_{G^*}[X] - \mathbb{E}_{G^* \setminus e}[X] \geqslant \varepsilon\, \mathbb{E}[X]/(Cn^2p^2\log(1/p))$ and $\mathbb{E}[X] \geqslant (1-o(1))n^3p^3/6$, we thus get

$$\frac{\varepsilon np}{7C\log(1/p)} \leqslant N(K_3, G^*; e) + N(K_{1,2}, G^*; e) \cdot p + np^2.$$

Since $p \ll 1$ implies that $np^2 \ll np/\log(1/p)$, we find that either

$$N(K_3, G^*; e) \geqslant \frac{\varepsilon np}{15C\log(1/p)} \quad \text{or} \quad N(K_{1,2}, G^*; e) \geqslant \frac{\varepsilon n}{15C\log(1/p)}. \tag{9}$$

Since $N(K_3, G^*; uv) \leqslant \min\{\deg_{G^*} u, \deg_{G^*} v\}$ and $N(K_{1,2}, G^*; e) \leqslant \deg_{G^*} u + \deg_{G^*} v$, the first inequality in (9) implies that e contained in A_{G^*} whereas the second inequality implies that e has an endpoint in B_{G^*}, as claimed.

Our key observation implies that for fixed sets $B \subseteq A \subseteq [\![n]\!]$ with $|A| = a$ and $|B| = b$, there are at most $\binom{a^2/2+bn}{m}$ cores G^* with m edges that satisfy $A_{G^*} \subseteq A$ and $B_{G^*} \subseteq B$. We can thus (generously) upper bound the number of cores with m edges by

$$\binom{n}{a}\binom{n}{b}\binom{a^2/2+bn}{m}.$$

Recalling the inequality $\binom{x}{y} \leqslant (ex/y)^y$, valid for all nonnegative integers x and y, we may conclude that the number of cores with m edges is at most

$$n^{\frac{120Cm\log(1/p)}{\varepsilon np}} \cdot \left(\frac{e(60C)^2m\big(\log(1/p)\big)^2}{2\varepsilon^2n^2p^2} + \frac{e60C\log(1/p)}{\varepsilon}\right)^m.$$

Since $p \gg n^{-1}\log n$, the first factor is at most $e^{o(m\log(1/p))}$. Using $m \leqslant Cn^2p^2\log(1/p)$, the second factor is at most $e^{O(m\log\log(1/p))} = e^{o(m\log(1/p))}$. This shows that the number of cores with m edges is indeed at most $(1/p)^{\varepsilon m}$, as claimed. □

3. The main technical result: 'entropic stability implies localisation'

The goal of this section is to state a general result that allows one, in many cases of interest, to reduce the problem of determining the precise asymptotics of the logarithmic upper tail probability of a polynomial (with nonnegative coefficients) of independent Bernoulli random variables to a counting problem. Since the main technical lemmas also apply to non-product measures on the hypercube, we phrase the basic definitions in this broader context.

We denote by Y a random variable taking values in the discrete N-dimensional cube $\{0,1\}^N$ and by $X = X(Y)$ a real-valued, increasing function of Y with positive expectation. Given a subset $I \subseteq [\![N]\!]$, we write $Y_I = \prod_{i \in I} Y_i$ for the indicator random variable of the event $\big\{Y_i = 1 \text{ for all } i \in I\big\}$. Using the shorthand notation $\mathbb{E}_I[X] = \mathbb{E}[X \mid Y_I = 1]$,[3] we define a function $\Phi_X \colon \mathbb{R} \to \mathbb{R}_{\geqslant 0} \cup \{\infty\}$ by[4]

$$\Phi_X(\delta) = \min\big\{-\log \mathbb{P}(Y_I = 1) : I \subseteq [\![N]\!] \text{ and } \mathbb{E}_I[X] \geqslant (1+\delta)\,\mathbb{E}[X]\big\}. \tag{10}$$

It is easy to see that Φ_X is a nondecreasing function satisfying $\Phi_X(\delta) > 0$ for all $\delta > 0$. We say that a function $X \colon \{0,1\}^N \to \mathbb{R}_{\geqslant 0}$ is a polynomial with nonnegative coefficients and total degree at most d if it admits a representation $X = \sum_{I \subseteq [\![N]\!]} \alpha_I Y_I$, where each coefficient α_I is nonnegative and $\alpha_I = 0$ whenever $|I| > d$.

The following statement is the main technical result of our work.

[3]Strictly speaking, $\mathbb{E}_I[X]$ is well-defined only if $\mathbb{P}(Y_I = 1) > 0$. However, the value of $\mathbb{E}_I[X]$ for sets I with $\mathbb{P}(Y_I = 1) = 0$ does not affect any of our statements.

[4]We use the standard convention that $\min \varnothing = \infty$.

Theorem 3.1. *For every positive integer d and all positive real numbers ε and δ with $\varepsilon < 1/2$, there is a positive $K = K(d, \varepsilon, \delta)$ such that the following holds. Let Y be a sequence of N independent* $\mathrm{Ber}(p)$ *random variables for some $p \in (0, 1-\varepsilon]$ and let $X = X(Y)$ be a nonzero polynomial with nonnegative coefficients and total degree at most d such that $\Phi_X(\delta - \varepsilon) \geqslant K \log(1/p)$. Denote by $\mathcal{I}^*$ the collection of all subsets $I \subseteq [\![N]\!]$ satisfying*

(C1) $\mathbb{E}_I[X] \geqslant (1 + \delta - \varepsilon)\,\mathbb{E}[X]$,
(C2) $|I| \leqslant K \cdot \Phi_X(\delta + \varepsilon)$, *and*
(C3) $\min_{i \in I} \big(\mathbb{E}_I[X] - \mathbb{E}_{I \setminus \{i\}}[X]\big) \geqslant \mathbb{E}[X] / \big(K \cdot \Phi_X(\delta + \varepsilon)\big)$,

and assume that for every $m \in \mathbb{N}$, there are at most $(1/p)^{\varepsilon m/2}$ sets of size m in $\mathcal{I}^*$. Then

$$(1 - \varepsilon)\Phi_X(\delta - \varepsilon) \leqslant -\log \mathbb{P}\big(X \geqslant (1 + \delta)\,\mathbb{E}[X]\big) \leqslant (1 + \varepsilon)\Phi_X(\delta + \varepsilon) \tag{11}$$

and, writing $\mathcal{J}^*$ for the collection of those $I \in \mathcal{I}^*$ with $-\log \mathbb{P}(Y_I = 1) \leqslant (1 + \varepsilon)\Phi_X(\delta + \varepsilon)$,

$$\mathbb{P}\big(X \geqslant (1 + \delta)\,\mathbb{E}[X] \text{ and } Y_I = 0 \text{ for all } I \in \mathcal{J}^*\big) \leqslant \varepsilon \mathbb{P}\big(X \geqslant (1 + \delta)\,\mathbb{E}[X]\big). \tag{12}$$

Remark 3.2. Observe that (11) gives tight bounds on the logarithmic upper tail probability of X, provided that $\Phi_X(\delta) = \big(I(\delta) + o(1)\big) f(N, p)$ for a continuous, positive function I and some function f. Equation (12) states that the upper tail event is (almost) contained in the event that $Y_I = 1$ for some $I \in \mathcal{J}^*$; note that each such I is a near-minimiser of $\Phi_X(\delta - \varepsilon)$. In some cases, it is possible to classify these near-minimisers and thereby obtain a rough structural characterisation of the upper tail event.

Remark 3.3. The assumption $\Phi_X(\delta - \varepsilon) \geqslant K \log(1/p)$ means that conditioning on $Y_I = 1$ for any constant-size subset $I \subseteq [\![N]\!]$ cannot increase the expected value of X by more than $(\delta - \varepsilon)\,\mathbb{E}[X]$; it is very easy to verify this for the applications we have in mind. The more onerous task is verifying the upper bound on the number of *cores* (as we call the elements of $\mathcal{I}^*$) of a given size. In fact, a large part of this paper is dedicated to counting cores. Frequently, there are very few minimisers of $\Phi_X(\delta)$, for every $\delta > 0$. Our assumption on the number of cores quantifies the notion that there are rather few near-minimisers as well. We call this property 'entropic stability'.

Remark 3.4. In the following sections, we will compute the logarithmic upper tail probabilities in various settings by estimating the function Φ_X and verifying that the random variable X in question satisfies the assumptions of Theorem 3.1. However, there are natural contexts (for example, the upper tail problem for Δ-regular, bipartite graphs at sufficiently low densities) in which the entropic stability assumption of the theorem is not satisfied, but localisation is believed to occur. Nevertheless, we expect that the methods used to prove Theorem 3.1 can be successfully deployed in some of these cases.

The upper bound on $-\log \mathbb{P}\big(X \geqslant (1 + \delta)\,\mathbb{E}[X]\big)$ stated in (11) will follow easily from the following simple lemma.

Lemma 3.5. *Let Y be a random variable taking values in $\{0, 1\}^N$ and let $X = X(Y)$ be a real-valued function of Y. Suppose that $\mathbb{E}[X] > 0$ and that $X \leqslant M$ always. Then for all positive ε and δ,*

$$-\log \mathbb{P}\big(X \geqslant (1 + \delta)\,\mathbb{E}[X]\big) \leqslant \Phi_X(\delta + \varepsilon) + \log\left(\frac{M}{\varepsilon\,\mathbb{E}[X]}\right).$$

Proof. Let $t = (1 + \delta)\,\mathbb{E}[X]$. If $\Phi_X(\delta + \varepsilon) = \infty$, then the assertion of the lemma is vacuous. Otherwise, there exists a set $I \subseteq [\![N]\!]$ with $-\log \mathbb{P}(Y_I = 1) = \Phi_X(\delta + \varepsilon)$ and $\mathbb{E}_I[X] \geqslant t + \varepsilon\,\mathbb{E}[X]$. As $\mathbb{E}_I[X] \leqslant M \cdot \mathbb{P}(X \geqslant t \mid Y_I = 1) + t$, it follows that

$$\mathbb{P}(X \geqslant t) \geqslant \mathbb{P}(Y_I = 1) \cdot \mathbb{P}(X \geqslant t \mid Y_I = 1) \geqslant \mathbb{P}(Y_I = 1) \cdot \frac{\varepsilon\,\mathbb{E}[X]}{M}.$$

Taking the negative logarithm of both sides gives the assertion of the lemma. □

The next lemma lies at the heart of the matter. In very broad terms, it states that the upper tail event $\{X \geqslant (1+\delta)\,\mathbb{E}[X]\}$, viewed as a subset of the cube $\{0,1\}^N$, may be covered almost completely by a union of subcubes of small codimension, where, crucially, the average value of X on each of these subcubes is at least $(1+\delta-\varepsilon)\,\mathbb{E}[X]$. The proof uses a variant of the moment argument of Janson, Oleszkiewicz, and Ruciński [38].

Lemma 3.6. *Let Y be a random variable taking values in $\{0,1\}^N$ and let $X = X(Y)$ be a nonzero polynomial with nonnegative coefficients and total degree at most d. Then for every positive integer ℓ and all positive real numbers ε and δ,*

$$\mathbb{P}\big(X \geqslant (1+\delta)\,\mathbb{E}[X] \text{ and } Y_I = 0 \text{ for all } I \in \mathcal{I}\big) \leqslant \left(\frac{1+\delta-\varepsilon}{1+\delta}\right)^{\ell},$$

where $\mathcal{I} = \big\{I \subseteq [\![N]\!] : |I| \leqslant d\ell \text{ and } \mathbb{E}_I[X] \geqslant (1+\delta-\varepsilon)\,\mathbb{E}[X]\big\}$.

Proof. Given $S \subseteq [\![N]\!]$, let Z_S be the indicator random variable of the event that $Y_I = 0$ for all $I \in \mathcal{I}$ with $I \subseteq S$. Note that $I' \subseteq I$ implies $Z_I \leqslant Z_{I'}$ and let $Z = Z_{[\![N]\!]}$. Since $XZ \geqslant 0$ and $Z^\ell = Z$, Markov's inequality gives

$$\mathbb{P}\big(X \geqslant (1+\delta)\,\mathbb{E}[X] \text{ and } Z = 1\big) = \mathbb{P}\big(XZ \geqslant (1+\delta)\,\mathbb{E}[X]\big) \leqslant \frac{\mathbb{E}[X^\ell Z]}{\big((1+\delta)\,\mathbb{E}[X]\big)^{\ell}}. \tag{13}$$

Write $X = \sum_I \alpha_I Y_I$, where the sum ranges over all subsets $I \subseteq [\![N]\!]$, each coefficient α_I is nonnegative, and $\alpha_I = 0$ unless $|I| \leqslant d$. Then for every $k \in [\![\ell]\!]$,

$$\begin{aligned}
\mathbb{E}[X^k Z] &= \sum_{I_1,\ldots,I_k} \alpha_{I_1}\cdots\alpha_{I_k}\,\mathbb{E}[Y_{I_1}\cdots Y_{I_k}\cdot Z] \\
&\leqslant \sum_{I_1,\ldots,I_k} \alpha_{I_1}\cdots\alpha_{I_k}\,\mathbb{E}[Y_{I_1}\cdots Y_{I_k}\cdot Z_{I_1\cup\cdots\cup I_k}] \\
&\leqslant \sum_{I_1,\ldots,I_{k-1}} \alpha_{I_1}\cdots\alpha_{I_{k-1}}\,\mathbb{E}[Y_{I_1}\cdots Y_{I_{k-1}}\cdot Z_{I_1\cup\cdots\cup I_{k-1}}]\cdot\mathbb{E}[X \mid Y_{I_1}\cdots Y_{I_{k-1}}\cdot Z_{I_1\cup\cdots\cup I_{k-1}} = 1],
\end{aligned}$$

where we may let the third sum range only over sequences $I_1,\ldots,I_{k-1}$ for which the event $\big\{Y_{I_1}\cdots Y_{I_{k-1}}\cdot Z_{I_1\cup\cdots\cup I_{k-1}} = 1\big\}$ has a positive probability of occurring. Note that for any such sequence, $Y_{I_1}\cdots Y_{I_{k-1}}\cdot Z_{I_1\cup\cdots\cup I_{k-1}} = Y_{I_1}\cdots Y_{I_{k-1}}$ and $I_1\cup\cdots\cup I_{k-1} \notin \mathcal{I}$. Since $|I_1\cup\cdots\cup I_{k-1}| \leqslant d(k-1) \leqslant d\ell$, then

$$\mathbb{E}[X \mid Y_{I_1}\cdots Y_{I_{k-1}} = 1] = \mathbb{E}_{I_1\cup\cdots\cup I_k}[X] < (1+\delta-\varepsilon)\,\mathbb{E}[X],$$

as otherwise $I_1\cup\cdots\cup I_{k-1}$ would belong to $\mathcal{I}$. It follows that

$$\begin{aligned}
&\sum_{I_1,\ldots,I_k} \alpha_{I_1}\cdots\alpha_{I_k}\,\mathbb{E}[Y_{I_1}\cdots Y_{I_k}\cdot Z_{I_1\cup\cdots\cup I_k}] \\
&\qquad < (1+\delta-\varepsilon)\,\mathbb{E}[X]\cdot\sum_{I_1,\ldots,I_{k-1}} \alpha_{I_1}\cdots\alpha_{I_{k-1}}\,\mathbb{E}[Y_{I_1}\cdots Y_{I_{k-1}}\cdot Z_{I_1\cup\cdots\cup I_{k-1}}].
\end{aligned}$$

By induction, we see that $\mathbb{E}[X^\ell Z] < \big((1+\delta-\varepsilon)\,\mathbb{E}[X]\big)^{\ell}$. Substituting this inequality into (13) completes the proof. □

The following easy lemma will be used to relate the family $\mathcal{I}$ from the statement of Lemma 3.6 to the family $\mathcal{I}^*$ of cores.

Lemma 3.7. *Let Y be a random variable taking values in $\{0,1\}^N$ and let $X = X(Y)$ be a real-valued function of Y. Then for every $I \subseteq [\![N]\!]$ and every nonnegative real number s, there exists some $I^* \subseteq I$ such that*

(i) $\mathbb{E}_{I^*}[X] \geqslant \mathbb{E}_I[X] - s$ *and*
(ii) $\min_{i\in I^*}\big(\mathbb{E}_{I^*}[X] - \mathbb{E}_{I^*\setminus\{i\}}[X]\big) \geqslant s/|I|$.

Proof. Define a sequence $I = I_0 \supseteq I_1 \supseteq \dots \supseteq I_r = I^*$ by repeatedly setting $I_{k+1} = I_k \setminus \{i\}$ for some $i \in I_k$ such that $\mathbb{E}_I[X] - \mathbb{E}_{I\setminus\{i\}}[X] < s/|I|$, as long as such an i exists. By construction, the set I^* satisfies (ii). Finally, since $r \leqslant |I|$, we have

$$\mathbb{E}_I[X] - \mathbb{E}_{I^*}[X] = \sum_{k=0}^{r-1} \big(\mathbb{E}_{I_k}[X] - \mathbb{E}_{I_{k+1}}[X]\big) \leqslant \max\{0, s\} \leqslant s,$$

which is (i). □

Proof of Theorem 3.1. Let $t = (1+\delta)\,\mathbb{E}[X]$. We first prove the upper bound in (11). Let $\mathbf{1}$ denote the N-dimensional all-ones vector. Since X is an increasing function of Y, then $X \leqslant X(\mathbf{1})$ always. In particular, Lemma 3.5 implies that

$$-\log \mathbb{P}(X \geqslant t) \leqslant \Phi_X(\delta+\varepsilon) + \log\left(\frac{X(\mathbf{1})}{\varepsilon\,\mathbb{E}[X]}\right).$$

As X has total degree at most d and nonnegative coefficients, we have $\mathbb{E}[X] \geqslant X(\mathbf{1}) \cdot p^d$ and thus

$$-\log \mathbb{P}(X \geqslant t) \leqslant \Phi_X(\delta+\varepsilon) + \log\big(1/(\varepsilon p^d)\big) \leqslant (1+\varepsilon/8) \cdot \Phi_X(\delta+\varepsilon), \tag{14}$$

where the second inequality holds provided that K is sufficiently large, as we have assumed that $\Phi_X(\delta+\varepsilon) \geqslant K\log(1/p)$ and $p \leqslant 1-\varepsilon$.

For the rest of the proof let $\ell = \lceil \varepsilon(3d)^{-1}K \cdot \Phi_X(\delta+\varepsilon)\rceil$ and define

$$\mathcal{I} = \big\{I \subseteq [\![N]\!] : |I| \leqslant d\ell \text{ and } \mathbb{E}_I[X] \geqslant (1+\delta-\varepsilon/2)\,\mathbb{E}[X]\big\}.$$

It follows from Lemma 3.6 (invoked with ε replaced by $\varepsilon/2$) that

$$\mathbb{P}(X \geqslant t \text{ and } Y_I = 0 \text{ for all } I \in \mathcal{I}) \leqslant \left(1 - \frac{\varepsilon/2}{1+\delta}\right)^{\ell}.$$

Since we have already shown that $\mathbb{P}(X \geqslant t) \geqslant \exp\big(-(1+\varepsilon)\Phi_X(\delta+\varepsilon)\big)$, see (14), we find that letting K be sufficiently large ensures

$$\mathbb{P}(X \geqslant t \text{ and } Y_I = 0 \text{ for all } I \in \mathcal{I}) \leqslant \big(1 - \varepsilon/(2+2\delta)\big)^{\ell} \leqslant (\varepsilon/2) \cdot \mathbb{P}(X \geqslant t).$$

Note next that every $I \in \mathcal{I}$ satisfies $|I| \leqslant d\ell \leqslant (\varepsilon K/2) \cdot \Phi_X(\delta+\varepsilon)$ and hence, by Lemma 3.7 applied with $s = \varepsilon\,\mathbb{E}[X]/2$, there is a subset $I^* \subseteq I$ satisfying the conditions (C1), (C2), and (C3). It follows that

$$\mathbb{P}(X \geqslant t \text{ and } Y_{I^*} = 0 \text{ for all } I^* \in \mathcal{I}^*) \leqslant (\varepsilon/2) \cdot \mathbb{P}(X \geqslant t). \tag{15}$$

Let $\mathcal{I}^*_m := \{I^* \in \mathcal{I}^* : |I^*| = m\}$ and recall that we assume $|\mathcal{I}^*_m| \leqslant (1/p)^{\varepsilon m/2}$ for all $m \in \mathbb{N}$.

We now prove the upper bound in (11). It follows from (15) that

$$\mathbb{P}(X \geqslant t) \leqslant (1-\varepsilon/2)^{-1} \cdot \mathbb{P}\big(Y_{I^*} = 1 \text{ for some } I^* \in \mathcal{I}^*\big).$$

Moreover, the definitions of $\mathcal{I}^*$ and $\Phi_X(\delta-\varepsilon)$ imply that every core $I^* \in \mathcal{I}^*$ satisfies $|I^*|\log(1/p) = -\log \mathbb{P}(Y_{I^*} = 1) \geqslant \Phi_X(\delta-\varepsilon)$, see (C1). Hence, taking the union bound over all cores and using $|\mathcal{I}^*_m| \leqslant (1/p)^{\varepsilon m/2}$, we find that

$$\begin{aligned}\mathbb{P}(X \geqslant t) &\leqslant (1-\varepsilon/2)^{-1} \sum_{I^* \in \mathcal{I}^*} p^{|I^*|} \leqslant (1-\varepsilon/2)^{-1} \sum_m |\mathcal{I}^*_m| \cdot p^m \\ &\leqslant (1-\varepsilon/2)^{-1} \sum_{m=\frac{\Phi_X(\delta-\varepsilon)}{\log(1/p)}}^{\infty} p^{(1-\varepsilon/2)m} = \frac{e^{-(1-\varepsilon/2)\Phi_X(\delta-\varepsilon)}}{(1-\varepsilon/2)(1-p^{1-\varepsilon/2})}.\end{aligned}$$

Taking logarithms and using $p \leqslant 1-\varepsilon$ and $\Phi_X(\delta-\varepsilon) \geqslant K\log(1/p)$, we see that a large enough choice of K ensures that $-\log \mathbb{P}(X \geqslant t) \geqslant (1-\varepsilon)\Phi_X(\delta-\varepsilon)$, as required.

Finally, let us prove (12). Using (15), we obtain

$$\mathbb{P}(X \geqslant t \text{ and } Y_{I^*} = 0 \text{ for all } I^* \in \mathcal{J}^*) \leqslant (\varepsilon/2) \cdot \mathbb{P}(X \geqslant t) + \mathbb{P}(Y_{I^*} = 1 \text{ for some } I^* \in \mathcal{I}^* \setminus \mathcal{J}^*).$$

Noting that every $I^* \in \mathcal{I}^* \setminus \mathcal{J}^*$ satisfies $|I^*|\log(1/p) = -\mathbb{P}(Y_{I^*} = 1) > (1+\varepsilon)\Phi_X(\delta+\varepsilon)$, we may employ a union bound again to show that

$$\mathbb{P}(Y_{I^*} = 1 \text{ for some } I^* \in \mathcal{I}^* \setminus \mathcal{J}^*) \leqslant \sum_{I^* \in \mathcal{I}^* \setminus \mathcal{J}^*} p^{|I^*|} \leqslant \frac{e^{-(1-\varepsilon/2)(1+\varepsilon)\Phi_X(\delta+\varepsilon)}}{1-p^{1-\varepsilon/2}}.$$

In order to complete the proof, it now suffices to show that

$$\frac{e^{-(1+\varepsilon)(1-\varepsilon/2)\Phi_X(\delta+\varepsilon)}}{1-p^{1-\varepsilon/2}} \leqslant (\varepsilon/2)\cdot \mathbb{P}(X \geqslant t). \tag{16}$$

To see that this inequality holds, note first that $(1+\varepsilon)(1-\varepsilon/2) > 1+\varepsilon/4$ as $\varepsilon < 1/2$ and therefore, by (14),

$$e^{-(1+\varepsilon)(1-\varepsilon/2)\Phi_X(\delta+\varepsilon)} \leqslant \mathbb{P}(X \geqslant t)\cdot e^{-(\varepsilon/8)\Phi_X(\delta+\varepsilon)}.$$

As $p \leqslant 1-\varepsilon$ and $\Phi_X(\delta+\varepsilon) \geqslant K\log(1/p)$, we can choose K so large that $e^{-(\varepsilon/8)\Phi_X(\delta+\varepsilon)}/(1-p^{1-\varepsilon/2}) \leqslant \varepsilon/2$, proving (16). □

4. Arithmetic progressions in random sets of integers

Fix an integer $k \geqslant 3$ and let $X = X_{N,p}^{k\text{-AP}}$ be the number of k-term arithmetic progressions (k-APs) in the random set $[\![N]\!]_p$. The goal of this section is to study the upper tail of X in the regime where Theorem 3.1 is applicable. In particular, we will prove Theorem 1.3, which we restate here for convenience.

Theorem 1.3. *Let $k \geqslant 3$ be an integer and let $X = X_{N,p}^{k\text{-}AP}$ denote the number of k-term arithmetic progressions in $[\![N]\!]_p$. Then, for every fixed positive constant δ and all $p = p(N)$ satisfying $N^{-1}\log N \ll p^{k/2} \ll 1$,*

$$\lim_{N\to\infty} \frac{-\log \mathbb{P}\big(X \geqslant (1+\delta)\,\mathbb{E}[X]\big)}{Np^{k/2}\log(1/p)} = \sqrt{\delta}.$$

To prove the theorem, we will use Theorem 3.1 to relate $-\log \mathbb{P}\big(X \geqslant (1+\delta)\,\mathbb{E}[X]\big)$ to the solution of the optimisation problem

$$\Phi_X(\delta) = \min\big\{|I|\log(1/p) : I \subseteq [\![N]\!] \text{ and } \mathbb{E}_I[X] \geqslant (1+\delta)\,\mathbb{E}[X]\big\}.$$

More precisely, we shall prove the following statement, which is the main result of this section.

Proposition 4.1. *For every integer $k \geqslant 3$ and all positive real numbers ε and δ, there exists a positive constant C such that the following holds. Suppose that $N \in \mathbb{N}$ and $p \in (0,1)$ satisfy $CN^{-1}\log N \leqslant p^{k/2} \leqslant 1/C$. Then $X = X_{N,p}^{k\text{-}AP}$ satisfies*

$$(1-\varepsilon)\Phi_X(\delta-\varepsilon) \leqslant -\log \mathbb{P}\big(X \geqslant (1+\delta)\,\mathbb{E}[X]\big) \leqslant (1+\varepsilon)\Phi_X(\delta+\varepsilon).$$

In order to derive Theorem 1.3 from the proposition, it suffices to estimate the function Φ_X. For this, we use the following extremal result about the largest number of k-APs in a set of integers of a given size, proved in the case $k = 3$ by Green–Sisask [33] and later extended by Bhattacharya, Ganguly, Shao, and Zhao [9] to arbitrary $k \geqslant 3$; the corresponding statement in the case where $k \in \{1,2\}$ is trivial. For a set $I \subseteq \mathbb{Z}$, we denote by $A_k(I)$ the number of k-APs in I. Recall that we only count k-APs with positive common difference.

Theorem 4.2 ([9, 33]). *For every positive integer k and $I \subseteq \mathbb{Z}$, we have $A_k(I) \leqslant A_k\big([\![|I|]\!]\big)$.*

We reproduce the proof here for the sake of completeness.

Proof. We prove the statement by induction on k. The cases $k = 1$ and $k = 2$ are trivial as $A_k(I) = A_k\big([\![|I|]\!]\big)$ for every set I, so we may assume that $k \geqslant 3$. Suppose that $|I| = n$ and let $a_1, \ldots, a_n$ be the elements of I listed in increasing order. We partition the set of k-APs in I into two parts depending on the location of the $(k-1)$st element. More precisely, we let $m = \lceil (k-2)n/(k-1) \rceil$, let

$$\mathcal{A}_1 = \big\{(i_1, \dots, i_k) \in [\![n]\!]^k : (a_{i_1}, \dots, a_{i_k}) \text{ is a } k\text{-AP and } i_{k-1} \leqslant m\big\},$$

and let $\mathcal{A}_2$ comprise the remaining k-APs (that is, ones with $i_{k-1} > m$). Since the removal of the kth term from a progression in $\mathcal{A}_1$ maps it to a $(k-1)$-AP contained the set $\{a_1, \dots, a_m\}$, then $|\mathcal{A}_1| \leqslant A_{k-1}(\{a_1, \dots, a_m\}) \leqslant A_{k-1}\big([\![m]\!]\big)$, by the induction hypothesis. On the other hand, we observe that for every $i > m$, there are at most $n - i$ arithmetic progression of length k such that $i_{k-1} = i$ and thus

$$|\mathcal{A}_2| \leqslant \sum_{i=m+1}^{n} n - i.$$

In order to complete the proof, it is sufficient to verify that our choice of m ensures that

$$A_{k-1}\big([\![m]\!]\big) + \sum_{i=m+1}^{n} n - i = A_k([\![n]\!]).$$

Indeed, m satisfies the following two inequalities:

$$m + \left\lfloor \frac{m-1}{k-2} \right\rfloor \leqslant n \qquad \text{and} \qquad m + 1 - (k-2)(n-m-1) \geqslant 1.$$

The first inequality implies that extending any arithmetic progression $(i_1, \dots, i_{k-1})$ contained in $[\![m]\!]$ by adjoining to it the element $i_k = 2i_{k-1} - i_{k-2}$ yields a k-AP contained in $[\![n]\!]$, whereas the second inequality implies that $\sum_{i=m+1}^{n} n - i$ is precisely the number of k-APs in $[\![n]\!]$ whose $(k-1)$st term exceeds m. □

For future reference, let us note that $A_k\big([\![i]\!]\big) - A_k\big([\![i-1]\!]\big) = \lfloor \frac{i-1}{k-1} \rfloor$ for all positive integers i and $k \geqslant 2$ and, consequently,

$$A_k([\![m]\!]) = \sum_{i=1}^{m} \left\lfloor \frac{i-1}{k-1} \right\rfloor = \frac{m^2}{2(k-1)} - \frac{(k-1)m}{2} \pm k^2. \tag{17}$$

Using Theorem 4.2, it is not difficult to compute the asymptotic value of $\Phi_X(\delta)$ and complete the proof of Theorem 1.3.

Proposition 4.3. *For every integer $k \geqslant 3$ and all positive real numbers ε and δ, there exists a positive constant C such that the following holds. Suppose that $N \in \mathbb{N}$ and $p \in (0,1)$ satisfy $CN^{-1} \leqslant p^{k/2} \leqslant 1/C$. Then $X = X_{N,p}^{k\text{-}AP}$ satisfies*

$$1 - \varepsilon \leqslant \frac{\Phi_X(\delta)}{\sqrt{\delta} \cdot Np^{k/2} \log(1/p)} \leqslant 1 + \varepsilon.$$

Proof. Without loss of generality, we may assume that $\varepsilon \leqslant 1$. Given a subset $I \subseteq [\![N]\!]$, let $a_j(I)$ denote the number of k-APs in $[\![N]\!]$ that intersect I in exactly j elements. Note that

$$\mathbb{E}_I[X] = \sum_{j=0}^{k} a_j(I) p^{k-j} \qquad \text{and that} \qquad \mathbb{E}[X] = A_k\big([\![N]\!]\big)p^k = \sum_{j=0}^{k} a_j(I) p^k. \tag{18}$$

It follows that $\mathbb{E}_I[X] - \mathbb{E}[X] \geqslant (1-p^k)a_k(I) = (1-p^k)A_k(I)$ for every $I \subseteq [\![N]\!]$. In particular, whenever $(1-p^k)A_k([\![m]\!]) \geqslant \delta p^k A_k\big([\![N]\!]\big)$, then $\mathbb{E}_{[\![m]\!]}[X] \geqslant (1+\delta)\,\mathbb{E}[X]$. Therefore,

$$\Phi_X(\delta) \leqslant \min\left\{ m \log(1/p) : A_k\big([\![m]\!]\big) \geqslant \frac{\delta p^k A_k\big([\![N]\!]\big)}{1-p^k} \right\} \leqslant (1+\varepsilon) \cdot \sqrt{\delta} \cdot Np^{k/2} \log(1/p),$$

where the last inequality follows from (17) and our assumption $N^2 p^k \geqslant C^2$ for a sufficiently large constant C. It remains to prove the matching lower bound.

Suppose that I is a smallest subset of $[\![N]\!]$ with $\mathbb{E}_I[X] \geqslant (1+\delta)\,\mathbb{E}[X]$. Then (18) implies

$$\delta A_k\big([\![N]\!]\big)p^k = \delta\,\mathbb{E}[X] \leqslant \mathbb{E}_I[X] - \mathbb{E}[X] \leqslant \sum_{j=1}^{k} a_j(I)p^{k-j}. \tag{19}$$

Since every pair of distinct numbers in $[\![N]\!]$ is contained in at most $\binom{k}{2}$ arithmetic progressions of length k, it follows that $a_1(I) \leqslant |I| \cdot Nk^2$ and $\sum_{j=2}^{k-1} a_j(I) \leqslant |I|^2 \cdot k^2$. Since we already know that $|I| \leqslant 2\sqrt{\delta} \cdot Np^{k/2}$, then (19) gives

$$\delta A_k([\![N]\!])p^k \leqslant 2\sqrt{\delta}k^2N^2p^{3k/2-1} + 4\delta k^2N^2p^{k+1} + A_k(I).$$

We now invoke Theorem 4.2 and (17) to obtain

$$(1-\varepsilon)\cdot\frac{\delta N^2p^k}{2(k-1)} \leqslant \delta A_k([\![N]\!])p^k - 2\sqrt{\delta}k^2N^2p^{3k/2-1} - 4\delta k^2N^2p^{k+1} \leqslant A_k([\![I]\!]) \leqslant \frac{|I|^2}{2(k-1)},$$

where we use the assumptions $p \leqslant C^{-2/k}$ and $N^2p^k \geqslant C^2$ for a large enough C. Thus $\Phi_X(\delta) = |I|\log(1/p) \geqslant (1-\varepsilon)\cdot\sqrt{\delta}\cdot Np^{k/2}\log(1/p)$, as required. $\square$

4.1. **Janson's inequality.** It remains to prove Proposition 4.1. The proof uses the following version of Janson's inequality for hypergeometric random variables. It follows from the (original version of) Janson's inequality for binomial distributions [37, Theorem 1] and the fact that the median of a binomial random variable whose mean is an integer is equal to its mean. Our argument is an adaptation of [5, Lemma 3.1].

Lemma 4.4. *Suppose that $\{B_\alpha\}_{\alpha\in A}$ is a family of subsets of a t-element set Ω. Let $s \in \{0,\dots,t\}$ and let*

$$\mu = \sum_{\alpha\in A}\left(\frac{s}{t}\right)^{|B_\alpha|} \qquad \text{and} \qquad \Delta = \sum_{\alpha\sim\beta}\left(\frac{s}{t}\right)^{|B_\alpha\cup B_\beta|},$$

where the second sum is over all ordered pairs $(\alpha,\beta)\in A^2$ such that $\alpha\neq\beta$ and $B_\alpha\cap B_\beta\neq\varnothing$. Let S be the uniformly chosen random s-element subset of Ω and let Z denote the number of $\alpha\in A$ such that $B_\alpha\subseteq S$. Then for every $\varepsilon\in(0,1]$,

$$\mathbb{P}\big(Z \leqslant (1-\varepsilon)\mu\big) \leqslant 2\exp\left(-\frac{\varepsilon^2}{2}\cdot\frac{\mu^2}{\mu+\Delta}\right).$$

Proof. For every $k\in\{0,\dots,t\}$, let S_k be the uniformly chosen random k-element subset of Ω and let Z_k denote the number of $\alpha\in A$ such that $B_\alpha\subseteq S_k$, so that $Z = Z_s$, and note that there exists a natural coupling under which $Z_k \leqslant Z_{k+1}$ for every k. Let S' be the (s/t)-random subset of Ω, that is the random subset of Ω formed by keeping each of its elements with probability s/t, independently of others, and let Z' denote the number of $\alpha\in A$ such that $B_\alpha\subseteq S'$. Since $\mathbb{E}\left[Z' \mid |S'|\right] = Z_{|S'|}$, then the stochastic ordering of the Z_ks implies that, for any number M, the function $k\mapsto\mathbb{P}(Z'\leqslant M \mid |S'|=k)$ is decreasing. Hence,

$$\begin{aligned}\mathbb{P}\big(Z' \leqslant (1-\varepsilon)\mu\big) &= \sum_{k=0}^{t}\mathbb{P}\big(Z'\leqslant(1-\varepsilon)\mu \mid |S'|=k\big)\cdot\mathbb{P}(|S'|=k)\\ &\geqslant \mathbb{P}\big(Z'\leqslant(1-\varepsilon)\mu \mid |S'|=s\big)\cdot\mathbb{P}(|S'|\leqslant s)\\ &= \mathbb{P}\big(Z\leqslant(1-\varepsilon)\mu\big)\cdot\mathbb{P}(|S'|\leqslant s) \geqslant \mathbb{P}\big(Z\leqslant(1-\varepsilon)\mu\big)/2,\end{aligned}$$

where the last inequality follows from the well-known fact that if np is an integer, then it is the median of the binomial distribution with parameters n and p. We can now invoke the classical version of Janson's inequality and conclude that

$$\mathbb{P}\big(Z\leqslant(1-\varepsilon)\mu\big) \leqslant 2\mathbb{P}\big(Z'\leqslant(1-\varepsilon)\mu\big) \leqslant 2\exp\left(-\frac{\varepsilon^2}{2}\cdot\frac{\mu^2}{\mu+\Delta}\right). \qquad \square$$

4.2. **Proof of Proposition 4.1.** We may assume without loss of generality that ε is sufficiently small, say $\varepsilon < \min\{1/2, \delta/2\}$. Note also that the case $N \leqslant 2$ is trivial; indeed, in that case X is identically zero and thus $\log \mathbb{P}[X \geqslant (1+\delta)\,\mathbb{E}[X]] = 0 = \Phi_X(\delta)$ for every $\delta \in \mathbb{R}$. We may therefore assume that $N \geqslant 3$, which, in turn, implies that $N^2p^k \geqslant C^2$.

Denote by Y_i the indicator random variable of the event that $i \in [\![N]\!]_p$. Then $Y = (Y_1, \ldots, Y_N)$ is a vector of independent $\mathrm{Ber}(p)$ random variables and X is a nonzero polynomial with nonnegative coefficients and total degree at most k in the coordinates of Y. Let $K = K(k, \varepsilon, \delta)$ be the constant given by Theorem 3.1. The proposition follows once we verify that X satisfies the various assumptions of the theorem.

First, our assumption on p implies that $p \leqslant 1 - \varepsilon$ whenever C is large enough. Second, it follows from Proposition 4.3 and the inequality $N^2p^k \geqslant C^2$ that, whenever C is large enough, $\Phi_X(\delta - \varepsilon) \geqslant \Phi_X(\delta/2) \geqslant K \log(1/p)$. Recall that a subset $I \subseteq [\![N]\!]$ is called a core if

(C1) $\mathbb{E}_I[X] \geqslant (1 + \delta - \varepsilon)\,\mathbb{E}[X]$,
(C2) $|I| \leqslant K \cdot \Phi_X(\delta + \varepsilon)$, and
(C3) $\min_{i\in I}\big(\mathbb{E}_I[X] - \mathbb{E}_{I\setminus\{i\}}[X]\big) \geqslant \mathbb{E}[X]/\big(K \cdot \Phi_X(\delta+\varepsilon)\big)$.

The final assumption of Theorem 3.1 is that, for every integer m, there are at most $(1/p)^{\varepsilon m/2}$ cores of size m.

In order to count the cores, we must first unravel the meaning of (C1), (C2), and (C3), and show that each core enjoys a simple combinatorial property. Proposition 4.3 supplies a constant $K' = K'(K, k, \varepsilon, \delta)$ such that, whenever C is sufficiently large,

$$4kK \cdot \Phi_X(\delta+\varepsilon) \leqslant 4kK \cdot (1+\varepsilon)\sqrt{\delta+\varepsilon} \cdot Np^{k/2}\log(1/p) \leqslant K' \cdot Np^{k/2}\log(1/p). \tag{20}$$

Given a set $I \subseteq [\![N]\!]$ and an $i \in [\![N]\!]$, we write $A_k(I; i)$ for the number of k-term arithmetic progressions in $I \cup \{i\}$ that contain the element i. The proof of the following claim is similar to the argument used to prove Proposition 4.3.

Claim 4.5. *For every core I of size m and all $i \in I$,*

$$A_k(I; i) \geqslant \frac{Np^{k/2}}{K' \log(1/p)}.$$

Proof. Given an $i \in I$, let $a_j(I; i)$ denote the number of k-APs in $[\![N]\!]$ that intersect I in exactly j elements, one of which is i. With this notation, $A_k(I; i) = a_k(I; i)$ and we may write $\mathbb{E}_I[X] - \mathbb{E}_{I\setminus\{i\}}[X] = \sum_{j=1}^{k} a_j(I; i) \cdot p^{k-j}(1-p)$. Since every pair of distinct numbers in $[\![N]\!]$ is contained in at most $\binom{k}{2}$ arithmetic progressions of length k, we have $a_1(I; i) \leqslant Nk^2$ and $\sum_{j=2}^{k-1} a_j(I; i) \leqslant mk^2$. In particular, as $m \leqslant K \cdot \Phi_X(\delta+\varepsilon)$ by (C2), we get

$$\mathbb{E}_I[X] - \mathbb{E}_{I\setminus\{i\}}[X] \leqslant k^2Np^{k-1} + k^2K \cdot \Phi_X(\delta+\varepsilon) \cdot p + A_k(I; i).$$

On the other hand, it follows from (C3) that

$$\mathbb{E}_I[X] - \mathbb{E}_{I\setminus\{i\}}[X] \geqslant \frac{\mathbb{E}[X]}{K \cdot \Phi_X(\delta+\varepsilon)}.$$

By (17), we have $\mathbb{E}[X] = A_k([\![N]\!])p^k \geqslant N^2p^k/(2k)$, since $N \geqslant C$ and C is large. Combining the upper and lower bounds on $\mathbb{E}_I[X] - \mathbb{E}_{I\setminus\{i\}}[X]$ and using (20), we obtain

$$A_k(I; i) \geqslant \frac{2Np^{k/2}}{K' \log(1/p)} - k^2Np^{k-1} - kK'Np^{k/2+1}\log(1/p).$$

Since $k \geqslant 3$ and $p \leqslant C^{-2/k}$ for a large enough C, we deduce the assertion of the claim. $\square$

For the remainder of the proof, fix some integer m satisfying $1 \leqslant m \leqslant K \cdot \Phi_X(\delta+\varepsilon)$ and let K'' be a sufficiently large positive constant depending on K' and k (but not on C). For a subset

$I' \subseteq [\![N]\!]$ and an integer $i \in [\![N]\!] \setminus I'$, we shall say that i is *rich* with respect to I' if

$$A_k(I';i) \geqslant \frac{Np^{k/2}}{K'' \log(1/p)} \cdot \left(\frac{|I'|}{m}\right)^{k-1}. \tag{21}$$

Moreover, given a sequence $(i_1, \dots, i_m)$ of m distinct elements of $[\![N]\!]$, we shall say that an index $m' \in [\![m]\!]$ is *rich* if $i_{m'}$ is rich with respect to the set $\{i_1, \dots, i_{m'-1}\}$.

We first observe that for every $I' \subseteq [\![N]\!]$, there are relatively few integers $i \in [\![N]\!] \setminus I'$ that are rich with respect to I'. Indeed, since there are at most $k|I'|^2$ arithmetic progressions P of length k in $[\![N]\!]$ for which $|I' \cap P| = k-1$, then

$$\big|\{i \in [\![N]\!] \setminus I' : i \text{ is rich w.r.t. } I'\}\big| \cdot \frac{Np^{k/2}}{K'' \log(1/p)} \cdot \left(\frac{|I'|}{m}\right)^{k-1} \leqslant \sum_{i \in [\![N]\!] \setminus I'} A_k(I';i) \leqslant k|I'|^2.$$

Consequently, as $m \leqslant K' \cdot Np^{k/2} \log(1/p)$ by (20),

$$\big|\{i \in [\![N]\!] \setminus I' : i \text{ is rich w.r.t. } I'\}\big| \leqslant kK'K'' \cdot m\big(\log(1/p)\big)^2 \cdot \left(\frac{m}{|I'|}\right)^{k-3}. \tag{22}$$

The key property that allows us to control the number of cores I of size m is that, in a large proportion of orderings of the members of I, almost all indices are rich. This property implies that, if one builds an (ordered) core element by element, then, very often, one must choose the next element from the small set of integers that are rich with respect to the previously chosen ones. From this, it will be easy to obtain an upper bound on the number of cores of a given size.

Claim 4.6. *Suppose that I is a core of size m. Then there are at least $m!/2$ orderings $(i_1, \dots, i_m)$ of the elements of I such that all but at most*

$$\left(\frac{K'' \log(1/p)}{Np^{k/2}}\right)^{\frac{1}{k-1}} \cdot m \tag{23}$$

indices $m' \in [\![m]\!]$ are rich.

Proof. Let $(i_1, \dots, i_m)$ be a uniformly chosen random ordering of the elements of I. Fix integers $m' \in [\![m]\!]$ and $i \in I$ and condition on the event $\{i_{m'} = i\}$. Under this conditioning, the set $\{i_1, \dots, i_{m'-1}\}$ is a uniformly random $(m'-1)$-element subset of $I \setminus \{i\}$. Therefore, we may use Janson's inequality for the hypergeometric distribution (Lemma 4.4) to get an upper bound for the probability that the given m' is not rich. It follows from the definition that $m' = 1$ is trivially rich, so assume $m' \geqslant 2$. Let $\mathcal{B}_i$ be the collection of all $(k-1)$-element subsets of $I \setminus \{i\}$ that form a k-AP with i. Define

$$\mu_{m'}(i) = \sum_{B \in \mathcal{B}_i} \left(\frac{m'-1}{m-1}\right)^{|B|} = A_k(I;i) \cdot \left(\frac{m'-1}{m-1}\right)^{k-1}$$

and, writing $B \sim B'$ to mean that $B \neq B'$ and $B \cap B' \neq \varnothing$,

$$\Delta_{m'}(i) = \sum_{\substack{B, B' \in \mathcal{B}_i \\ B \sim B'}} \left(\frac{m'-1}{m-1}\right)^{|B \cup B'|}.$$

Since for a given $B \in \mathcal{B}_i$, there are fewer than k^3 sets $B' \in \mathcal{B}_i$ such that $B \cap B' \neq \varnothing$, we have $\Delta_{m'}(i) \leqslant \mu_{m'}(i) \cdot k^3$. It follows from Claim 4.5 that

$$\mu_{m'}(i) \geqslant \frac{Np^{k/2}}{K' \log(1/p)} \cdot \left(\frac{m'-1}{m}\right)^{k-1},$$

which, provided that K'' is sufficiently large, is at least twice as large as the right-hand side of (21) with $|I'| = m' - 1$. Hence, by Lemma 4.4 with $\varepsilon = 1/2$,

$$\begin{aligned}\mathbb{P}\left(m' \text{ is not rich} \mid i_{m'} = i\right) &\leqslant 2\exp\left(-\frac{\mu_{m'}(i)^2}{8\big(\mu_{m'}(i) + \Delta_{m'}(i)\big)}\right)\\ &\leqslant 2\exp\left(-\frac{\mu_{m'}(i)}{9k^3}\right)\\ &\leqslant 2\exp\left(-\frac{Np^{k/2}}{9k^3K'\log(1/p)}\cdot\left(\frac{m'-1}{m}\right)^{k-1}\right).\end{aligned}$$

Since this upper bound is independent of i, then one may replace the conditional probability above with the unconditional one. Letting $\mathcal{X} \subseteq [\![m]\!]$ denote the (random) set of non-rich indices, we then find that

$$\mathbb{E}\left[|\mathcal{X}|\right] \leqslant 2\sum_{m'=2}^{m}\exp\left(-\frac{Np^{k/2}}{9k^3K'\log(1/p)}\cdot\left(\frac{m'-1}{m}\right)^{k-1}\right).$$

Since for every $\alpha > 0$, we have

$$\sum_{m'=2}^{m}\exp\left(-\left(\alpha\cdot\frac{m'-1}{m}\right)^{k-1}\right) \leqslant \int_0^\infty e^{-(\alpha x/m)^{k-1}}\,dx = \frac{m}{\alpha}\int_0^\infty e^{-y^{k-1}}\,dy \leqslant \frac{2m}{\alpha}$$

we obtain

$$\mathbb{E}\left[|\mathcal{X}|\right] \leqslant 4\cdot\left(\frac{9k^3K'\log(1/p)}{Np^{k/2}}\right)^{\frac{1}{k-1}}\cdot m.$$

The assertion of the claim now follows from Markov's inequality, provided that K'' is sufficiently large. $\square$

Equipped with the above facts, we can now prove the desired upper bound on the number of cores of size m. For a set $\mathcal{X} \subseteq [\![m]\!]$, let $\mathcal{S}_m(\mathcal{X})$ denote the family of all sequences of m distinct elements of $[\![N]\!]$ such that every index $m' \notin \mathcal{X}$ is rich. To control the number of sequences in $\mathcal{S}_m(\mathcal{X})$, note that we can pick the first element of the sequence arbitrarily and, for every subsequent index m', bound the number of possible values for the m'th element of the sequence either by appealing to (22), if $m' \notin \mathcal{X}$, or simply by N, otherwise. Thus,

$$\begin{aligned}\frac{|\mathcal{S}_m(\mathcal{X})|}{m!} &\leqslant \frac{1}{m!}\cdot N\cdot N^{|\mathcal{X}|}\cdot\prod_{m'=2}^{m}\left(kK'K''\cdot m\big(\log(1/p)\big)^2\cdot\left(\frac{m}{m'-1}\right)^{k-3}\right)\\ &\leqslant N\cdot N^{|\mathcal{X}|}\cdot\left(kK'K''\cdot\big(\log(1/p)\big)^2\right)^m\cdot\prod_{m'=1}^{m}\left(\frac{m}{m'}\right)^{k-2}.\end{aligned}$$

Since $\prod_{m'=1}^{m}\left(\frac{m}{m'}\right)^{k-2} \leqslant e^{(k-2)m}$, we find that, whenever C is sufficiently large,

$$\frac{|\mathcal{S}_m(\mathcal{X})|}{m!} \leqslant e^{(|\mathcal{X}|+1)\log N}\cdot e^{3m\log\log(1/p)}.$$

Finally, denote by $\mathcal{I}_m^*$ the set of all cores of size m. Claim 4.6 implies that

$$|\mathcal{I}_m^*| \leqslant \frac{2}{m!}\sum_{\mathcal{X}}\big|\mathcal{S}_m(\mathcal{X})\big| \leqslant 2^{m+1}\cdot\max_{\mathcal{X}}\frac{\big|\mathcal{S}_m(\mathcal{X})\big|}{m!},$$

where the sum and the maximum range over all $\mathcal{X} \subseteq [\![N]\!]$ of size at most $\left(\frac{K''\log(1/p)}{Np^{k/2}}\right)^{\frac{1}{k-1}} m$. Hence,

$$|\mathcal{I}_m^*| \leqslant 2^{m+1}\cdot\exp\left(\left(\frac{K''\log(1/p)}{Np^{k/2}}\right)^{\frac{1}{k-1}}\cdot m\cdot\log N + \log N\right)\cdot e^{3m\log\log(1/p)}.$$

Since we have assumed that $Np^{k/2} \geqslant C \log N$ and $p \leqslant C^{-2/k}$, if C is sufficiently large, then the above inequality implies that $|\mathcal{I}_m^*| \leqslant (1/p)^{\varepsilon m/2}$. This completes the proof of Proposition 4.1.

5. Counting small subgraphs—a graph-theoretic interlude

Recall that $\mathrm{Emb}(J, G)$ denotes the set of embeddings of J into G and, for every edge uv of G, $\mathrm{Emb}(J, G; uv)$ denotes the subset of $\mathrm{Emb}(J, G)$ containing all embeddings that map an edge of J to uv.

Definition 5.1. A *fractional independent set* in a graph J is an assignment $\alpha \colon V(J) \to [0, 1]$ that satisfies $\alpha_u + \alpha_v \leqslant 1$ for every edge uv of J. The *fractional independence number* of J, denoted by α_J^*, is the largest value of $\sum_{v \in V(J)} \alpha_v$ among all fractional independent sets α in J.

The following result is folklore.

Lemma 5.2. *Every graph J admits a fractional independent set α with $\sum_{v \in V(J)} \alpha_v = \alpha_J^*$ such that $\alpha_v \in \{0, \frac{1}{2}, 1\}$ for every $v \in V(J)$. Moreover, there is a partition $V(J) = V_1 \cup V_2$ with $|V_1|/2 + |V_2| = \alpha_J^*$ such that V_1 can be covered by a collection of vertex-disjoint edges and cycles of J.*

Proof. Let J' be the bipartite double cover of J, that is, the graph with vertex set $V(J) \times \{1, 2\}$ whose edges are all pairs $\{(u, 1), (v, 2)\}$ such that $uv \in E(J)$. Moreover, let $\pi \colon V(J) \times \{1, 2\} \to V(J)$ be the projection onto the first coordinate. The Kőnig–Egerváry theorem (see, e.g., [12, Theorem 8.32] or [26, Theorem 2.1.1]) implies that J' contains a matching M' and an independent set I' such that $|I'| + |M'| = v_{J'}$. Define $\alpha \colon V(J) \to \{0, \frac{1}{2}, 1\}$ by letting $\alpha_v = |\pi^{-1}(v) \cap I'|/2$ for every $v \in V(J)$. Since I' is an independent set in J', one can see that α is a fractional independent set with $\sum_{v \in V(J)} \alpha_v = |I'|/2$. In particular, we have

$$v_{J'} - |M'| = 2 \sum_{v \in V(J)} \alpha_v \leqslant 2\alpha_J^*. \tag{24}$$

Since π induces a projection of J' onto J, we can define $M = \pi(M')$ to be the image of the matching M'. Since $M \subseteq J$, we have $\alpha_J^* \leqslant \alpha_M^*$. Moreover, as M' is a matching in J', we see that M has maximum degree at most two and thus each nontrivial connected component of M is either a cycle or a path. Let $V_2 \subseteq V(J)$ comprise all isolated vertices of M and one arbitrarily chosen endpoint of each path of even length; let $V_1 = V(J) \setminus V_2$. By construction, each connected component of $M[V_1]$ is either a cycle or a path of odd length. Since the fractional independence number of every cycle and every path of odd length is exactly half its number of vertices, it follows that $\alpha_M^* \leqslant \alpha_{M[V_1]}^* + |V_2| = |V_1|/2 + |V_2|$. It is clear that V_1 can be covered with vertex-disjoint edges and cycles of M and thus also of J. We now claim that $|V_1| \geqslant |M'|$. To see this, fix a connected component L of M and observe that $\pi^{-1}(L)$ has at most e_L edges unless L is a single edge, in which case $\pi^{-1}(L)$ has at most two edges. Therefore,

$$e_{\pi^{-1}(L)} \leqslant \begin{cases} v_L - 1 & \text{if } L \text{ is a path of length at least two,} \\ v_L & \text{otherwise.} \end{cases}$$

Let $\mathcal{C}(M)$ denote the nontrivial connected components of M. We have

$$\begin{aligned} |M'| = \sum_{L \in \mathcal{C}(M)} e_{\pi^{-1}(L)} &\leqslant \sum_{L \in \mathcal{C}(M)} v_L - \mathbb{1}[L \text{ is a path of length at least two}] \\ &\leqslant \sum_{L \in \mathcal{C}(M)} v_L - \mathbb{1}[L \text{ is a path of even length}] = |V_1|. \end{aligned}$$

Consequently, (24) shows that

$$|V_1| + 2|V_2| = 2v_J - |V_1| \leqslant 2v_J - |M'| = v_{J'} - |M'| = 2 \sum_{v \in V(J)} \alpha_v \leqslant 2\alpha_J^* \leqslant 2\alpha_M^* \leqslant |V_1| + 2|V_2|,$$

and so $\sum_{v \in V(J)} \alpha_v = \alpha_J^* = |V_1|/2 + |V_2|$. □

Lemma 5.3. *Suppose that J is a nonempty subgraph of a connected, Δ-regular graph H. Then*

$$e_J \leqslant \Delta \cdot (v_J - \alpha_J^*) \leqslant \Delta \cdot \alpha_J^*.$$

If the first inequality is tight, then

(Q1) $J = H$ or

(Q2) J admits a bipartition $V(J) = A \cup B$ such that $\deg_J a = \Delta$ for all $a \in A$.

If both inequalities are tight, then $J = H$.

Proof. By Lemma 5.2, J has a fractional independent set α such that $\alpha_v \in \{0, \frac{1}{2}, 1\}$. Then

$$e_J \leqslant \sum_{uv \in E(J)} (2 - \alpha_u - \alpha_v) = \sum_{v \in V(J)} (1 - \alpha_v) \deg_J v \leqslant \Delta \cdot \sum_{v \in V(J)} (1 - \alpha_v) = \Delta \cdot (v_J - \alpha_J^*), \quad (25)$$

which is the first inequality. For the second inequality, note that the function $\alpha \colon V(J) \to [0,1]$ defined by $\alpha_v = 1/2$ is a fractional independent set, so $\alpha_J^* \geqslant v_J/2$.

Assume now that $e_J = \Delta \cdot (v_J - \alpha_J^*)$. Then both inequalities in (25) are equalities; this implies $\alpha_u + \alpha_v = 1$ for every edge $uv \in E(J)$ and $\deg_J(v) = \Delta$ whenever $\alpha_v \neq 1$. Let A, B, and C denote the sets of vertices that α maps to 0, 1, and 1/2, respectively. Each vertex in $A \cup C$ has degree Δ and each edge of J has either both endpoints in C or one endpoint in each of A and B. In particular, if C is not empty, then it induces a Δ-regular graph and hence $C = V(J)$ and $J = H$, as H is connected and Δ-regular. Otherwise, if C is empty, then $A \cup B$ is a bipartition of J and all vertices of A have degree Δ.

Lastly, suppose that $e_J = \Delta \cdot \alpha_J^*$, which implies $e_J = \Delta \cdot (v_J - \alpha_J^*)$. Let A, B, C be the same partition as above. If C is nonempty, then J is Δ-regular, and we are done. Otherwise,

$$|A| = e_J/\Delta = \alpha_J^* = v_J - e_J/\Delta = v_J - |A| = |B|.$$

Therefore, every vertex of B has degree Δ and $J = H$. □

Remark 5.4. Since every graph J is a subgraph of the complete graph of v_J vertices, Lemma 5.3 implies that

$$e_J \leqslant (v_J - 1) \cdot (v_J - \alpha_J^*).$$

Moreover, equality holds if and only if J is complete, J is empty, or $J = K_{1, v_J - 1}$.

The following result is due to Alon [1]. We provide a short proof for the sake of completeness.

Lemma 5.5. *Let C_ℓ denote the cycle of length ℓ. For every $\ell \geqslant 3$ and every graph G,*

$$|\operatorname{Emb}(C_\ell, G)| \leqslant (2e_G)^{\ell/2}.$$

Remark 5.6. If ℓ is even, this follows immediately from the fact that C_ℓ contains a perfect matching of $\ell/2$ edges. If ℓ is prime, there is also a very short and pretty proof using the monotonicity of L^p norms; see [53] for this proof and more precise estimates. The proof presented below works for all $\ell \geqslant 3$.

Proof. For each edge $e \in E(G)$, denote by c_e the number of copies of C_ℓ in G that contain the edge e. Since $\sum_{e \in E(G)} c_e = \ell \cdot N(C_\ell, G)$, where $N(C_\ell, G)$ is the number of copies of C_ℓ in G, it follows from the Cauchy–Schwarz inequality that

$$|\operatorname{Emb}(C_\ell, G)|^2 = \big(2\ell \cdot N(C_\ell, G)\big)^2 \leqslant 2e_G \cdot \sum_{e \in E(G)} 2c_e^2.$$

Let C_ℓ^* be the graph obtained from gluing two copies of C_ℓ along an edge. In other words, C_ℓ^* is obtained from the cycle of length $2\ell - 2$ by adding to it one longest chord. Observe that if (L_1, L_2) is an ordered pair of copies of C_ℓ in G, both containing e, then there are at exactly two homomorphisms $\varphi \colon V(C_\ell^*) \to V(G)$ that map the two vertices of degree three in C_ℓ^* onto the endpoints of e and the two copies of C_ℓ in C_ℓ^* onto L_1 and L_2, respectively. Letting $\mathrm{Hom}(C_\ell^*, G)$ be the collection of all homomorphisms from C_ℓ^* to G, we may conclude that

$$\sum_{e \in E(G)} 2c_e^2 \leqslant |\operatorname{Hom}(C_\ell^*, G)| \leqslant |\operatorname{Hom}\big((\ell-1)\cdot K_2, G\big)| \leqslant (2e_G)^{\ell-1},$$

where the second inequality holds because C_ℓ^* contains a perfect matching of $\ell - 1$ edges. □

The following theorem of Janson, Oleszkiewicz, and Ruciński [38] is a straightforward consequence of Lemmas 5.2 and 5.5. A closely related bound that does not depend on the number of vertices in G was obtained earlier by Alon [1] (see also [30] for a short proof).

Theorem 5.7 ([38])**.** *For every nonempty graph J without isolated vertices and every graph G with n vertices,*

$$|\operatorname{Emb}(J, G)| \leqslant (2e_G)^{v_J - \alpha_J^*} \cdot \min\{2e_G, n\}^{2\alpha_J^* - v_J}$$

Proof. By Lemma 5.2, there is a partition of $V(J)$ into V_1 and V_2 such that $|V_1|/2 + |V_2| = \alpha_J^*$ and V_1 can be covered by a collection $\mathcal{C}$ of vertex-disjoint edges and cycles of J. Let J' be the spanning subgraph comprising the edges and cycles of $\mathcal{C}$ and one edge incident to every vertex in V_2. We claim that

$$|\operatorname{Emb}(J', G)| \leqslant \prod_{C \in \mathcal{C}} |\operatorname{Emb}(C, G)| \cdot \min\{2e_G, n\}^{|V_2|}.$$

Indeed, every embedding of $J'[V_1]$ decomposes into embeddings of the graphs in $\mathcal{C}$, and there are at most $\min\{2e_G, n\}$ possible images for every vertex of V_2. By Lemma 5.5, for every cycle $C \in \mathcal{C}$,

$$|\operatorname{Emb}(C, G)| \leqslant (2e_G)^{v_C/2};$$

the same inequality holds when C is a single edge. Since every embedding of J into G is also an embedding of J', we deduce that

$$|\operatorname{Emb}(J, G)| \leqslant \prod_{C \in \mathcal{C}} (2e_G)^{v_C/2} \cdot \min\{2e_G, n\}^{|V_2|} = (2e_G)^{|V_1|/2} \cdot \min\{2e_G, n\}^{|V_2|}.$$

Since $|V_1|/2 = v_J - \alpha_J^*$ and $|V_2| = 2\alpha_J^* - v_J$, this completes the proof. □

Observe that, when J is the complete graph, then Theorem 5.7 yields the upper bound $|\operatorname{Emb}(K_r, G)| \leqslant (2e_G)^{r/2}$. This is a weak version of a more precise result due to Erdős–Hanani [29] and also follows from the Kruskal–Katona theorem [44, 49]. One can see that the upper bound $(2e_G)^{r/2}$ is asymptotically optimal if G contains a clique comprising all of its edges. Our next theorem states that, when the upper bound given in Theorem 5.7 is nearly tight, then G resembles such a graph, in the sense that it must contain a subgraph of density $1 - o(1)$ covering nearly all of its edges. This could be proved by appealing to a stability version of the Kruskal–Katona theorem due to Keevash [45]. The proof we present below is elementary.

Theorem 5.8. *Suppose that $r \geqslant 3$. If a graph G satisfies*

$$|\operatorname{Emb}(K_r, G)| \geqslant (1 - \varepsilon) \cdot (2e_G)^{r/2},$$

for some $\varepsilon \geqslant e_G^{-1/2}$, then G has a subgraph G' with minimum degree at least $(1 - 4\varepsilon^{1/2}) \cdot (2e_G)^{1/2}$.

Proof. The assertion of the theorem follows once we establish the case $r = 3$ and an analogous property for the path with four vertices (and three edges), which we denote by P_4. Indeed, if r is odd, then K_r contains a spanning subgraph that is the disjoint union of K_3 and a matching of size $(r-3)/2$. Thus,

$$|\operatorname{Emb}(K_r, G)| \leqslant |\operatorname{Emb}(K_3, G)| \cdot |\operatorname{Emb}(K_2, G)|^{(r-3)/2} = |\operatorname{Emb}(K_3, G)| \cdot (2e_G)^{(r-3)/2}$$

and hence $|\operatorname{Emb}(K_3, G)| \geqslant (1-\varepsilon)\cdot(2e_G)^{3/2}$. Analogously, if r is even, then K_r contains a subgraph that is the disjoint union of P_4 and a matching of size $(r-4)/2$. Thus,

$$|\operatorname{Emb}(K_r, G)| \leqslant |\operatorname{Emb}(P_4, G)| \cdot |\operatorname{Emb}(K_2, G)|^{(r-4)/2} = |\operatorname{Emb}(P_4, G)| \cdot (2e_G)^{(r-4)/2},$$

which implies that $|\operatorname{Emb}(P_4, G)| \geqslant (1-\varepsilon)\cdot(2e_G)^2$. Therefore, it suffices to prove the following two claims.

Claim 5.9. *If $|\operatorname{Emb}(P_4, G)| \geqslant (1-\varepsilon)\cdot(2e_G)^2$, for some positive ε, then G has a subgraph G' with minimum degree at least $(1-2\varepsilon^{1/2})(2e_G)^{1/2}$.*

Claim 5.10. *If $|\operatorname{Emb}(K_3, G)| \geqslant (1-\varepsilon)\cdot(2e_G)^{3/2}$, for some $\varepsilon \geqslant e_G^{-1/2}$, then G has a subgraph G' with minimum degree at least $(1-4\varepsilon^{1/2})(2e_G)^{1/2}$.*

Proof of Claim 5.9. We may assume that $\varepsilon < 1/4$, as otherwise the assertion of the claim is trivially satisfied. Let F be the graph with vertex set $E(G)$ whose edges are all pairs $\{uv, xy\}$ such that the set $\{u, v, x, y\}$ induces a K_4 in G. Let $\mathbb{1}_G$ be the indicator function of the edge set of G and note that

$$\begin{aligned} |\operatorname{Emb}(P_4, G)| &= \sum_{\substack{\{uv,xy\}\subseteq E(G)\\ \{u,v\}\cap\{x,y\}=\varnothing}} 2\cdot\big(\mathbb{1}_G(ux)+\mathbb{1}_G(uy)+\mathbb{1}_G(vx)+\mathbb{1}_G(vy)\big)\\ &\leqslant 8e_F + 6\left(\binom{e_G}{2} - e_F\right) \leqslant 3e_G^2 + 2e_F. \end{aligned}$$

In particular, our assumption implies that $e_F \geqslant (1/2 - 2\varepsilon)\cdot e_G^2$.

Let F' be the subgraph obtained from F by iteratively removing vertices whose degree is smaller than $(1-2\varepsilon^{1/2})\cdot e_G$. We claim that fewer than $2\varepsilon^{1/2}\cdot e_G$ vertices are removed this way and, consequently, the graph F' is nonempty and its minimum degree is at least $(1-2\varepsilon^{1/2})\cdot e_G$. Suppose that this were not true. We would then have

$$e_F \leqslant \binom{(1-2\varepsilon^{1/2})\cdot e_G}{2} + 2\varepsilon^{1/2}\cdot e_G\cdot(1-2\varepsilon^{1/2})\cdot e_G < \left(\frac{1}{2} - 2\varepsilon\right)\cdot e_G^2 \leqslant e_F,$$

a contradiction.

Finally, let G' be the subgraph of G induced by the set of endpoints of the edges from $V(F')$. Let u be an arbitrary vertex of G'. There must be another vertex v of G' such that $uv \in V(F')$. Since $\deg_{F'} uv \geqslant (1-2\varepsilon^{1/2})\cdot e_G$, the common neighbourhood of u and v in G' induces a subgraph with at least $(1-2\varepsilon^{1/2})\cdot e_G$ edges in G. In particular,

$$\deg_{G'} u \geqslant \deg_{G'}(u, v) \geqslant \sqrt{2\cdot(1-2\varepsilon^{1/2})\cdot e_G} \geqslant (1-2\varepsilon^{1/2})\cdot(2e_G)^{1/2}.$$

Since u was arbitrary, we obtain the desired lower bound on the minimum degree of G'. □

Proof of Claim 5.9. We may assume that $\varepsilon < 1/16$, as otherwise the assertion of the claim is trivially satisfied. For every edge $e \in E(G)$, let t_e denote the number of copies of K_3 in G that contain the edge e. Observe that, for each $e \in E(G)$, there are at least $2t_e(t_e-1)$ embeddings of P_4 into G that map the middle edge of P_4 onto e. Since $\sum_{e\in E(G)} t_e = |\operatorname{Emb}(K_3, G)|/2$ and the function $t \mapsto 2t(t-1)$ is convex, we conclude that

$$|\operatorname{Emb}(P_4, G)| \geqslant \sum_{e\in E(G)} 2t_e(t_e-1) \geqslant e_G\cdot\frac{|\operatorname{Emb}(K_3, G)|}{e_G}\cdot\left(\frac{|\operatorname{Emb}(K_3, G)|}{2e_G} - 1\right).$$

Our assumptions imply that

$$1 \leqslant \varepsilon\cdot e_G^{1/2} \leqslant \varepsilon\cdot\frac{(1-\varepsilon)\cdot(2e_G)^{3/2}}{2e_G} \leqslant \varepsilon\cdot\frac{|\operatorname{Emb}(K_3, G)|}{2e_G}$$

and consequently,

$$|\operatorname{Emb}(P_4, G)| \geqslant (1-\varepsilon) \cdot \frac{|\operatorname{Emb}(K_3, G)|^2}{2e_G} \geqslant (1-\varepsilon)^3 \cdot (2e_G)^2 \geqslant (1-3\varepsilon) \cdot (2e_G)^2,$$

It now follows from Claim 5.9 that G contains a subgraph G' with minimum degree at least

$$\big(1 - 2 \cdot (3\varepsilon)^{1/2}\big) \cdot (2e_G)^{1/2} \geqslant (1 - 4\varepsilon^{1/2}) \cdot (2e_G)^{1/2},$$

as claimed. □□

Our next lemma gives a tight upper bound on the number of stars $K_{1,s}$ in a given bipartite graph, as well as a structural characterisation of the bipartite graphs that are close to achieving this bound. This lemma and Theorem 5.8 above constitute the main combinatorial ingredient in the proof of Theorem 1.8. Given a graph G and a set U of vertices of G, we let $\operatorname{Emb}_U(K_{1,s}, G)$ denote the set of embeddings of $K_{1,s}$ into G that map the centre vertex to a vertex of U.

Lemma 5.11. *Let $s \geqslant 2$ be an integer and suppose that G is a bipartite graph with parts U and V and at most $q|V|$ edges, for some $q \in (0, |U|]$. Then the following holds:*

(i)

$$|\operatorname{Emb}_U(K_{1,s}, G)| \leqslant \big(\lfloor q \rfloor + \{q\}^s\big)|V|^s.$$

(ii) For every positive ε, there exists a positive η such that, if

$$|\operatorname{Emb}_U(K_{1,s}, G)| \geqslant (1-\eta) \cdot \big(\lfloor q \rfloor + \{q\}^s\big)|V|^s,$$

then there is a subset $W \subseteq U$ of size $\lceil q \rceil$ such that $e_G(W, V) \geqslant (1-\varepsilon)q|V|$, and a further subset $W' \subseteq W$ of size at least $\lfloor (1-\varepsilon)|W| \rfloor$ such that $\deg_G u \geqslant (1-\varepsilon)|V|$ for every $u \in W'$.

Proof. We will use the following inequality, valid for any two numbers x and y with $x \geqslant y \geqslant 1$:

$$(x+1)^s + (y-1)^s - x^s - y^s \geqslant (x+1-y) \cdot (x+1)^{s-2}. \tag{26}$$

It is clear that

$$|\operatorname{Emb}_U(K_{1,s}, G)| = (\deg_G u)(\deg_G u - 1) \cdots (\deg_G u - s + 1) \leqslant \sum_{u \in U} (\deg_G u)^s. \tag{27}$$

Let $m = |U|$ and, given a sequence $\mathbf{d} = (d_1, \ldots, d_m)$, define

$$S(\mathbf{d}) = \sum_{i=1}^{m} d_i^s.$$

By the degree sequence of a bipartite graph with parts U and V, we will mean the sequence of degrees of the vertices in U, listed in a nonincreasing order. Thus (27) implies that $|\operatorname{Emb}_U(K_{1,s}, G)| \leqslant S(\mathbf{d}_G)$, where $\mathbf{d}_G$ is the degree sequence of G. Let $m_0 = \lfloor e_G/|V| \rfloor$ and define $\mathbf{d}_{\max} = (d_1^*, \ldots, d_m^*)$ by

$$d_i^* = \begin{cases} |V| & \text{if } i \leqslant m_0, \\ \{e_G/|V|\} \cdot |V| & \text{if } i = m_0 + 1, \\ 0 & \text{otherwise.} \end{cases}$$

Note that $\sum_{i=1}^m d_i^* = e_G$; in particular, $\{e_G/|V|\} \cdot |V|$ is an integer. We claim that $\mathbf{d}_{\max}$ maximises S over all degree sequences whose sum is e_G. Indeed, for any other such degree sequence $\mathbf{d}' = (d'_1, \ldots, d'_m)$, there must be two distinct indices i and j such that $0 < d'_i \leqslant d'_j < |V|$. Let $\mathbf{d}''$ be the degree sequence obtained from $\mathbf{d}'$ by decreasing d'_i by one and increasing d'_j by one (and reordering the degrees, if necessary). It follows from (26) that

$$S(\mathbf{d}'') - S(\mathbf{d}') = (d'_j + 1)^s + (d'_i - 1)^s - (d'_j)^s - (d'_i)^s \geqslant (d'_j - d'_i + 1) \cdot (d'_j + 1)^{s-2} \geqslant 1.$$

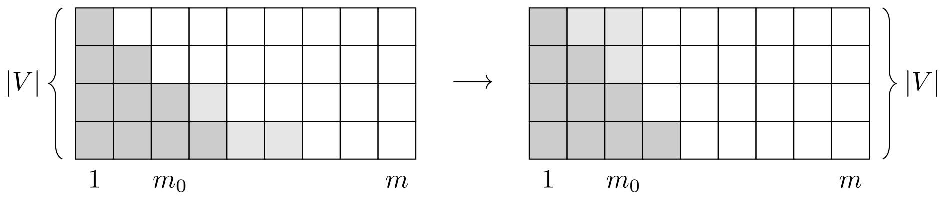

FIGURE 4. The degree sequence on the left can be turned into the degree sequence on the right by moving mass from columns $i \geqslant m_0 + 1$ to columns $j \leqslant m_0 + 1$.

Therefore,

$$|\operatorname{Emb}_U(K_{1,s}, G)| \leqslant S(\mathbf{d}_G) \leqslant S(\mathbf{d}_{\max}) = \big(\lfloor e_G/|V| \rfloor + \{e_G/|V|\}^s\big)|V|^s.$$

Since $e_G \leqslant q|V|$, this completes the proof of the first part of the lemma.

For the second stipulation of the lemma, fix a positive ε. Let Ω be the set of degree sequences of bipartite graphs G with parts U and V for which there is a subset $W \subseteq U$ of size $\lceil e_G/|V| \rceil$ such that $e_G(W, V) \geqslant (1-\varepsilon)e_G$ and, additionally, a further subset $W' \subseteq W$ of size $\lfloor (1-\varepsilon)|W| \rfloor$ such that $\deg_G u \geqslant (1-\varepsilon)|V|$ for each $u \in W'$.

It suffices to show that any degree sequence $\mathbf{d} \notin \Omega$ whose sum is e_G satisfies $S(\mathbf{d}) < (1-\eta)S(\mathbf{d}_{\max})$, for some positive η that depends only on ε. Let $\mathbf{d} = (d_1, \ldots, d_m)$. The crucial observation is that we may obtain $\mathbf{d}_{\max}$ from $\mathbf{d}$ by successively increasing some d_i with $i \leqslant m_0 + 1$ by one and, simultaneously, decreasing some d_j with $j \geqslant m_0 + 1$ by one (see Figure 4). Note that, when doing so, we perform at least

$$\sum_{i=1}^{m_0} (|V| - d_i) + \max\left\{d^*_{m_0+1} - d_{m_0+1}, 0\right\}$$

such operations. We split further analysis into two cases.

First, assume that $\sum_{i=1}^{m_0}(|V| - d_i) \geqslant \varepsilon^2 e_G$; this implies that $m_0 \geqslant 1$. In this case, in at least $\varepsilon^2 e_G/2$ steps of the above procedure, we will be increasing some d_i with $i \leqslant m_0$ which is, at this time, already at least $\lfloor (|V| + d_i)/2 \rfloor \geqslant (|V| + d_{m_0})/2 - 1$, while decreasing some d_j with $j \geqslant m_0 + 1$ which is at most d_{m_0}. Inequality (26) implies that

$$S(\mathbf{d}_{\max}) - S(\mathbf{d}) \geqslant \frac{\varepsilon^2 e_G}{2} \cdot \left(\frac{|V| + d_{m_0}}{2} - d_{m_0}\right) \cdot \left(\frac{|V| + d_{m_0}}{2}\right)^{s-2}.$$

Since our assumption implies that $|V| - d_{m_0} \geqslant \varepsilon^2 e_G/m_0$, it follows that

$$S(\mathbf{d}_{\max}) - S(\mathbf{d}) \geqslant \frac{\varepsilon^2 e_G}{2} \cdot \frac{\varepsilon^2 e_G}{2m_0} \cdot \left(\frac{|V|}{2}\right)^{s-2} = \frac{\varepsilon^4}{2^{s+1}} \cdot \left(\frac{e_G}{m_0|V|}\right)^2 \cdot 2m_0|V|^s.$$

Finally, we note that $e_G \geqslant m_0|V|$ and that $S(\mathbf{d}_{\max}) \leqslant (m_0+1)|V|^s \leqslant 2m_0|V|^s$, and thus

$$S(\mathbf{d}_{\max}) - S(\mathbf{d}) \geqslant \frac{\varepsilon^4}{2^{s+1}} \cdot S(\mathbf{d}_{\max}),$$

proving $S(\mathbf{d}) < (1-\eta)S(\mathbf{d}_{\max})$ for some positive η.

Assume now that $\sum_{i=1}^{m_0}(|V| - d_i) < \varepsilon^2 e_G$. Let W be the set comprising the $\lceil e_G/|V| \rceil$ vertices with largest degrees in U. Suppose first that $e_G(W, V) \geqslant (1-\varepsilon)e_G$, but every set of $\lfloor (1-\varepsilon)|W| \rfloor$ vertices of W contains a vertex of degree smaller than $(1-\varepsilon)|V|$. In this case, $(1-\varepsilon)|W| \geqslant 1$, as otherwise the latter condition is vacuously false, and $d_i < (1-\varepsilon)|V|$ whenever $i \geqslant \lfloor (1-\varepsilon)|W| \rfloor$. Then

$$\sum_{i=1}^{m_0}(|V| - d_i) > \big(m_0 + 1 - \lfloor (1-\varepsilon)|W| \rfloor\big) \cdot \varepsilon|V| \geqslant \big(|W| - \lfloor (1-\varepsilon)|W| \rfloor\big) \cdot \varepsilon|V| \geqslant \varepsilon^2 e_G,$$

contradicting our assumption. Thus, since $\mathbf{d} \notin \Omega$, we may assume that $e_G(W,V) < (1-\varepsilon)e_G$. Therefore,

$$(1-\varepsilon)e_G > \sum_{i=1}^{|W|} d_i \geqslant \sum_{i=1}^{m_0} d_i = m_0|V| - \sum_{i=1}^{m_0}(|V| - d_i) > m_0|V| - \varepsilon^2 e_G.$$

This means, in particular, that $m_0 < e_G/|V|$ and hence $|W| = m_0 + 1$. Moreover,

$$m_0|V| + d^*_{m_0+1} = e_G > \sum_{i=1}^{|W|} d_i + \varepsilon e_G > m_0|V| + d_{m_0+1} + (\varepsilon - \varepsilon^2)e_G,$$

which implies that $d^*_{m_0+1} - d_{m_0+1} > (\varepsilon - \varepsilon^2)e_G$. Therefore, there exist at least $(\varepsilon - \varepsilon^2)e_G/2$ steps in which we increase d_{m_0+1} at a time where it is already at least $\lfloor (\varepsilon - \varepsilon^2)e_G/2 \rfloor$. Inequality (26) implies that

$$S(\mathbf{d}_{\max}) - S(\mathbf{d}) \geqslant \left(\frac{(\varepsilon - \varepsilon^2)e_G}{2} \right)^s.$$

However, we trivially have $S(\mathbf{d}_{\max}) \leqslant e_G^s$ and thus

$$S(\mathbf{d}_{\max}) - S(\mathbf{d}) \geqslant \frac{(\varepsilon - \varepsilon^2)^s}{2^s} \cdot S(\mathbf{d}_{\max}),$$

completing the proof. □

We remark that the extremal structures given by Theorem 5.8 and Lemma 5.11(ii) are quite different and, in a sense, incompatible. This has the following technically important consequence: if a graph G simultaneously contains many copies of K_r and many copies of $K_{1,r-1}$, then it can be split into two edge-disjoint graphs, one containing nearly all the copies of K_r and the other containing nearly all the copies of $K_{1,r-1}$. The following lemma formalises this statement; its proof is similar to an argument of Lubetzky and Zhao [51].

Lemma 5.12. *For every integer $r \geqslant 3$ and positive real number ε, there is a positive η such that the following holds. Let G be a graph on n vertices with $e_G \leqslant \eta n^2$. Then there exists a partition $V(G) = U \cup V$ satisfying $|U| \leqslant \varepsilon n$,*

$$|\operatorname{Emb}(K_r, G[V])| \geqslant |\operatorname{Emb}(K_r, G)| - \varepsilon e_G^{r/2},$$

and

$$|\operatorname{Emb}_U(K_{1,r-1}, G[U,V])| \geqslant |\operatorname{Emb}(K_{1,r-1}, G)| - \varepsilon e_G n^{r-2}.$$

Proof. Assume that η is sufficiently small, let U be the set of vertices in G with degree at least $\eta^{1/3} n$, and let $V = V(G) \setminus U$. Note that $|U| \leqslant 2e_G/(\eta^{1/3}n) \leqslant 2\eta^{2/3}n \leqslant \varepsilon n$.

Every embedding of K_r into G that maps a vertex of K_r to a vertex of U can be specified by first choosing a vertex v of K_r, then a vertex of U that v will be mapped to, and finally an embedding of K_{r-1} into G. Using Theorem 5.7, we thus obtain

$$|\operatorname{Emb}(K_r, G)| - |\operatorname{Emb}(K_r, G[V])| \leqslant r \cdot |U| \cdot |\operatorname{Emb}(K_{r-1}, G)| \leqslant r \cdot \frac{2e_G}{\eta^{1/3}n} \cdot (2e_G)^{(r-1)/2}.$$

Since $e_G^{1/2} \leqslant \eta^{1/2} n$, this implies the first assertion of the lemma.

Next, note that

$$|\operatorname{Emb}(K_{1,r-1}, G)| = |\operatorname{Emb}_U(K_{1,r-1}, G[U,V])| + t_1 + t_2, \tag{28}$$

where t_1 is the number of embeddings of $K_{1,r-1}$ into G that map the centre vertex and at least one leaf of $K_{1,r-1}$ to U and t_2 is the number of embeddings that map the centre vertex of $K_{1,r-1}$ to V. We have

$$t_1 \leqslant (r-1) \cdot |U|^2 \cdot n^{r-2} \leqslant (r-1) \left(\frac{2e_G}{\eta^{1/3}n} \right)^2 n^{r-2} \leqslant \varepsilon e_G n^{r-2}/2,$$

as $e_G \leqslant \eta n^2$. Finally, in order to bound t_2, observe that every embedding counted by t_2 can be specified by first choosing a leaf a of $K_{1,r-1}$, then choosing the image e of the edge of $K_{1,r-1}$ incident with a, then choosing the endpoint $v \in V$ of e that is the image of the centre vertex of $K_{1,r-1}$, and finally choosing the images of the remaining $r-2$ leaves of $K_{1,r-1}$ among the neighbours of v in G. Since every vertex $v \in V$ has degree at most $\eta^{1/3}n$, it follows that

$$t_2 \leqslant (r-1) \cdot e_G \cdot 2 \cdot (\eta^{1/3}n)^{r-2} \leqslant \varepsilon e_G n^{r-2}/2.$$

Together with (28), these bounds on t_1 and t_2 imply the second assertion of the lemma. □

We conclude this section with three lemmas that bound $|\operatorname{Emb}(J, G; uv)|$ from above.

Lemma 5.13. *Suppose that H is a Δ-regular graph. For every graph G and each $uv \in E(G)$,*

$$|\operatorname{Emb}(H, G; uv)| \leqslant 4e_H \cdot (2e_G)^{\frac{v_H}{2} - \frac{2\Delta-1}{\Delta}} \cdot (4 \deg_G u \cdot \deg_G v)^{\frac{\Delta-1}{\Delta}}.$$

Lemma 5.14. *Suppose that J is a nonempty, connected graph with maximum degree Δ that admits a bipartition $V(J) = A \cup B$ such that $|A| < |B|$ and $\deg_J a = \Delta$ for every $a \in A$. For every graph G and every $uv \in E(G)$,*

$$|\operatorname{Emb}(J, G; uv)| \leqslant e_J \cdot (\deg_G u + \deg_G v) \cdot (2e_G)^{|A|-1} \cdot \big(\min\{e_G, v_G\}\big)^{|B|-|A|-1}.$$

Lemma 5.15. *Suppose that H is a Δ-regular graph. For every graph G and every $G' \subseteq G$,*

$$\sum_{uv \in E(G')} |\operatorname{Emb}(H, G; uv)| \leqslant e_H \cdot (2e_G)^{v_H/2} \cdot \left(\frac{e_{G'}}{e_G}\right)^{1/\Delta}.$$

Our proofs of Lemmas 5.13 and 5.15 are relatively straightforward adaptations of the elegant entropy argument of Friedgut and Kahn [30] (see also the excellent survey of Galvin [31]). They will be derived from the following somewhat abstract form of the main result of [30]. The proof of Lemma 5.14 is elementary.

Lemma 5.16. *Suppose that H is a Δ-regular graph. Let $\mathcal{E}$ be a family of embeddings of H into a graph G and, for every edge ab of H, let*

$$\mathcal{E}_{ab} = \{\varphi(ab) : \varphi \in \mathcal{E}\}.$$

Then

$$|\mathcal{E}| \leqslant \prod_{ab \in E(H)} \big(2|\mathcal{E}_{ab}|\big)^{1/\Delta}.$$

Proof. Let $\bar{\varphi}\colon V(H) \to V(G)$ be a uniformly chosen random element of $\mathcal{E}$. Write $H(Z)$ for the entropy of a discrete random variable Z and observe that $H(\bar{\varphi}) = \log|\mathcal{E}|$. Since H is Δ-regular, Shearer's inequality [21] implies that

$$H(\bar{\varphi}) \leqslant \frac{1}{\Delta} \cdot \sum_{ab \in E(H)} H\big(\bar{\varphi}(a), \bar{\varphi}(b)\big).$$

The random variable $(\bar{\varphi}(a), \bar{\varphi}(b))$ can take on at most $2|\mathcal{E}_{ab}|$ values, as it an ordered pair of vertices that make up the edge $\varphi(ab)$. Using the fact that the entropy of any distribution on a set is at most that of the uniform distribution on that set, it follows that $H\big(\bar{\varphi}(a), \bar{\varphi}(b)\big) \leqslant \log(2|\mathcal{E}_{ab}|)$. This implies the assertion of the lemma. □

Proof of Lemma 5.13. Given an ordered pair (c, d) of adjacent vertices of H, let $\mathcal{E}^{(c,d)}$ be the family of embeddings φ of H into G such that $\varphi(c) = u$ and $\varphi(d) = v$. For a given edge ab of H, define $\mathcal{E}_{ab}^{(c,d)} = \{\varphi(ab) : \varphi \in \mathcal{E}^{(c,d)}\}$ as in the statement of Lemma 5.16. Observe that

$$|\mathcal{E}_{ab}^{(c,d)}| \leqslant \begin{cases} e_G & \text{if } \{a,b\} \cap \{c,d\} = \varnothing, \\ \deg_G u & \text{if } \{a,b\} \cap \{c,d\} = \{c\}, \\ \deg_G v & \text{if } \{a,b\} \cap \{c,d\} = \{d\}, \\ 1 & \text{if } \{a,b\} = \{c,d\}. \end{cases}$$

Invoking Lemma 5.16 to bound $|\mathcal{E}^{(c,d)}|$ from above and summing over all $2e_H$ pairs (c,d) of adjacent vertices of H, the claimed upper bound on the number of embeddings of H into G that use the edge uv. ☐

Proof of Lemma 5.14. We fix an edge ab of J, where $a \in A$ and $b \in B$, and count the embeddings φ of J into G such that $\varphi(ab) = uv$. To this end, we first show that $J - b$ contains a matching M that saturates A. To see this, note that for every nonempty $S \subseteq A$,

$$|S| \cdot \Delta = \sum_{c \in S} \deg_J c \leqslant \sum_{d \in N_J(S)} \deg_J d \leqslant |N_J(S)| \cdot \Delta,$$

yielding $|N_J(S)| \geqslant |S|$. Moreover, this inequality is strict unless the subgraph of J induced by $S \cup N_J(S)$ is Δ-regular. However, the latter is impossible since, as J is connected, the only Δ-regular subgraph of J could be J itself, but our assumption $|A| < |B|$ implies that J is not regular. Hence $|N_{J-b}(S)| \geqslant |N_J(S)| - 1 \geqslant |S|$, verifying Hall's condition. Now, given a matching $M \subseteq J - b$ that saturates A, we may bound the number of embeddings φ as above in the following way. Let c be the neighbour of a in M. There are at most $\deg_G u + \deg_G v$ embeddings of $J[\{a,b,c\}]$ into G that map ab to uv. Each of them admits at most $(2e_G)^{|M|-1}$ extensions to an embedding of $J[V(M) \cup \{b\}]$. Each of those embeddings can be extended to an embedding of J in at most $\min\{e_G, v_G\}^{|B \setminus (\{b\} \cup V(M))|}$ ways. Since $|M| = |A|$ and $|B \setminus (\{b\} \cup V(M))| = |B| - |A| - 1$, summing over all $ab \in E(J)$ gives the claimed bound on $|\operatorname{Emb}(J, G; uv)|$. ☐

Proof of Lemma 5.15. Given an edge cd of H, let $\mathcal{E}^{cd}$ be the family of embeddings of H into G that map cd onto an edge of G'. Define $\mathcal{E}_{ab}^{cd} = \{\varphi(ab) : \varphi \in \mathcal{E}^{cd}\}$ as in the statement of Lemma 5.16 and observe that

$$|\mathcal{E}_{ab}^{cd}| \leqslant \begin{cases} e_{G'} & \text{if } \{a,b\} = \{c,d\}, \\ e_G & \text{otherwise}. \end{cases}$$

Invoking Lemma 5.16 to bound $|\mathcal{E}^{cd}|$ from above and summing over all edges cd of H gives the claimed upper bound. ☐

6. Cliques in random graphs

Fix an integer $r \geqslant 3$ and let $X = X_{n,p}^{K_r}$ be the number of r-vertex cliques in the random graph $G_{n,p}$. In this section, we shall use Theorem 3.1 not only to determine the logarithmic upper tail probability of X but also to provide a detailed description of the upper tail event. Before we restate the two theorems that will be proved in this section, we discuss the combinatorial constructions that are responsible for the localisation phenomenon in more detail.

As was shown in [51], when $n^{-1} \ll p^{(r-1)/2} \ll 1$, there are essentially two optimal strategies for planting a subgraph inside $G_{n,p}$ that increases the expected number of copies of K_r by the required $\delta\,\mathbb{E}[X]$. The first, and most straightforward, involves planting a clique with $\delta^{1/r} n p^{(r-1)/2}$ vertices; note that our assumption on p implies that this expression is unbounded and thus we may implicitly assume that it is an integer. Note that such a clique has close to

$$\frac{\delta^{2/r} n^2 p^{r-1}}{2}$$

edges and contains approximately $\delta\binom{n}{r}p^{\binom{r}{2}} = \delta\,\mathbb{E}[X]$ copies of K_r. If np^{r-1} is bounded from below, then there is an alternative strategy that competes with planting a clique. By a *hub* of order $\lfloor \delta np^{r-1}/r \rfloor + 1$, we mean a subgraph of $G_{n,p}$ constructed as follows. Let U be a set of $\lfloor \delta np^{r-1}/r \rfloor$ vertices of $G_{n,p}$ and let u be another vertex that lies outside of U. Connect every vertex in U to every vertex outside of $U \cup \{u\}$ and connect u to some $\{\delta np^{r-1}/r\}^{1/(r-1)} \cdot n$ such vertices. Note that every hub has close to

$$\Big(\lfloor \delta np^{r-1}/r \rfloor + \{\delta np^{r-1}/r\}^{1/(r-1)}\Big) \cdot n$$

edges, which is $\Theta(n^2p^{r-1})$, as np^{r-1} is bounded from below. Unlike in the previous construction, the hub itself contains *no* copies of K_r. However, as $p \ll 1$, planting a hub creates approximately $\lfloor \delta np^{r-1}/r \rfloor \cdot \binom{n}{r-1}$ copies of the star graph $K_{1,r-1}$ whose centre vertex lies in U and approximately $\{\delta np^{r-1}/r\} \cdot \binom{n}{r-1}$ copies of $K_{1,r-1}$ whose centre vertex is u. The total number of planted copies of $K_{1,r-1}$ is thus approximately $\delta p^{r-1} \cdot \binom{n}{r} = \delta\,\mathbb{E}[X] \cdot p^{-\binom{r-1}{2}}$. Since each of the planted copies of $K_{1,r-1}$ lies in a copy of K_r that now appears in $G_{n,p}$ with probability $p^{\binom{r-1}{2}}$, one expects to see approximately $\delta\,\mathbb{E}[X]$ such extra copies of K_r. We remark that if np^{r-1} is large, then the contribution of the single vertex u becomes negligible and the hub construction can be described more concisely as connecting some $\delta np^{r-1}/r$ vertices to all the others, using $\delta n^2p^{r-1}/r$ edges.

We prove that, for a vast majority of values of p in the range of interest, the logarithmic upper tail probability of X corresponds to one of the two strategies described above. In particular, we show that

(i) If $np^{r-1} \to 0$, then the logarithmic upper tail probability is asymptotically equal to the 'cost' of planting the smallest clique that has $\delta\,\mathbb{E}[X]$ copies of K_r.
(ii) If $np^{r-1} \to \infty$, then the logarithmic upper tail probability is asymptotically equal to the 'cost' of planting either a clique as above (when $\delta^{2/r} \leqslant \delta/r$) or the smallest hub that has $\delta\,\mathbb{E}[X] \cdot p^{-\binom{r-1}{2}}$ copies of $K_{1,r-1}$ (when $\delta^{2/r} \geqslant \delta/r$).

Note that in the regime $np^{r-1} \to \infty$, we may approximate the number of edges planted in the hub construction by $\delta n^2p^{r-1}/r$. However, when $np^{r-1} \to c$ for some constant $c \in (0,\infty)$, this approximation is no longer valid and we are forced to account for the lack of smoothness that stems from the integral and fractional parts of $\delta c/r$. As a result, we find that, for certain values of the parameters δ and c, the logarithmic upper tail probability corresponds to a mixture of the first and second strategies: it is equal to the cost of planting a graph comprising both a hub and a clique, each contributing a nonnegligible proportion of the (expected) extra $\delta\,\mathbb{E}[X]$ copies of K_r.

Finally, suppose that one conditions $G_{n,p}$ on the upper tail event $\{X \geqslant (1+\delta)\,\mathbb{E}[X]\}$. We prove that, with probability close to one, the conditioned random graph contains a subgraph that very closely resembles the graph described by the optimal strategy (for the particular values of n, p, and δ). For example, in cases where the logarithmic upper tail probability corresponds to planting a clique, we show that $G_{n,p}$ conditioned on the event $\{X \geqslant (1+\delta)\,\mathbb{E}[X]\}$ contains a set of $\delta^{1/r}np^{(r-1)/2}$ vertices that induces an 'almost-clique', that is, a subgraph of density $1-o(1)$.

We now turn to the details. As in the introduction, we define continuous functions $\psi_r \colon (0,\infty)^2 \times [0,1] \to (0,\infty)$ and $\varphi_r \colon (0,\infty)^2 \to (0,\infty)$ by

$$\psi_r(\delta,c,x) = \frac{\big(\delta(1-x)\big)^{2/r}}{2} + \frac{\lfloor x\delta c/r \rfloor + \{x\delta c/r\}^{\frac{1}{r-1}}}{c} \quad \text{and} \quad \varphi_r(\delta,c) = \min_{x\in[0,1]} \psi_r(\delta,c,x).$$

Note that $\psi_r(\delta, np^{r-1}, x) \cdot n^2p^{r-1}$ is approximately the number of edges in the disjoint union of a clique with $\delta xnp^{(r-1)/2}$ vertices and a hub of order $\lfloor \delta(1-x)np^{r-1}/r \rfloor + 1$. Recalling the discussion above, $\varphi_r(\delta, np^{r-1}) \cdot n^2p^{r-1}$ then represents the smallest number of edges among all combinations of clique and hub that yield an expected $\delta\,\mathbb{E}[X]$ copies of K_{r-1}.

In order to handle the three cases $np^{r-1} \to 0$, $np^{r-1} \to c \in (0,\infty)$, and $np^{r-1} \to \infty$ in a unified manner, it will be convenient to extend φ_r to a continuous function $\varphi_r \colon (0,\infty) \times [0,\infty] \to (0,\infty)$.

This extension may be defined by noting that

$$\lim_{c\to 0}\varphi_r(\delta,c)=\delta^{2/r}/2 \qquad \text{and} \qquad \lim_{c\to\infty}\varphi_r(\delta,c)=\min\{\delta^{2/r}/2,\delta/r\}$$

uniformly as functions of δ. For every $\delta>0$ and $c\in[0,\infty]$, we then define the set

$$\bar{X}_r(\delta,c)=\{x\in[0,1]:\varphi_r(\delta,c)=\lim_{c'\to c}\psi_r(\delta,c',x)\}$$

of (asymptotic) minimisers to $x\mapsto\psi_r(\delta,c,x)$. One can check that this set is nonempty for any δ and c, though it might contain more than one element. The following lemma describes the set of possible minimisers in more detail.

Lemma 6.1. *Let $r\geqslant 3$ be an integer and let δ and c be positive real numbers. Let $x^*=r\lfloor\delta c/r\rfloor/(\delta c)$. Then*

$$\bar{X}_r(\delta,0)=\{0\},\quad \bar{X}_r(\delta,c)\subseteq\{0,x^*,1\},\quad \textit{and}\quad \bar{X}_r(\delta,\infty)\subseteq\{0,1\}.$$

Proof. The statement for $\bar{X}_r(\delta,0)$ holds because $\lim_{c\to0}\varphi_r(\delta,c,x)=\infty$ whenever $x>0$. The statement for $\bar{X}_r(\delta,\infty)$ follows because

$$\lim_{c\to\infty}\psi_r(\delta,c,x)=(\delta(1-x))^{2/r}/2+x\delta/r$$

and the right-hand side is strictly concave in $x\in[0,1]$. In particular, since

$$\varphi_r(\delta,\infty)=\min\{\delta^{2/r},\delta/r\}=\min_{x\in[0,1]}(\delta(1-x))^{2/r}/2+x\delta/r,$$

then $\varphi_r(\delta,\infty)=\lim_{c\to\infty}\psi_r(\delta,c,x)$ only if x is one of the endpoints of the interval $[0,1]$. Assume now that $c\in(0,\infty)$. Treating r, δ, and c as fixed, consider the function $f\colon[0,1]\to(0,\infty)$ defined by $f(x)=\psi_r(\delta,c,x)$. Since ψ_r is continuous, we have $x\in\bar{X}_r(\delta,c)$ if and only if x is a minimiser of $f(x)$ in $[0,1]$. We cover the domain of f with essentially disjoint intervals as follows:

$$[0,1]=\frac{r}{\delta c}[0,1]\cup\frac{r}{\delta c}[1,2]\cup\cdots\cup\frac{r}{\delta c}[\lfloor\delta c/r\rfloor-1,\lfloor\delta c/r\rfloor]\cup\frac{r}{\delta c}[\lfloor\delta c/r\rfloor,\delta c/r].$$

Observe that f is strictly concave on each of these intervals and thus f can achieve its minimum value only at an endpoint of one of the intervals, that is, either at some x for which $\delta xc/r\in\mathbb{Z}$ or at $x=1$. Define the function $h\colon[0,1]\to(0,\infty)$ by

$$h(x)=\frac{\big(\delta(1-x)\big)^{2/r}}{2}+\frac{x\delta}{r}$$

and observe that h is strictly concave and that $h(x)=f(x)$ whenever $\delta xc/r\in\mathbb{Z}$. It follows that the minimum of $f(x)$ is achieved either at $x=1$ or at the smallest or largest x satisfying $\delta xc/r\in\mathbb{Z}$. Since the latter two points are $x=0$ and $x=r\lfloor\delta c/r\rfloor/(\delta c)=x^*$, respectively, we may conclude that the minimiser is in the set $\{0,x^*,1\}$, as desired. $\square$

Let us now state the two main results of this section. The following is a straightforward reformulation of Theorem 1.7.

Theorem 6.2. *Let $r\geqslant3$ be an integer and let $X=X_{n,p}^{K_r}$ denote the number of r-vertex cliques in the random graph $G_{n,p}$. Then, for every fixed positive constant δ and all $p=p(n)$ such that $n^{-1}(\log n)^{\frac{1}{r-2}}\ll p^{\frac{r-1}{2}}\ll 1$ and $\lim_{n\to\infty}np^{r-1}=c\in[0,\infty]$,*

$$\lim_{n\to\infty}\frac{-\log\mathbb{P}\big(X\geqslant(1+\delta)\,\mathbb{E}[X]\big)}{n^2p^{r-1}\log(1/p)}=\varphi_r(\delta,c).$$

Recall the following three events:

(i) Let $\mathrm{UT}(\delta)$ be the upper tail event $\{X\geqslant(1+\delta)\,\mathbb{E}[X]\}$.
(ii) Let $\mathrm{Clique}_\varepsilon(x)$ be the event that $G_{n,p}$ contains a set $U\subseteq V(G)$ of size at least $(1-\varepsilon)x^{1/r}np^{(r-1)/2}$ such that $G[U]$ has minimum degree at least $(1-\varepsilon)|U|$.

(iii) Let $\mathrm{Hub}_\varepsilon(x)$ be the event that $G_{n,p}$ contains a set $U \subseteq V(G)$ such that at least $\lfloor (1-\varepsilon)|U| \rfloor$ vertices in U have degree at least $(1-\varepsilon)n$ and

$$e(U, V(G) \setminus U) \geqslant (1-\varepsilon)n\big(\lfloor xnp^{r-1}/r \rfloor + \{xnp^{r-1}/r\}^{1/(r-1)}\big).$$

Together with Lemma 6.1, the following theorem directly implies Theorem 1.8.

Theorem 6.3. *Let $r \geqslant 3$ be an integer and let $X = X_{n,p}^{K_r}$ denote the number of r-vertex cliques in the random graph $G_{n,p}$. Then, for every fixed positive constant δ and all $p = p(n)$ such that $n^{-1}(\log n)^{\frac{1}{r-2}} \ll p^{\frac{r-1}{2}} \ll 1$ and $\lim_{n\to\infty} np^{r-1} = c \in [0,\infty]$,*

$$\lim_{n\to\infty} \mathbb{P}\Big(\bigcup_{x \in \bar{X}(\delta,c)} \mathrm{Clique}_\varepsilon(\delta(1-x)) \cap \mathrm{Hub}_\varepsilon(\delta x) \mid \mathrm{UT}(\delta) \Big) = 1.$$

In order to prove Theorems 6.2 and 6.3, we will first relate $-\log \mathbb{P}\big(X \geqslant (1+\delta)\,\mathbb{E}[X]\big)$ to the solutions of the optimisation problem

$$\Phi_X(\delta) = \min\big\{ e_G \log(1/p) : G \subseteq K_n \text{ and } \mathbb{E}_G[X] \geqslant (1+\delta)\,\mathbb{E}[X] \big\},$$

where $\mathbb{E}_G[X] = \mathbb{E}[X \mid G \subseteq G_{n,p}]$. For every $\varepsilon > 0$, we define the event

$$\begin{aligned}\text{Near-min}(\varepsilon) = \{G_{n,p} \text{ contains a subgraph } G \text{ such that } e_G \leqslant (1+\varepsilon)\Phi_X(\delta+\varepsilon) \text{ and}\\ \mathbb{E}_G[X] \geqslant (1+\delta-\varepsilon)\,\mathbb{E}[X]\}.\end{aligned}$$

Using Theorem 3.1, we shall prove the following result.

Proposition 6.4. *For every integer $r \geqslant 3$ and all positive reals ε and δ, there exists a positive constant C such that the following holds. Suppose that an integer n and $p \in (0,1)$ satisfy $Cn^{-1}(\log n)^{1/(r-2)} \leqslant p^{(r-1)/2} \leqslant 1/C$. Then $X = X_{n,p}^{K_r}$ satisfies*

$$(1-\varepsilon)\Phi_X(\delta-\varepsilon) \leqslant -\log \mathbb{P}\big(X \geqslant (1+\delta)\,\mathbb{E}[X]\big) \leqslant (1+\varepsilon)\Phi_X(\delta+\varepsilon).$$

Furthermore,

$$\mathbb{P}\big(\text{Near-min}(\varepsilon) \mid X \geqslant (1+\delta)\,\mathbb{E}[X]\big) \geqslant 1-\varepsilon.$$

In order to complete the proof of Theorem 6.2, we must also evaluate the asymptotic value of the function Φ_X. This is the content of our second proposition.

Proposition 6.5. *Let $r \geqslant 3$ be an integer and let $X = X_{n,p}^{K_r}$. Then, for every fixed positive real δ and all $p = p(n)$ such that $n^{-1} \ll p^{(r-1)/2} \ll 1$ and $\lim_{n\to\infty} np^{r-1} = c \in [0,\infty]$,*

$$\lim_{n\to\infty} \frac{\Phi_X(\delta)}{n^2 p^{r-1} \log(1/p)} = \varphi_r(\delta, c).$$

Finally, to prove Theorem 6.3, we characterise the near-minimisers of the optimisation problem for $\Phi_X(\delta)$. This is what we do in the final proposition of this section.

Proposition 6.6. *Let $r \geqslant 3$ be an integer and let $X = X_{n,p}^{K_r}$. For all fixed $\varepsilon, \delta > 0$ and $c \in [0,\infty]$, there exists some positive constant η such that the following holds. Assume $p = p(n)$ is such that $n^{-1} \ll p^{(r-1)/2} \ll 1$ and $\lim_{n\to\infty} np^{r-1} = c$. Then*

$$\text{Near-min}(\eta) \subseteq \bigcup_{x \in \bar{X}(\delta,c)} \mathrm{Clique}_\varepsilon(\delta(1-x)) \cap \mathrm{Hub}_\varepsilon(\delta x)$$

whenever n is sufficiently large.

The propositions above readily imply Theorems 6.2 and 6.3.

6.1. **Proof of Proposition 6.4.** We may assume without loss of generality that ε is sufficiently small, say $\varepsilon < \min\{1/2, \delta/2\}$. Note also that the case $n < r$ is trivial; indeed, in that case X is identically zero and thus $-\log \mathbb{P}\big(X \geqslant (1+\delta)\,\mathbb{E}[X]\big) = 0 = \Phi_X(\delta)$ for every $\delta \in \mathbb{R}$, and Near-min(ε) holds vacuously. We may therefore assume that $n \geqslant r \geqslant 3$, which, in turn, implies that $n \geqslant C$.

Set $N = \binom{n}{2}$ and let $Y = (Y_1, \dots, Y_N)$ be the sequence of indicator random variables of the events that $e \in E(G_{n,p})$, where e ranges over $\binom{\llbracket n \rrbracket}{2}$ in some arbitrary order. Observe that Y is a vector of independent $\mathrm{Ber}(p)$ random variables and that X is a nonzero polynomial with nonnegative coefficients and total degree at most $\binom{r}{2}$ in the coordinates of Y. Let $K = K(r, \varepsilon, \delta)$ be the constant given by Theorem 3.1. We shall show that X satisfies the various assumptions of the theorem; the theorem then implies both assertions of the proposition.

First, our assumption on p implies that $p \leqslant 1 - \varepsilon$, provided that C is sufficiently large.

Recall that $N(J,G)$ denotes the number of copies of J in G. Note that for all $J \subseteq K_r$ without isolated vertices and all $G \subseteq K_n$, we can trivially bound $N(J,G)$ from above by $e_G^{e_J}$. It follows that

$$\begin{aligned}\mathbb{E}_G[X] - \mathbb{E}[X] &\leqslant \sum_{\varnothing \neq J \subseteq K_r} N(J,G) \cdot n^{r-v_J} p^{\binom{r}{2}-e_J} \leqslant 2^{\binom{r}{2}} \cdot \max_{\varnothing \neq J \subseteq K_r} e_G^{e_J} \cdot n^{r-v_J} p^{\binom{r}{2}-\binom{v_J}{2}} \\ &\leqslant (2e_G)^{\binom{r}{2}} \cdot \frac{n^r p^{\binom{r}{2}}}{\min_{2 \leqslant k \leqslant r} n^k p^{\binom{k}{2}}},\end{aligned}$$

where the sum ranges over all nonempty subgraphs $J \subseteq K_r$ without isolated vertices. Since $\mathbb{E}[X] = \Theta\big(n^r p^{\binom{r}{2}}\big)$ and our assumption on p implies that $n^k p^{\binom{k}{2}} \geqslant C^{2/(r-1)}$ for each $k \in \{2, \dots, r\}$, then the right-hand side above is at most $(\delta/2)\,\mathbb{E}[X]$ unless $e_G \geqslant K$, provided that C is sufficiently large. Therefore, $\Phi_X(\delta - \varepsilon) \geqslant \Phi_X(\delta/2) \geqslant K \log(1/p)$. Furthermore, since a clique with $\lceil (1+2\delta)^{1/r} n p^{(r-1)/2} \rceil$ vertices contains at least $(1+\delta+\varepsilon)\,\mathbb{E}[X]$ copies of K_r and has fewer than $K' n^2 p^{r-1}$ edges, for some constant $K' = K'(\delta)$, we deduce that $\Phi_X(\delta+\varepsilon) \leqslant K' n^2 p^{r-1} \log(1/p)$.

Recall that a graph $G^* \subseteq K_n$ is a *core* if it satisfies the following three conditions:

(C1) $\mathbb{E}_{G^*}[X] \geqslant (1+\delta-\varepsilon)\,\mathbb{E}[X]$,
(C2) $e_{G^*} \leqslant K \cdot \Phi_X(\delta+\varepsilon)$, and
(C3) $\min_{e \in E(G^*)} \big(\mathbb{E}_{G^*}[X] - \mathbb{E}_{G^* \setminus e}[X]\big) \geqslant \mathbb{E}[X] / \big(K \cdot \Phi_X(\delta+\varepsilon)\big)$.

Our goal is to show that, for every integer m, there are at most $(1/p)^{\varepsilon m/2}$ cores with m edges. The key observation is that, for any core G^* and any edge uv of G^*, either u and v have many common neighbours or the sum of the degrees of u and v is large. More precisely, letting $\deg_{G^*}(u,v)$ denote the number of common neighbours of u and v in G^*, we shall establish the following statement.

Claim 6.7. *There exists a positive constant $\eta = \eta(\delta, r, K)$ such that, for every core G^* and each edge $uv \in E(G^*)$, either*

$$\deg_{G^*}(u,v) \geqslant \frac{\eta n p^{(r-1)/2}}{\big(\log(1/p)\big)^{1/(r-2)}} \qquad \text{or} \qquad \deg_{G^*} u + \deg_{G^*} v \geqslant \frac{\eta n}{\log(1/p)}.$$

The above claim readily implies the desired bound on the number of cores with m edges. Note first that this number is zero whenever $m > KK'n^2p^{r-1}\log(1/p)$, see (C2), so we may assume that $m \leqslant KK'n^2p^{r-1}\log(1/p)$. Given a core G^*, we denote by A_{G^*} the set of vertices of G^* with degree at least $\eta n p^{(r-1)/2} / \big(\log(1/p)\big)^{1/(r-2)}$ and by $B_{G^*} \subseteq A_{G^*}$ the set of vertices of G^* with degree at least $\eta n / \log(1/p)$. Since G^* has m edges, then

$$|A_{G^*}| \leqslant a := \frac{2m\big(\log(1/p)\big)^{1/(r-2)}}{\eta n p^{(r-1)/2}} \quad \text{and} \quad |B_{G^*}| \leqslant b := \frac{2m \log(1/p)}{\eta n}$$

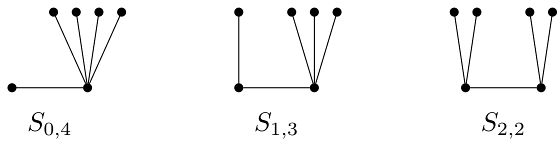

FIGURE 5. The double stars with six vertices.

Claim 6.7 states that every edge of G^* is either fully contained in A_{G^*} or has at least one endpoint in B_{G^*}. In particular, for fixed sets $B \subseteq A \subseteq [\![n]\!]$ with $|A| = a$ and $|B| = b$, there are at most $\binom{a^2/2+bn}{m}$ cores G^* with m edges that satisfy $A_{G^*} \subseteq A$ and $B_{G^*} \subseteq B$. We can thus (generously) upper bound the number of cores with m edges by

$$\binom{n}{a}\binom{n}{b}\binom{a^2/2+bn}{m}.$$

Recalling the inequality $\binom{x}{y} \leqslant (ex/y)^y$, valid for all nonnegative integers x and y, we may conclude that the number of cores with m edges is at most

$$n^{\frac{4m(\log(1/p))^{1/(r-2)}}{\eta n p^{(r-1)/2}}} \cdot \left(\frac{2em\big(\log(1/p)\big)^{2/(r-2)}}{\eta^2 n^2 p^{r-1}} + \frac{2e\log(1/p)}{\eta} \right)^m .$$

Since $p^{(r-1)/2} \geqslant Cn^{-1}(\log n)^{1/(r-2)}$, the first factor is at most $e^{\varepsilon m \log(1/p)/4}$. Since we have assumed that $m \leqslant KK'n^2p^{r-1}\log(1/p)$, the second factor is at most $e^{O(m\log\log(1/p))} \leqslant e^{\varepsilon m \log(1/p)/4}$, since $1/p \geqslant C^{2/(r-1)}$. This shows that the number of cores with m edges is indeed at most $(1/p)^{\varepsilon m/2}$, as claimed.

Proof of Claim 6.7. For any nonempty $J \subseteq K_r$, we shall let $N(J, G^*; uv)$ denote the number of copies of J in G^* that contain the edge uv. For the sake of brevity, we set $m_{\max} = KK' \cdot n^2p^{r-1}\log(1/p)$. Observe that

$$\mathbb{E}_{G^*}[X] - \mathbb{E}_{G^* \setminus uv}[X] \leqslant \sum_{\varnothing \neq J \subseteq K_r} N(J, G^*; uv) \cdot n^{r - v_J} p^{\binom{r}{2} - e_J},$$

where J ranges over all nonempty subgraphs of K_r that have no isolated vertices. Since $\mathbb{E}[X] = \binom{n}{r}p^{\binom{r}{2}}$ and $\Phi_X(\delta+\varepsilon) \leqslant K'n^2p^{r-1}\log(1/p) = m_{\max}/K$, it follows from (C3) in the definition of the core that

$$\sum_{\varnothing \neq J \subseteq K_r} \frac{N(J, G^*; uv)}{n^{v_J}p^{e_J}} \geqslant \frac{\mathbb{E}[X]}{K \cdot \Phi_X(\delta+\varepsilon)} \geqslant \frac{\gamma}{m_{\max}},$$

where $\gamma = \gamma(r)$ is a constant that depends only on r.

For every edge $ab \in E(J)$, let $\mathrm{Emb}(J, G^*; ab, uv)$ denote the set of embeddings of J into G^* that map ab to uv. Then the above inequality implies that there is a nonempty $J \subseteq K_r$ with no isolated vertices, an edge $ab \in E(J)$, and a constant $\gamma' = \gamma'(r)$ such that

$$\frac{|\,\mathrm{Emb}(J, G^*; ab, uv)|}{n^{v_J}p^{e_J}} \geqslant \frac{\gamma'}{m_{\max}}. \tag{29}$$

For nonnegative integers i and j, let $S_{i,j}$ denote the graph obtained from a copy of $K_{1,i}$ and a copy of $K_{1,j}$ by joining their centres (vertices of degrees i and j, respectively) by an edge; see Figure 5 for an illustration. As the graphs $K_{1,i}$ are often called *stars*, we shall refer to the $S_{i,j}$ as *double stars*. Moreover, we shall call an edge of $S_{i,j}$ whose endpoints have degrees $i+1$ and $j+1$ a *centre edge*. Note that if $i, j > 0$, then $S_{i,j}$ has only one centre edge; otherwise, $S_{i,j}$ is just a star graph and each of its edges is a centre edge.

We shall now show that, unless $\deg_{G^*}(u,v) \geqslant np^{(r-1)/2}$, any graph J that satisfies (29) for some $ab \in E(J)$ must be either K_r or a double star with r vertices. We first show how this fact implies the assertion of the claim. Assume first that

$$|\operatorname{Emb}(K_r, G^*; ab, uv)| \geqslant \frac{\gamma' n^r p^{\binom{r}{2}}}{m_{\max}} = \frac{\gamma'}{KK'} \cdot \frac{n^{r-2} p^{\binom{r-1}{2}}}{\log(1/p)}.$$

Since $|\operatorname{Emb}(K_r, G^*; ab, uv)| \leqslant \deg_{G^*}(u,v)^{r-2}$, we conclude that

$$\deg_{G^*}(u,v) \geqslant \left(\frac{\gamma'}{KK'}\right)^{1/(r-2)} \cdot \frac{np^{(r-1)/2}}{\big(\log(1/p)\big)^{1/(r-2)}},$$

as claimed. Next, assume that, for some i and j with $i + j = r - 2$,

$$|\operatorname{Emb}(S_{i,j}, G^*; ab, uv)| \geqslant \frac{\gamma' n^r p^{i+j+1}}{m_{\max}} = \frac{\gamma'}{KK'} \cdot \frac{n^{r-2}}{\log(1/p)}.$$

If ab is a centre edge of $S_{i,j}$, then $|\operatorname{Emb}(S_{i,j}, G; ab, uv)| \leqslant (\deg_{G^*} u + \deg_{G^*} v)^{i+j} \leqslant (\deg_{G^*} u + \deg_{G^*} v) \cdot (2n)^{r-3}$. Otherwise, we have $i, j > 0$ and so $|\operatorname{Emb}(S_{i,j}, G; ab, uv)| \leqslant (\deg_{G^*} u + \deg_{G^*} v)^{\min\{i,j\}} \cdot n^{\max\{i,j\}} \leqslant (\deg_{G^*} u + \deg_{G^*} v) \cdot (2n)^{r-3}$ as well. Thus, in both cases,

$$\deg_{G^*} u + \deg_{G^*} v \geqslant \frac{\gamma'}{2^{r-3} KK'} \cdot \frac{n}{\log(1/p)},$$

which completes the proof of the claim.

It remains to prove our assertion. We first consider the special case $r = 3$. The only nonempty subgraph of K_3 with no isolated vertices that is not isomorphic to a double star with three vertices is K_2. However, as $p \leqslant C^{-2/(r-1)} = C^{-1}$, we have

$$\frac{|\operatorname{Emb}(K_2, G^*; ab, uv)|}{n^2 p} = \frac{2}{n^2 p} \leqslant \frac{2 \log C}{C n^2 p^2 \log(1/p)} < \frac{\gamma'}{m_{\max}}$$

whenever C is sufficiently large, which contradicts (29). We henceforth assume that $r \geqslant 4$.

By way of contradiction, suppose that J is neither K_r nor a double star with r vertices and that $\deg_{G^*}(u,v) < np^{(r-1)/2}$. Let ab be an arbitrary edge of J, let J_{ab} be the subgraph of J induced by $V(J) \setminus \{a,b\}$, and let α^*_{ab} be the fractional independence number of J_{ab}. By Lemma 5.2, there is a partition of $V(J_{ab})$ into V_1 and V_2 such that

- (P1) $|V_1|/2 + |V_2| = \alpha^*_{ab}$,
- (P2) V_1 can be covered by a collection of vertex-disjoint edges and cycles of J_{ab}.

Among all partitions satisfying (P1) and (P2), choose one that maximises the number of common neighbours of a and b in V_2, that is, the cardinality of the set

$$X = \{c \in V_2 : ac, bc \in E(J)\}.$$

Finally, let $v_1 = |V_1|$, $v_2 = |V_2|$, and $x = |X|$.

We now observe that for every partition satisfying (P1) and (P2), we have

$$|\operatorname{Emb}(J, G^*; ab, uv)| \leqslant 2 \cdot \deg_{G^*}(u,v)^x \cdot (2e_{G^*})^{v_1/2} \cdot \min\{2e_{G^*}, n\}^{v_2 - x}. \tag{30}$$

To see this, note first that there are two embeddings of ab onto uv. Since X lies in the common neighbourhood of a and b, each such embedding can be extended to an embedding of $J[\{a,b\} \cup X]$ in at most $\deg_{G^*}(u,v)^x$ ways. Next, since V_1 can be covered by cycles and edges of J (by property (P2)), Lemma 5.5 implies that every embedding of $J[\{a,b\} \cup X]$ can be extended in at most $(2e_{G^*})^{v_1/2}$ ways to an embedding of $J[\{a,b\} \cup X \cup V_1]$. Finally, since J contains no isolated vertices, any embedding of $J[\{a,b\} \cup X \cup V_1]$ can be extended in at most $\min\{2e_{G^*}, n\}^{v_2 - x}$ ways to an embedding of J.

Combining (29), (30), the inequality $e_{G^*} \leqslant m_{\max}$, which follows from (C2), and the assumption $\deg_{G^*}(u,v) < np^{(r-1)/2}$, we deduce that

$$\begin{aligned}\gamma' n^{v_J} p^{e_J} &\leqslant \left(np^{(r-1)/2}\right)^x \cdot (2m_{\max})^{v_1/2+1} \cdot \min\{2m_{\max}, n\}^{v_2-x} \\ &= \frac{1}{\left(KK'\log(1/p)\right)^{x/2}} \cdot (2m_{\max})^{v_1/2+x/2+1} \cdot \min\{2m_{\max}, n\}^{v_2-x}.\end{aligned}$$

On the other hand,

$$n^{v_J} p^{e_J} = \left(n^2 p^{r-1}\right)^{e_J/(r-1)} \cdot n^{v_J - 2e_J/(r-1)} = \left(\frac{m_{\max}}{KK'\log(1/p)}\right)^{e_J/(r-1)} \cdot n^{v_J-2e_J/(r-1)}.$$

Therefore, for some constant $K'' = K''(K, K', r)$, we must have

$$(m_{\max})^{v_1/2+x/2-e_J/(r-1)+1} \cdot n^{2e_J/(r-1)-v_J} \cdot \min\{m_{\max}, n\}^{v_2-x} \geqslant \left(K''\log(1/p)\right)^{-K''}. \tag{31}$$

In order to reach the desired contradiction, it suffices to prove that there is some positive $\sigma = \sigma(r)$ such that the left-hand side of (31) is bounded from above by $\max\{n^{-1}, KK' \cdot p^{r-1}\log(1/p)\}^{\sigma}$. We first observe that such an upper bound is implied by the following inequality:

$$\max\left\{v_1/2 + v_2 - x/2 - e_J/(r-1) + 1,\, 0\right\} < v_J - 2e_J/(r-1). \tag{32}$$

Indeed, if $m_{\max} \leqslant n$, then the left-hand side of (31) is upper bounded by

$$\begin{aligned}&(m_{\max})^{v_1/2+v_2-x/2-e_J/(r-1)+1} \cdot n^{2e_J/(r-1)-v_J} \\ &\qquad\leqslant \max\left\{n^{2e_J/(r-1)-v_J}, n^{v_1/2+v_2-x/2+1-(v_J-e_J/(r-1))}\right\}.\end{aligned}$$

On the other hand, if $n < m_{\max} = KK' \cdot n^2 p^{r-1}\log(1/p)$, then the left-hand side of (31) becomes

$$\begin{aligned}&\left(KK' \cdot n^2 p^{r-1}\log(1/p)\right)^{v_1/2+x/2-e_J/(r-1)+1} \cdot n^{2e_J/(r-1)-v_J+v_2-x} \\ &\qquad= \left(KK' \cdot p^{r-1}\log(1/p)\right)^{v_J-e_J/(r-1)-(v_1/2+v_2-x/2+1)},\end{aligned}$$

using $v_1 + v_2 = v_{J_{ab}} = v_J - 2$. Note that (32) guarantees that, in both cases, the left-hand side of (31) is at most $\max\{n^{-1}, KK' \cdot p^{r-1}\log(1/p)\}^{\sigma}$, for a suitable positive $\sigma = \sigma(r)$.

In order to complete the proof of Claim 6.7, we now prove inequality (32). Recall that $V(J_{ab}) = V_1 \cup V_2$ is a partition that satisfies (P1) and (P2) that maximises the cardinality of the set $X = \{c \in V_2 : ac, bc \in E(J)\}$. Since, by the definition of X, each vertex in $V_2 \setminus X$ has at most one neighbour in $\{a, b\}$,

$$e_J \leqslant e_{J_{ab}} + 2v_1 + v_2 + x + 1 = e_{J_{ab}} + 3v_J - 2\alpha^*_{ab} + x - 5 \tag{33}$$

and equality holds only if every vertex in V_1 is adjacent to both a and b and $e(V_2, \{a,b\}) = v_2$. Moreover, Lemma 5.3 and the remark that follows it give

$$e_{J_{ab}} \leqslant (v_{J_{ab}} - 1)(v_{J_{ab}} - \alpha^*_{ab}) = (v_J - 3)(v_J - \alpha^*_{ab} - 2), \tag{34}$$

where equality holds only if J_{ab} is complete, empty, or isomorphic to K_{1,v_J-3}. Putting (33) and (34) together yields the inequality

$$e_J \leqslant (v_J - 1)(v_J - \alpha^*_{ab} - 1) + x. \tag{35}$$

Moreover, inequality (35) is strict unless both (33) and (34) hold with equality. Rearranging (35) gives the inequality

$$\frac{e_J}{r-1} \leqslant \frac{v_J - 1}{r-1} \cdot (v_J - \alpha^*_{ab} - 1) + \frac{x}{r-1} \leqslant v_J - \alpha^*_{ab} - 1 + \frac{x}{2}. \tag{36}$$

Since $\alpha^*_{ab} \leqslant v_{J_{ab}} = v_J - 2$ and $r \geqslant 4$, the second inequality in (36) is strict unless $v_J = r$ and $x = 0$. Consequently, the left-hand and the right-hand sides of (36) can be equal only if the following conditions are satisfied for every partition $V(J_{ab}) = V_1 \cup V_2$ with properties (P1) and (P2):

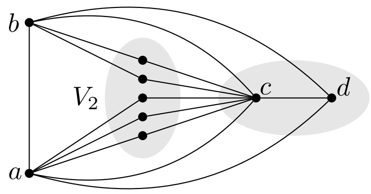

FIGURE 6. Illustration for the case $J_{ab} = K_{1,r-3}$ and $r > 4$. Any partition $V(J_{ab}) = V_1' \cup V_2'$ obtained by exchanging d with a vertex from V_2 satisfies (P1) and (P2) but violates (D3).

(D1) J_{ab} is either K_{r-2}, E_{r-2} (the empty graph with $r-2$ vertices), or $K_{1,r-3}$;
(D2) every vertex in V_1 is adjacent to both a and b;
(D3) every vertex of V_2 is adjacent to exactly one of a and b.

We now show that our assumptions preclude (D1)–(D3) holding simultaneously. Indeed, note first that, if $J_{ab} = K_{r-2}$ (which also includes the case $J_{ab} = K_{1,r-3}$ and $r = 4$), then $\alpha^*_{ab} = v_{J_{ab}}/2$. Since $v_1/2 + v_2 = \alpha^*_{ab}$ and $v_1 + v_2 = v_{J_{ab}}$, this implies $V_1 = V(J_{ab})$. Then (D2) shows that $J = K_r$, a contradiction. Second, if $J_{ab} = E_{r-2}$, then $\alpha^*_{ab} = r-2$. Since $v_1/2 + v_2 = \alpha^*_{ab}$ and $v_1 + v_2 = r-2$, this implies $V_2 = V(J_{ab})$ and it follows from (D3) that J is a double star whose centre edge is ab, another contradiction. Finally, suppose that $J_{ab} = K_{1,r-3}$ and $r > 4$. Since $\alpha^*_{ab} = v_1/2 + v_2 = r-3$ and $v_1 + v_2 = r-2$, we see that $v_1 = 2$ and $v_2 = r-4$. Property (P2) implies that $V_1 = \{c, d\}$, where c is the vertex of degree $r-3$ in J_{ab} and d is one of its neighbours. Since $d \in V_1$, it must be adjacent to both a and b (see Figure 6 for an illustration). Let e be an arbitrary vertex of V_2; there is at least one such vertex as $v_2 = r-4 \geqslant 1$. The partition of $V(J_{ab})$ into $V_1' = \{c, e\}$ and $V_2' = (V_2 \setminus \{e\}) \cup \{d\}$ satisfies both conditions (P1) and (P2) but the set $X' = \{c' \in V_2' : ac', vc' \in E(J_{ab})\}$ is nonempty (as it contains the vertex d); this contradicts property (D3) for the partition $V(J) = V_1' \cup V_2'$.

To summarise, at least one of the inequalities in (36) is strict. Since $J \neq K_r$, then not every vertex of J has degree $r-1$ and thus $2e_J < (r-1) \cdot v_J$. We conclude that

$$\max\left\{\alpha^*_{ab} - e_J/(r-1) - x/2 + 1, \, 0\right\} < v_J - 2e_J/(r-1).$$

Since $\alpha^*_{ab} = v_1/2 + v_2$, this is exactly (32). □

6.2. **Proof of Proposition 6.5.** We begin by showing that

$$\limsup_{n\to\infty} \frac{\Phi_X(\delta)}{n^2 p^{r-1} \log(1/p)} \leqslant \varphi_r(\delta, c). \tag{37}$$

For every small $\varepsilon > 0$ and sufficiently large n, we shall construct a graph G with vertex set $[\![n]\!]$ and at most $(\varphi_r(\delta+\varepsilon, c) + \varepsilon) \cdot n^2 p^{r-1}$ edges that satisfies $\mathbb{E}_G[X] \geqslant (1+\delta)\,\mathbb{E}[X]$. The existence of such a graph and the continuity of φ_r will imply that

$$\limsup_{n\to\infty} \frac{\Phi_X(\delta)}{n^2 p^{r-1} \log(1/p)} \leqslant \lim_{\varepsilon\to 0} \varphi_r(\delta+\varepsilon, c) + \varepsilon = \varphi_r(\delta, c),$$

as required.

Let $x \in \bar{X}(\delta+\varepsilon, c)$ and note that it follows from Lemma 6.1 that $x = 0$ if $c = 0$. Set

$$\ell_1 = \big((\delta+\varepsilon)(1-x)\big)^{1/r} np^{(r-1)/2} \qquad \text{and} \qquad \ell_2 = x(\delta+\varepsilon)np^{r-1}/r.$$

and fix an arbitrary partition $[\![n]\!] = \{u\} \cup U_1 \cup U_2 \cup U_3$, where $|U_1| = \lfloor \ell_1 \rfloor$ and $|U_2| = \lfloor \ell_2 \rfloor$. Let G be the union of the clique on U_1, the complete bipartite graph between U_2 and U_3, and an arbitrary star, centred at u, with $\lfloor\{\ell_2\}^{1/(r-1)}|U_3|\rfloor$ edges whose non-u endpoints are in U_3. We have

$$\begin{aligned}
e_G &\leqslant \binom{\ell_1}{2} + \lfloor\ell_2\rfloor\cdot|U_3| + \{\ell_2\}^{1/(r-1)}\cdot|U_3| \\
&\leqslant \frac{\big((\delta+\varepsilon)(1-x)\big)^{2/r}n^2p^{r-1}}{2} + \big(\lfloor\ell_2\rfloor + \{\ell_2\}^{1/(r-1)}\big)\cdot n \\
&\leqslant \left(\frac{\big((\delta+\varepsilon)(1-x)\big)^{2/r}}{2} + \frac{\lfloor\ell_2\rfloor+\{\ell_2\}^{1/(r-1)}}{np^{r-1}}\right)\cdot n^2p^{r-1} \\
&= \varphi_r(\delta+\varepsilon, np^{r-1}, x)\cdot n^2p^{r-1},
\end{aligned}$$

and our choice of x ensures that $\varphi_r(\delta+\varepsilon, np^{r-1}, x) \leqslant \varphi_r(\delta+\varepsilon, c)+\varepsilon$ for all large enough n.

It remains to show that $\mathbb{E}_G[X] \geqslant (1+\delta)\,\mathbb{E}[X]$. To this end, observe first that:

(i) The complete graph $G[U_1]$ contains $\binom{\lfloor\ell_1\rfloor}{r}$ copies of K_r.
(ii) The complete bipartite graph $G[U_2, U_3]$ contains $\lfloor\ell_2\rfloor\cdot\binom{|U_3|}{r-1}$ copies of $K_{1,r-1}$ whose centre vertex lies in U_2.
(iii) The star $G[u, U_3]$ contains $\binom{\lfloor\{\ell_2\}^{1/(r-1)}|U_3|\rfloor}{r-1}$ copies of $K_{1,r-1}$.

In particular,

$$\begin{aligned}
\mathbb{E}_G[X]-\mathbb{E}[X] &= \sum_{\varnothing\neq J\subseteq K_r} N(J,G)\cdot\binom{n-v_J}{r-v_J}\cdot p^{\binom{r}{2}}(p^{-e_J}-1) \\
&\geqslant (1-p)\cdot\left[\binom{\lfloor\ell_1\rfloor}{r} + \left[\lfloor\ell_2\rfloor\cdot\binom{|U_3|}{r-1} + \binom{\lfloor\{\ell_2\}^{1/(r-1)}|U_3|\rfloor}{r-1}\right]\cdot p^{\binom{r}{2}-(r-1)}\right].
\end{aligned}$$

We now estimate the right-hand side of the above inequality. Since $np^{(r-1)/2}\to\infty$, then

$$\binom{\lfloor\ell_1\rfloor}{r} \geqslant \frac{(\ell_1-r)^r}{r!} \geqslant \big((\delta+\varepsilon)(1-x)-\varepsilon/2\big)\cdot\frac{n^rp^{\binom{r}{2}}}{r!},$$

provided that n is sufficiently large. In the case $c=0$, this is already sufficient, as $x=0$ and

$$\mathbb{E}_G[X]-\mathbb{E}[X] \geqslant (1-p)\cdot\binom{\lfloor\ell_1\rfloor}{r} \geqslant (1-p)\cdot(\delta+\varepsilon/2)\cdot\frac{n^rp^{\binom{r}{2}}}{r!} \geqslant \delta\,\mathbb{E}[X].$$

We may therefore assume that $c\in(0,\infty]$. Since $p\to 0$, and thus $|U_3|/n\to 1$, then

$$\begin{aligned}
\lfloor\ell_2\rfloor\cdot\binom{|U_3|}{r-1} + \binom{\lfloor\{\ell_2\}^{1/(r-1)}|U_3|\rfloor}{r-1} &\geqslant \lfloor\ell_2\rfloor\cdot\frac{(1-\varepsilon^2)\cdot n^{r-1}}{(r-1)!} + \frac{\big(\{\ell_2\}^{1/(r-1)}|U_3|-r\big)^{r-1}}{(r-1)!} \\
&\geqslant \frac{(1-\varepsilon^2)\cdot\ell_2\cdot n^{r-1}}{(r-1)!} - \frac{\varepsilon^2n^{r-1}}{(r-1)!} \\
&= \left((1-\varepsilon^2)x(\delta+\varepsilon) - \frac{\varepsilon^2 r}{np^{r-1}}\right)\cdot\frac{n^rp^{r-1}}{r!}.
\end{aligned}$$

Consequently,

$$\begin{aligned}
\mathbb{E}_G[X]-\mathbb{E}[X] &\geqslant (1-p)\cdot\left((\delta+\varepsilon)(1-x)-\varepsilon/2+(1-\varepsilon^2)x(\delta+\varepsilon)-\frac{\varepsilon^2r}{np^{r-1}}\right)\cdot\frac{n^rp^{\binom{r}{2}}}{r!} \\
&\geqslant (1-p)\cdot\left(\delta+\varepsilon/2-\varepsilon^2(\delta+\varepsilon)-\frac{\varepsilon^2r}{np^{r-1}}\right)\cdot\mathbb{E}[X] \geqslant \delta\,\mathbb{E}[X],
\end{aligned}$$

where the last inequality holds for all sufficiently small ε since $np^{r-1}\to c>0$. This completes the proof of (37).

It remains to prove the matching lower bound

$$\liminf_{n\to\infty}\frac{\Phi_X(\delta)}{n^2p^{r-1}\log(1/p)} \geqslant \varphi_r(\delta, c). \tag{38}$$

Fix $\varepsilon > 0$ small enough and suppose that n is sufficiently large. By the continuity of φ_r, it is enough to show that any graph G on n vertices satisfying

$$\mathbb{E}_G[X] \geqslant (1+\delta)\,\mathbb{E}[X]$$

has at least $(1-\varepsilon)\cdot\varphi_r(\delta-\varepsilon,c)\cdot n^2p^{r-1}$ edges. By way of contradiction, assume that $e_G < (1-\varepsilon)\cdot\varphi_r(\delta-\varepsilon,c)\cdot n^2p^{r-1}$. Note that, for all large enough n,

$$(\delta-\varepsilon/4)\cdot\frac{n^rp^{\binom{r}{2}}}{r!} \leqslant \delta\,\mathbb{E}[X] \leqslant \mathbb{E}_G[X]-\mathbb{E}[X] \leqslant \sum_{\varnothing\neq J\subseteq K_r} N(J,G)\cdot n^{r-v_J}\cdot p^{\binom{r}{2}-e_J},$$

where the sum ranges over the nonempty subgraphs J of K_r without isolated vertices, so

$$\sum_{\varnothing\neq J\subseteq K_r}\frac{N(J,G)}{n^{v_J}p^{e_J}} \geqslant \frac{\delta-\varepsilon/3}{r!}. \tag{39}$$

Using our assumed upper bound on e_G, Theorem 5.7 implies that

$$\begin{aligned}\frac{N(J,G)}{\delta n^{v_J}p^{e_J}} &\leqslant \frac{(2e_G)^{v_J-\alpha_J^*}\cdot\min\{2e_G,n\}^{2\alpha_J^*-v_J}}{n^{v_J}p^{e_J}}\\ &\leqslant C\cdot\frac{\min\{(n^2p^{r-1})^{\alpha_J^*},n^{v_J}p^{(r-1)(v_J-\alpha_J^*)}\}}{n^{v_J}p^{e_J}}\end{aligned}$$

for a suitable constant C. If the minimum is achieved by the first term, then $np^{r-1}\leqslant 1$, and thus

$$\frac{(n^2p^{r-1})^{\alpha_J^*}}{n^{v_J}p^{e_J}} = n^{\alpha_J^*-v_J+e_J/(r-1)}\cdot\left(np^{r-1}\right)^{\alpha_J^*-e_J/(r-1)} \leqslant n^{\alpha_J^*-v_J+e_J/(r-1)},$$

as $\alpha_J^*\geqslant v_J/2\geqslant e_J/(r-1)$ for every graph J with maximum degree at most $r-1$. A straightforward algebraic manipulation in the case where the minimum is achieved by the second term then shows that, in both cases,

$$\frac{N(J,G)}{n^{v_J}p^{e_J}} \leqslant C\cdot\min\{n^{-1},p^{r-1}\}^{v_J-\alpha_J^*-e_J/(r-1)}.$$

If $J\subseteq K_r$ is not equal to either K_r or $K_{1,r-1}$, then Lemma 5.3 and the remark that follows it imply that $v_J-\alpha_J^*-e_J/(r-1)>0$, and then the right-hand side goes to zero as $n\to\infty$. It thus follows from (39) that

$$\frac{N(K_r,G)}{n^rp^{\binom{r}{2}}}+\frac{N(K_{1,r-1},G)}{n^rp^{r-1}} \geqslant \frac{\delta-\varepsilon/2}{r!},$$

or, equivalently, since $|\operatorname{Aut}(K_r)|=r!$ and $|\operatorname{Aut}(K_{1,r-1})|=(r-1)!$,

$$|\operatorname{Emb}(K_r,G)|+r\cdot|\operatorname{Emb}(K_{1,r-1},G)|\cdot p^{\binom{r}{2}-r+1} \geqslant (\delta-\varepsilon/2)\cdot n^rp^{\binom{r}{2}}. \tag{40}$$

Recalling our assumption $e_G<(1-\varepsilon)\cdot\varphi_r(\delta-\varepsilon,c)\cdot n^2p^{r-1}\ll n^2$, it follows from Lemma 5.12 that there is a partition $V(G)=U\cup V$ such that $|U|\leqslant\varepsilon^2 n$,

$$|\operatorname{Emb}(K_r,G[V])| \geqslant |\operatorname{Emb}(K_r,G)|-\varepsilon^2e_G^{r/2} \geqslant |\operatorname{Emb}(K_r,G)|-\varepsilon n^rp^{\binom{r}{2}}/4,$$

and, recalling that $\operatorname{Emb}_U(K_{1,r-1},G)$ denotes the set of embeddings of $K_{1,r-1}$ into G that map the centre vertex of $K_{1,r-1}$ to a vertex of U,

$$\begin{aligned} r\cdot|\operatorname{Emb}_U(K_{1,r-1},G[U,V])|\cdot p^{\binom{r}{2}-r+1} &\geqslant r\cdot|\operatorname{Emb}(K_{1,r-1},G)|\cdot p^{\binom{r}{2}-r+1}-\varepsilon^2e_Gn^{r-2}\cdot p^{\binom{r}{2}-r+1}\\ &\geqslant r\cdot|\operatorname{Emb}(K_{1,r-1},G)|\cdot p^{\binom{r}{2}-r+1}-\varepsilon n^rp^{\binom{r}{2}}/4\end{aligned},$$

where the stated inequalities are valid if ε is sufficiently small. From this and (40), we obtain

$$|\operatorname{Emb}(K_r,G[V])|+r\cdot|\operatorname{Emb}_U(K_{1,r-1},G[U,V])|\cdot p^{\binom{r}{2}-r+1} \geqslant (\delta-\varepsilon)\cdot n^rp^{\binom{r}{2}};$$

consequently, there exists an $x\in[0,1]$ such that

$$|\operatorname{Emb}(K_r,G[V])| \geqslant (1-x)\cdot(\delta-\varepsilon)\cdot n^rp^{\binom{r}{2}}$$

and

$$r \cdot |\operatorname{Emb}_U(K_{1,r-1}, G[U,V])| \cdot p^{\binom{r}{2}-r+1} \geqslant x \cdot (\delta - \varepsilon) \cdot n^r p^{\binom{r}{2}}.$$

By Theorem 5.7 and Lemma 5.11, we thus obtain the bounds

$$\begin{cases} (2e_{G[V]})^{r/2} \geqslant (1-x) \cdot (\delta - \varepsilon) \cdot n^r p^{\binom{r}{2}} \\ (\lfloor e_{G[U,V]}/|V| \rfloor + \{e_{G[U,V]}/|V|\}^{r-1}) \cdot n^{r-1} \geqslant x \cdot (\delta - \varepsilon) \cdot n^r p^{r-1}/r, \end{cases}$$

and solving for $e_{G[V]}$ and $e_{G[U,V]}$, we get

$$\begin{cases} e_{G[V]} \geqslant \big((1-x) \cdot (\delta - \varepsilon)\big)^{2/r} \cdot \frac{n^2 p^{r-1}}{2} \\ e_{G[U,V]} \geqslant |V| \cdot \big(\lfloor x \cdot (\delta - \varepsilon) n p^{r-1}/r \rfloor + \{x \cdot (\delta - \varepsilon) n p^{r-1}/r\}^{1/(r-1)}\big). \end{cases}$$

Finally, since $|V| = n - |U| \geqslant (1-\varepsilon^2)n$, the definition of ψ_r shows that

$$e_G \geqslant e_{G[V]} + e_{G[U,V]} \geqslant (1-\varepsilon^2) \cdot \psi_r(\delta - \varepsilon, np^{r-1}, x) \cdot n^2 p^{r-1}.$$

As $\psi_r(\delta - \varepsilon, np^{r-1}, x) \geqslant \varphi_r(\delta - \varepsilon, np^{r-1}) \to \varphi_r(\delta - \varepsilon, c)$, this contradicts our assumption that $e_G < (1-\varepsilon) \cdot \varphi_r(\delta - \varepsilon, c) \cdot n^2 p^{r-1}$, provided that ε is sufficiently small and n is large enough.

6.3. **Proof of Proposition 6.6.** Fix $\varepsilon, \delta > 0$ and $c \in [0, \infty]$ and assume that $np^{r-1} \to c$. We fix three additional positive constants η, η', and γ, where γ is sufficiently small given the parameters of the proposition, η' is sufficiently small given γ, and finally η is sufficiently small given both η' and γ.

If Near-min(η) occurs, then $G_{n,p}$ contains a subgraph G such that $e_G \leqslant (1+\eta)\Phi_X(\delta+\eta)$ and $\mathbb{E}_G[X] \geqslant (1+\delta-\eta)\,\mathbb{E}[X]$. We claim that, if n is sufficiently large, then every such graph admits a partition $V(G) = U \cup V$ such that, for some $x \in \bar{X}(\delta, c)$,

(i) V contains a subset V' of size at least $(1-\varepsilon)\big(\delta(1-x)\big)^{1/r} np^{(r-1)/2}$ such that $G[V']$ has minimum degree at least $(1-\varepsilon)|V'|$.

(ii) U contains a set $W \subseteq U$ such that at least $\lfloor (1-\varepsilon)|W| \rfloor$ vertices in W have degree at least $(1-\varepsilon)n$ and

$$e(W,V) \geqslant (1-\varepsilon)n\big(\lfloor \delta x n p^{r-1}/r \rfloor + \{\delta x n p^{r-1}/r\}^{1/(r-1)}\big).$$

Note that these properties imply $\mathrm{Clique}_\varepsilon(\delta(1-x)) \cap \mathrm{Hub}_\varepsilon(\delta x)$.

By repeating the argument found in the proof of the lower bound in Proposition 6.5, we can find a partition $V(G) = U \cup V$ and some $x' \in [0,1]$ such that

$$\begin{cases} |\operatorname{Emb}(K_r, G[V])| \geqslant (1-x') \cdot (\delta - 2\eta) \cdot n^r p^{\binom{r}{2}} \\ r \cdot |\operatorname{Emb}_U(K_{1,r-1}, G[U,V])| \cdot p^{\binom{r}{2}-r+1} \geqslant x' \cdot (\delta - 2\eta) \cdot n^r p^{\binom{r}{2}} \end{cases} \tag{41}$$

and

$$\begin{cases} e_{G[V]} \geqslant \big((1-x') \cdot (\delta - 2\eta)\big)^{2/r} \cdot \frac{n^2 p^{r-1}}{2} \\ e_{G[U,V]} \geqslant \big(\lfloor x'(\delta - 2\eta) n p^{r-1}/r \rfloor + \{x'(\delta - 2\eta) n p^{r-1}/r\}^{1/(r-1)}\big) \cdot n. \end{cases} \tag{42}$$

It follows that $e_G \geqslant e_{G[V]} + e_{G[U,V]} \geqslant \psi_r(\delta - 2\eta, np^{r-1}, x') \cdot n^2 p^{r-1}$. Thus, by Proposition 6.5,

$$\psi_r(\delta - 2\eta, np^{r-1}, x') \leqslant \frac{e_G}{n^2 p^{r-1}} \leqslant \frac{(1+\eta)\Phi_X(\delta+\eta)}{n^2 p^{r-1}} \leqslant (1+2\eta)\varphi_r(\delta+\eta, c)\cdot \tag{43}$$

Our next claim tells us how to choose the constant η. Given a set $S \subseteq \mathbb{R}$, we write $B_{\eta'}(S) = \{x \in \mathbb{R} : \inf_{s \in S} |x - s| < \varepsilon\}$ for the η'-neighbourhood of S.

Claim 6.8. *We may choose $\eta = \eta(\eta') > 0$ such that $x' \in B_{\eta'}(\bar{X}_r(\delta, c))$ whenever n is sufficiently large.*

Proof. Suppose there is no such choice. Then, by invoking (43) with successively smaller values of η, we find that there is a subsequence (η_m, c_m, x_m) of points in $(0,\infty)^2 \times [0,1]$ such that $\eta_m \to 0$ and $c_m \to c$ as $m \to \infty$, and, for all m, we have $x_m \notin B_{\eta'}(\bar{X}_r(\delta, c))$ and

$$\varphi_r(\delta - 2\eta_m, c_m) \leqslant \psi_r(\delta - 2\eta_m, c_m, x_m) \leqslant (1 + 2\eta_m)\varphi_r(\delta + \eta_m, c).$$

Since both the left-hand and the right-hand sides above converge to $\varphi_r(\delta, c)$ as $m \to \infty$, then so must $\psi_r(\delta_m, c_m, x_m)$. Denote by $(\eta_\ell, c_\ell, x_\ell)$ a subsequence on which x_ℓ converges to some $x_\infty \in [0,1]$. We claim that $x_\infty \in \bar{X}_r(\delta, c)$, which contradicts the fact that $x_\ell \notin B_{\eta'}(\bar{X}_r(\delta, c))$ for all ℓ.

If $c = 0$, then we have $x_\infty = 0 \in \bar{X}_r(\delta, c)$, since otherwise the definition of ψ_r would imply that $\psi_r(\delta_\ell, c_\ell, x_\ell) \to \infty$. If $c \in (0, \infty)$, then continuity of ψ_r implies that $\varphi_r(\delta, c) = \lim_{\ell\to\infty} \psi_r(\delta_\ell, c_\ell, x_\ell) = \psi_r(\delta, c, x_\infty)$ and thus $x_\infty \in \bar{X}_r(\delta, c)$. Finally, if $c = \infty$, then $(\delta, x) \mapsto \psi_r(\delta, c_m, x)$ converges uniformly to the continuous function $(\delta, x) \mapsto (\delta x)^{2/r}/2 + \delta x/r$, which implies that

$$\lim_{\ell\to\infty} \psi_r(\delta_\ell, c_\ell, x_\ell) = \frac{(\delta x_\infty)^{2/r}}{2} + \frac{\delta x_\infty}{r} = \lim_{c'\to\infty} \psi(\delta, c', x_\infty),$$

so $x_\infty \in \bar{X}_r(\delta, c)$ in this case as well. □

Suppose now that $x' \in B_{\eta'}(\bar{X}(\delta, c))$. Since the right-hand sides of (41) and (42) are continuous in x', we may choose η and η' sufficiently small, as a function of γ, so that there is some $x \in \bar{X}_r(\delta, c)$ such that

$$\begin{cases} |\operatorname{Emb}(K_r, G[V])| \geqslant (1-x)\cdot(\delta-\gamma)\cdot n^r p^{\binom{r}{2}} \\ r\cdot|\operatorname{Emb}_U(K_{1,r-1}, G[U,V])|\cdot p^{\binom{r}{2}-r+1} \geqslant x\cdot(\delta-\gamma)\cdot n^r p^{\binom{r}{2}}, \end{cases}$$

$$\begin{cases} e_{G[V]} \geqslant \big((1-x)\cdot(\delta-\gamma)\big)^{2/r}\cdot \frac{n^2p^{r-1}}{2} \\ e_{G[U,V]} \geqslant \big(\lfloor x(\delta-\gamma)np^{r-1}/r\rfloor + \{x(\delta-\gamma)np^{r-1}/r\}^{1/(r-1)}\big)\cdot n. \end{cases}$$

Since $e_{G[V]} + e_{G[U,V]} \leqslant e_G \leqslant (1+2\eta)\psi_r(\delta+\eta, c, x')\cdot n^2p^{r-1}$ and since the lower bounds on $e_{G[V]}$ and $e_{G[U,V]}$ in (42) sum to $\psi_r(\delta - 2\eta, np^{r-1}, x')\cdot n^2p^{r-1}$, the two lower bounds on $e_{G[V]}$ and $e_{G[U,V]}$ stated above must be nearly tight. More precisely, the continuity of ψ_r implies that we may choose η and η' sufficiently small so that

$$\begin{cases} e_{G[V]} \leqslant \big((1-x)\cdot(\delta+\gamma)\big)^{2/r}\cdot \frac{n^2p^{r-1}}{2} \\ e_{G[U,V]} \leqslant \big(\lfloor x(\delta+\gamma)np^{r-1}/r\rfloor + \{x(\delta+\gamma)np^{r-1}/r\}^{1/(r-1)}\big)\cdot n. \end{cases}$$

Finally, if we choose γ sufficiently small, then the above statements for $G[V]$ and Theorem 5.8 yield a set $V' \subseteq V$ satisfying the conditions in (i). Similarly, the statements for $G[U,V]$ and Lemma 5.11(ii) yield a subset $W \subseteq U$ satisfying (ii).

7. Extensions to regular graphs

Fix a connected and Δ-regular graph H. In this section, we apply Theorem 3.1 to study the upper tail of the random variable $X = X_{n,p}^H$. In this setting, (10) may be rewritten as

$$\Phi_X(\delta) = \min\big\{e(G)\log(1/p) : G \subseteq K_n \text{ and } \mathbb{E}_G[X] \geqslant (1+\delta)\,\mathbb{E}[X]\big\},$$

where we use the notation $\mathbb{E}_G[X] = \mathbb{E}[X \mid G \subseteq G_{n,p}]$. Our main result in this section is the following:

Proposition 7.1. *For every $\Delta \geqslant 2$, every connected, nonbipartite, Δ-regular graph H, and all positive real numbers ε and δ, there exists a positive constant C such that the following holds. Suppose that an integer n and $p \in (0,1)$ satisfy $Cn^{-1}(\log n)^{\Delta v_H^2} \leqslant p^{\Delta/2} \leqslant 1/C$. Then $X = X_{n,p}^H$ satisfies*

$$(1-\varepsilon)\Phi_X(\delta-\varepsilon) \leqslant -\log\mathbb{P}\big(X \geqslant (1+\delta)\,\mathbb{E}[X]\big) \leqslant (1+\varepsilon)\Phi_X(\delta+\varepsilon).$$

Additionally, there is a positive constant $\xi = \xi(\Delta, \varepsilon)$ such that the assumption that H is nonbipartite is not necessary when $p^{\Delta/2} \geqslant n^{-1/2-\xi}$.

Bhattacharya, Ganguly, Lubetzky, and Zhao [8] solved the optimisation problem defining $\Phi_X(\delta)$ asymptotically. Explicitly, their results imply

$$\lim_{n\to\infty} \frac{\Phi_X(\delta)}{n^2p^\Delta \log(1/p)} = \begin{cases} \delta^{2/v_H}/2 & \text{if } np^\Delta \to 0, \\ \min\{\delta^{2/v_H}/2, \theta\} & \text{if } np^\Delta \to \infty, \end{cases}$$

where θ is the unique positive solution to $P_H(\theta) = 1 + \delta$ and P_H is the independence polynomial of H. Thus, Proposition 7.1 implies Theorem 1.5. The proof of Proposition 7.1 does not require these precise estimates, but only that $\Phi_X(\delta)$ is of order $n^2p^\Delta \log(1/p)$. The methods used in [8] can certainly show this, but we include a short proof for completeness.

Lemma 7.2. *For every $\Delta \geqslant 2$, every connected, Δ-regular graph H, and every positive real number δ, there exists a positive constant C such that the following holds. Assume $n \in \mathbb{N}$ and $p \in (0,1)$ are such that $Cn^{-1} \leqslant p^{\Delta/2} \leqslant 1/C$. Then $X = X_{n,p}^H$ satisfies*

$$1/C \leqslant \frac{\Phi_X(\delta)}{n^2p^\Delta \log(1/p)} \leqslant C.$$

Proof. The upper bound follows by noting that, if C is large enough, then a clique with $\lceil (1 + 2\delta)^{1/v_H} np^{\Delta/2}\rceil$ vertices contains at least $(1+\delta)\,\mathbb{E}[X]$ copies of H and has fewer than Cn^2p^Δ edges. For the lower bound, suppose that G is a graph with $\mathbb{E}_G[X] \geqslant (1+\delta)\,\mathbb{E}[X]$ with fewer than $C^{-1}n^2p^\Delta$ edges. Then

$$\frac{\delta}{2} \cdot \frac{n^{v_H}p^{e_H}}{|\operatorname{Aut}(H)|} \leqslant \delta\,\mathbb{E}[X] \leqslant \mathbb{E}_G[X] - \mathbb{E}[X] \leqslant \sum_{\varnothing \neq J \subseteq H} N(J,G) \cdot n^{v_H - v_J} p^{e_H - e_J},$$

where the sum ranges over all nonempty subgraphs J of H without isolated vertices. This implies that there is a nonempty subgraph $J \subseteq H$ without isolated vertices and a positive constant $\gamma = \gamma(H, \delta)$ such that

$$|\operatorname{Emb}(J,G)| \geqslant \gamma \cdot n^{v_J} p^{e_J}. \tag{44}$$

Theorem 5.7 implies that

$$|\operatorname{Emb}(J,G)| \leqslant (2e_G)^{v_J - \alpha_J^*} \cdot \min\{2e_G, n\}^{2\alpha_J^* - v_J}$$

and Lemma 5.3 yields $\alpha_J^* \leqslant v_J - e_J/\Delta$. Therefore, if $2e_G \leqslant n$, then $|\operatorname{Emb}(J,G)|$ is bounded from above by

$$(2e_G)^{\alpha_J^*} \leqslant (2e_G)^{\alpha_J^* - v_J + 2e_J/\Delta} \cdot n^{v_J - 2e_J/\Delta} \leqslant \left(\frac{2n^2p^\Delta}{C}\right)^{e_J/\Delta} n^{v_J - 2e_J/\Delta} = \frac{n^{v_J}p^{e_J}}{(C/2)^{e_J/\Delta}},$$

where the first inequality holds since $2e_J/\Delta \leqslant v_J$, as the maximum degree of J is at most Δ. If $n < 2e_G$, then $|\operatorname{Emb}(J,G)|$ is bounded from above by

$$\left(\frac{2n^2p^\Delta}{C}\right)^{v_J - \alpha_J^*} \cdot n^{2\alpha_J^* - v_J} = \frac{n^{v_J}p^{e_J}}{(C/2)^{v_J - \alpha_J^*}} \cdot p^{\Delta(v_J - \alpha_J^*) - e_J} \leqslant \frac{n^{v_J}p^{e_J}}{(C/2)^{v_J - \alpha_J^*}}.$$

In both cases, the obtained upper bound on $|\operatorname{Emb}(J,G)|$ contradicts (44) whenever C is large; indeed e_J/Δ and $v_J - \alpha_J^*$ are both positive, as J is nonempty. ◻

7.1. **Proof of Proposition 7.1.** Fix $\Delta \geqslant 2$, a nonempty, connected, Δ-regular graph H, and positive reals ε and δ. Without loss of generality, we may assume that $\varepsilon \leqslant \min\{1/3, \delta/2\}$. Let $X = X_{n,p}^H$ and assume $Cn^{-1}(\log n)^{\Delta v_H^2} \leqslant p^{\Delta/2} \leqslant 1/C$. Note that the case $n < v_H$ is trivial as then X is identically zero. We may therefore assume that $n \geqslant v_H \geqslant \Delta + 1 \geqslant 3$, which, in turn, implies that $n \geqslant C$.

Set $N = \binom{n}{2}$ and let $Y = (Y_1, \ldots, Y_N)$ be the sequence of indicator random variables of the events that $e \in E(G_{n,p})$, where e ranges over $\binom{\llbracket n \rrbracket}{2}$ in some arbitrary order. Then Y is a vector of independent $\mathrm{Ber}(p)$ random variables and X is a nonzero polynomial with nonnegative coefficients and total degree at most e_H in the coordinates of Y. Let $K = K(e_H, \varepsilon, \delta)$ be the constant whose existence is asserted by Theorem 3.1. To prove the proposition, it suffices to verify the assumptions of the theorem.

It follows from Lemma 7.2 and our assumptions on p that $p \leqslant 1-\varepsilon$ and $\Phi_X(\delta-\varepsilon) \geqslant K\log(1/p)$ for a large enough choice of C. It thus only remains to bound the number of cores of a given size. To this end, let $\mathcal{I}_m^*$ be the set of cores with m edges, that is, subgraphs $G^* \subseteq K_n$ such that

(C1) $\mathbb{E}_{G^*}[X] \geqslant (1 + \delta - \varepsilon)\,\mathbb{E}[X]$,
(C2) $e_{G^*} = m \leqslant K \cdot \Phi_X(\delta + \varepsilon)$, and
(C3) $\min_{e \in E(G^*)} \big(\mathbb{E}_{G^*}[X] - \mathbb{E}_{G^* \setminus e}[X]\big) \geqslant \mathbb{E}[X]/(K \cdot \Phi_X(\delta + \varepsilon))$.

Proposition 7.1 will follow once we prove

$$|\mathcal{I}_m^*| \leqslant (1/p)^{\varepsilon m/2} \quad \text{for all } m. \tag{45}$$

Observe that due to (C1), (C2), and the definition of Φ_X, it suffices to verify (45) for integers m such that $m_{\min} \leqslant m \leqslant m_{\max}$ with

$$m_{\min} := \Phi_X(\delta - \varepsilon)/\log(1/p) \geqslant n^2 p^{\Delta}/K' \text{ and}$$
$$m_{\max} := K \cdot \Phi_X(\delta + \varepsilon) \leqslant K' \cdot n^2 p^{\Delta} \log(1/p),$$

where the stated inequalities follow, for a suitable constant $K' = K'(H, \varepsilon, \delta)$, from Lemma 7.2, our bounds on p, and the assumption that C is sufficiently large.

The first step towards establishing (45) is to understand the combinatorial meaning of (C3). Suppose that $m_{\min} \leqslant m \leqslant m_{\max}$ and $G^* \in \mathcal{I}_m^*$. Recall that $N(J, G^*; e)$ denotes the number of copies of J in G^* that contain the edge e. Note that (C3) implies that, for every $e \in E(G^*)$,

$$\frac{\mathbb{E}[X]}{m_{\max}} \leqslant \mathbb{E}_{G^*}[X] - \mathbb{E}_{G^* \setminus e}[X] \leqslant \sum_{\varnothing \neq J \subseteq H} N(J, G^*; e) \cdot n^{v_H - v_J} p^{e_H - e_J}, \tag{46}$$

where the sum ranges over the nonempty subgraphs J of H without isolated vertices. Since $n \geqslant C$ for a large enough constant C, we can bound $\mathbb{E}[X] \geqslant \binom{n}{v_H} p^{e_H} \geqslant \frac{n^{v_H} p^{e_H}}{2v_H!}$ and, consequently, (46) implies that there is a positive constant $\gamma = \gamma(H)$ such that

$$\sum_{\varnothing \neq J \subseteq H} \frac{N(J, G^*; e)}{n^{v_J} p^{e_J}} \geqslant \frac{2\gamma}{m_{\max}} \qquad \text{for every } e \in E(G^*). \tag{47}$$

Definition 7.3. Let $\mathcal{Q}_H$ denote the family of all nonempty subgraphs $J \subseteq H$ without isolated vertices satisfying

(Q1) $J = H$ or
(Q2) J admits a bipartition $V(J) = A \cup B$ such that $\deg_J a = \Delta$ for all $a \in A$.

Our first claim is that, for the vast majority of $e \in E(G^*)$, the sum on the left-hand side of (47) is dominated by subgraphs $J \in \mathcal{Q}_H$. Let G_{exc}^* comprise all edges e of G^* such that

$$\sum_{J \in \mathcal{Q}_H} \frac{N(J, G^*; e)}{n^{v_J} p^{e_J}} < \frac{\gamma}{m_{\max}}.$$

Claim 7.4. *There is a positive constant $\sigma = \sigma(H)$ such that $e_{G_{\mathrm{exc}}^*} \leqslant p^{\sigma} \cdot m_{\min}$.*

Proof. Let m_{exc} denote the number of edges of G^*_{exc}. The definition of G^*_{exc} and (47) imply that

$$\sum_{e\in E(G^*_{\text{exc}})} \sum_{\substack{\varnothing\neq J\subseteq H\\ J\notin\mathcal{Q}_H}} \frac{N(J,G^*;e)}{n^{v_J}p^{e_J}} \geqslant \frac{\gamma\cdot m_{\text{exc}}}{m_{\max}} \geqslant \frac{\gamma\cdot m_{\text{exc}}}{m_{\min}\cdot (K')^2\cdot \log(1/p)}.$$

On the other hand, since $\sum_{e\in E(G^*)} N(J,G^*;e) \leqslant e_J\cdot|\operatorname{Emb}(J,G^*)|$ for every graph J, there must exist a nonempty subgraph $J\subseteq H$ without isolated vertices such that $J\notin\mathcal{Q}_H$ and

$$\frac{m_{\text{exc}}}{m_{\min}} \leqslant \frac{e_J\cdot (K')^2\cdot\log(1/p)\cdot 2^{e_H+v_H}}{\gamma}\cdot\frac{|\operatorname{Emb}(J,G^*)|}{n^{v_J}p^{e_J}}. \tag{48}$$

We now show that the right-hand side of (48) is at most p^σ, for some positive constant $\sigma=\sigma(H)$. To this end, recall that Theorem 5.7 states that

$$|\operatorname{Emb}(J,G^*)| \leqslant (2m)^{v_J-\alpha^*_J}\cdot\min\{2m,n\}^{2\alpha^*_J-v_J}. \tag{49}$$

Since J is a proper subgraph of a connected, Δ-regular graph, then it must contain a vertex of degree smaller than Δ and hence $v_J > 2e_J/\Delta$. Moreover, since $J\notin\mathcal{Q}_H$, Lemma 5.3 implies that $\alpha^*_J < v_J - e_J/\Delta$. Therefore, if $2m\leqslant n$, then there is a positive $\sigma=\sigma(H)$ such that the right-hand side of (49) can be bounded from above as follows:

$$\begin{aligned}(2m)^{\alpha^*_J} &\leqslant (2m)^{\alpha^*_J-v_J+2e_J/\Delta+2\sigma}\cdot n^{v_J-2e_J/\Delta-2\sigma}\\ &\leqslant (2m)^{e_J/\Delta}\cdot n^{v_J-2e_J/\Delta-2\sigma}\\ &\leqslant \big(2K'\cdot n^2p^\Delta\log(1/p)\big)^{e_J/\Delta}\, n^{v_J-2e_J/\Delta-2\sigma}\\ &= n^{-2\sigma}\cdot\big(2K'\cdot\log(1/p)\big)^{e_J/\Delta}\cdot n^{v_J}p^{e_J}.\end{aligned}$$

Similarly, if $n<2m$, there is a positive $\sigma=\sigma(H)$ such that the right-hand side of (49) is bounded from above by

$$\begin{aligned}(2m)^{v_J-\alpha^*_J}\cdot n^{2\alpha^*_J-v_J} &\leqslant \big(2K'\cdot n^2p^\Delta\log(1/p)\big)^{v_J-\alpha^*_J}\cdot n^{2\alpha^*_J-v_J}\\ &= p^{\Delta(v_J-\alpha^*_J)-e_J}\cdot\big(2K'\cdot\log(1/p)\big)^{v_J-\alpha^*_J}\cdot n^{v_J}p^{e_J}\\ &\leqslant p^{2\sigma}\cdot\big(2K'\cdot\log(1/p)\big)^{v_J-\alpha^*_J}\cdot n^{v_J}p^{e_J}.\end{aligned}$$

Since $p\leqslant C^{-2/\Delta}$ and $Cn^{-1}\leqslant p$, it follows that, in both cases,

$$|\operatorname{Emb}(J,G^*)| \leqslant p^\sigma\cdot n^{v_J}p^{e_J}\cdot\frac{\gamma}{e_J\cdot(K')^2\cdot\log(1/p)\cdot 2^{e_H+v_H}},$$

provided that C is sufficiently large. Substituting this inequality into (48) proves the claim. □

The next claim shows that the endpoints of every edge of $G^*\setminus G^*_{\text{exc}}$ satisfy a certain degree restriction.

Claim 7.5. *There is a positive $\gamma'=\gamma'(H,K')$ such that, for every edge uv of $G^*\setminus G^*_{\text{exc}}$, either*

$$\deg_{G^*}u\cdot\deg_{G^*}v \geqslant \frac{\gamma'}{\big(\log(1/p)\big)^{v_H}}\cdot m \qquad\text{or}\qquad \deg_{G^*}u+\deg_{G^*}v \geqslant \frac{\gamma'}{\big(\log(1/p)\big)^{v_H/2}}\cdot n.$$

Moreover, if the second inequality fails, then uv is contained in at least $\gamma' n^{v_H}p^{e_H}/m_{\max}$ copies of H in G^*.

Proof. Suppose that uv is an edge of $G^*\setminus G^*_{\text{exc}}$. It follows from (47) and the definition of G^*_{exc} that there is some $J\in\mathcal{Q}_H$ such that

$$|\operatorname{Emb}(J,G;uv)| \geqslant N(J,G;uv) \geqslant \frac{\gamma}{|\mathcal{Q}_H|}\cdot\frac{n^{v_J}p^{e_J}}{m_{\max}}. \tag{50}$$

Let $\mu = n^2p^\Delta/m$ and observe that $\mu \geqslant \big(K'\cdot\log(1/p)\big)^{-1}\cdot m_{\max}/m$. We split the remainder of the proof into two cases, depending on whether or not $J=H$.

Assume first $J = H$. Then, since $n^{v_H}p^{e_H} = \big(n^2p^\Delta\big)^{v_H/2} = (\mu m)^{v_H/2}$, Lemma 5.13 and (50) imply that

$$\begin{aligned}\big(4\deg_{G^*}u\cdot\deg_{G^*}v\big)^{\frac{\Delta-1}{\Delta}} &\geqslant |\operatorname{Emb}(J,G;uv)|\cdot\frac{(2m)^{\frac{2\Delta-1}{\Delta}-\frac{v_H}{2}}}{4e_H}\\ &\geqslant \frac{\gamma}{|\mathcal{Q}_H|}\cdot\frac{(\mu m)^{v_H/2}}{m_{\max}}\cdot\frac{(2m)^{\frac{2\Delta-1}{\Delta}-\frac{v_H}{2}}}{4e_H}\\ &= \frac{2^{\frac{2\Delta-1}{\Delta}-\frac{v_H}{2}}\gamma}{4e_H|\mathcal{Q}_H|}\cdot\mu^{v_H/2}\cdot\frac{m}{m_{\max}}\cdot m^{\frac{\Delta-1}{\Delta}}\\ &\geqslant (4\gamma')^{\frac{\Delta-1}{\Delta}}\cdot\big(\log(1/p)\big)^{-v_H/2}\cdot m^{\frac{\Delta-1}{\Delta}},\end{aligned}$$

for a suitable positive constant $\gamma' = \gamma'(H, K')$. Since $\Delta \geqslant 2$, this implies the claimed lower bound on $\deg_{G^*}u\cdot\deg_{G^*}v$.

Assume now that $J \neq H$. In this case, J admits a bipartition $V(J) = A\cup B$ such that every vertex in A has degree Δ. In particular, $e_J = |A|\cdot\Delta$; moreover, as $J\neq H$ and H is connected and Δ-regular, we also have $|B| > |A|$. Since $n^{v_J}p^{e_J} = \big(n^2p^\Delta\big)^{|A|}\cdot n^{|B|-|A|} = (\mu m)^{|A|}\cdot n^{|B|-|A|}$, it follows from Lemma 5.14 and (50) that

$$\begin{aligned}\deg_{G^*}u+\deg_{G^*}v &\geqslant \frac{|\operatorname{Emb}(J,G;uv)|}{e_J\cdot(2m)^{|A|-1}\cdot\big(\min\{m,n\}\big)^{|B|-|A|-1}}\\ &\geqslant \frac{\gamma}{|\mathcal{Q}_H|}\cdot\frac{(\mu m)^{|A|}\cdot n^{|B|-|A|}}{m_{\max}}\cdot\frac{1}{e_J\cdot(2m)^{|A|-1}\cdot\big(\min\{m,n\}\big)^{|B|-|A|-1}}\\ &\geqslant \frac{\gamma}{|\mathcal{Q}_H|}\cdot\frac{\mu^{|A|}m}{m_{\max}}\cdot\frac{n}{e_J\cdot 2^{|A|-1}}\\ &\geqslant \gamma'\cdot\big(\log(1/p)\big)^{-|A|}\cdot n\end{aligned}$$

for a suitable positive constant γ'. Since $|A| < v_H/2$, this implies the claimed lower bound on $\deg_{G^*}u+\deg_{G^*}v$. In particular, if the second inequality in the statement of the claim fails, then the above shows that $J = H$, and thus the second assertion of the claim follows from (50). $\square$

To prove (45), we will further distinguish between the cases $p^{\Delta/2} \geqslant n^{-1/2-\xi}$ and $p^{\Delta/2} \leqslant n^{-1/2-\xi}$, for a small constant ξ. The first case is easier and is handled by our next claim.

Claim 7.6. *There is a positive constant $\xi = \xi(\Delta,\varepsilon)$ such that* (45) *holds if $p^{\Delta/2} \geqslant n^{-1/2-\xi}$.*

Proof. To simplify the presentation, we shall prove (45) with 7ε instead of $\varepsilon/2$. Suppose that G^* is a core with m edges, where $m_{\min}\leqslant m\leqslant m_{\max}$. For each $k\in\mathbb{Z}$, we define the set

$$B_k = \big\{v\in V(G^*) : \deg_{G^*}v \geqslant e^k\sqrt{m}p^\varepsilon\big\}$$

and let $k_{\max}$ be the largest integer such that $e^{k_{\max}}\sqrt{m}\leqslant np^{2\varepsilon}$. Note that the bounds $n^2p^\Delta/K' \leqslant m \leqslant K'\cdot n^2p^\Delta\log(1/p)$ and our assumption $p\leqslant C^{-2/\Delta}$ imply that $0\leqslant k_{\max}\leqslant(\Delta/2)\cdot\log(1/p)$ whenever C is large enough.

We will first prove that Claim 7.5 implies that, for each edge uv of $G^*\setminus G^*_{\mathrm{exc}}$, either

(i) uv has an endpoint in $B_{k_{\max}}$ or
(ii) $u\in B_k$ and $v\in B_{-k}$ for some $k\in\{-k_{\max},\ldots,k_{\max}\}$.

Indeed, if we suppose that

$$\deg_{G^*}u+\deg_{G^*}v \geqslant \frac{\gamma'\cdot n}{(\log(1/p))^{v_H/2}},$$

then (i) follows immediately (with room to spare) because $\frac{\gamma'\cdot n}{(\log(1/p))^{v_H/2}} \geqslant np^{3\varepsilon}\geqslant 2e^{k_{\max}}\sqrt{m}p^\varepsilon$. On the other hand, if $u,v\notin B_{k_{\max}}$, then the above lower bound on $\deg_{G^*}u+\deg_{G^*}v$ does not hold and Claim 7.5 guarantees

$$\deg_{G^*} u \cdot \deg_{G^*} v \geqslant \frac{\gamma' \cdot m}{(\log(1/p))^{v_H}} \geqslant emp^{2\varepsilon},$$

which easily implies (ii); indeed the largest integer k such that $u, v \in B_k$ satisfies (ii).

To make use of this property of the edges of $G^* \setminus G^*_{\text{exc}}$, we also require upper bounds on the cardinalities of the sets B_k. To that end, observe that, for all $k \in \mathbb{Z}$,

$$2m \geqslant \sum_{v \in B_k} \deg_{G^*} v \geqslant |B_k| \cdot e^k \sqrt{m} p^{\varepsilon}$$

and hence

$$|B_k| \leqslant 2e^{-k} \sqrt{m} p^{-\varepsilon}. \tag{51}$$

We claim that there is a positive constant $\xi = \xi(\Delta, \varepsilon)$ such that $n^2 p^{\Delta} \geqslant 2K' np^{\varepsilon} \log n$ whenever $p^{\Delta/2} \geqslant n^{-1/2-\xi}$ and C is sufficiently large. Indeed, if $p^{\Delta/2} \geqslant n^{-1/3}$, then $n^2 p^{\Delta} \geqslant n^{4/3} \geqslant 2K' np^{\varepsilon} \log n$, provided that C is sufficiently large; otherwise, $p^{\varepsilon} \leqslant n^{-2\varepsilon/(3\Delta)}$ and our assumption implies that $n^2 p^{\Delta} \geqslant n^{1-2\xi} \geqslant K' np^{\varepsilon} \log n$, provided that ξ is sufficiently small and C is sufficiently large. Now inequality (51) and the choice of $k_{\max}$ imply

$$|B_{-k_{\max}}| \leqslant 2e^{k_{\max}} \sqrt{m} p^{-\varepsilon} \leqslant 2np^{\varepsilon} \leqslant \frac{m}{\log n} \tag{52}$$

and (since $e^{k_{\max}+1} > np^{2\varepsilon}$)

$$|B_{k_{\max}}| \cdot n \leqslant 2e^{-k_{\max}} \sqrt{m} p^{-\varepsilon} \cdot n < 2emp^{-3\varepsilon}. \tag{53}$$

Recall from Claim 7.4 that $e_{G^*_{\text{exc}}} \leqslant p^{\sigma} \cdot m_{\min}$ for a positive constant σ depending only on H. It follows that we may construct each $G^* \in \mathcal{I}^*_m$ as follows:

(1) Choose some $m_{\text{exc}} \leqslant p^{\sigma} \cdot m_{\min}$ and then choose m_{exc} edges of K_n to form G^*_{exc}.
(2) Choose the sets $B_{-k_{\max}}, \ldots, B_{k_{\max}}$ and then choose $m - m_{\text{exc}}$ edges from

$$\mathcal{B} = \big\{ uv \in E(K_n) : u \in B_{k_{\max}} \big\} \cup \bigcup_{k=0}^{k_{\max}} \big\{ uv \in E(K_n) : u \in B_k, v \in B_{-k} \big\}$$

to form $G^* \setminus G^*_{\text{exc}}$.

Since $B_{k_{\max}} \subseteq \cdots \subseteq B_{-k_{\max}}$ and $|B_{-k_{\max}}| \leqslant m/\log n$ by (52), the number of ways to choose the sets $B_{-k_{\max}}, \ldots, B_{k_{\max}}$ is at most

$$\big((2k_{\max} + 2) \cdot n\big)^{m/\log n} \leqslant e^{2m},$$

using the (very crude) bound $k_{\max} \leqslant n/4$. Moreover, inequalities (51) and (53) imply that

$$|\mathcal{B}| \leqslant |B_{k_{\max}}| \cdot n + \sum_{k=0}^{k_{\max}} |B_k| \cdot |B_{-k}| \leqslant 2emp^{-3\varepsilon} + (k_{\max} + 1) \cdot 4mp^{-2\varepsilon} \leqslant mp^{-4\varepsilon},$$

where we use $k_{\max} \leqslant (\Delta/2) \cdot \log(1/p)$ and $p^{\Delta/2} \leqslant 1/C$ for a large enough C. We conclude that

$$|\mathcal{I}^*_m| \leqslant e^{2m} \cdot \sum_{m_{\text{exc}}=0}^{p^{\sigma} \cdot m_{\min}} \binom{n^2}{m_{\text{exc}}} \cdot \binom{mp^{-4\varepsilon}}{m - m_{\text{exc}}}.$$

In order to bound the right-hand side above, we note that

$$\binom{mp^{-4\varepsilon}}{m - m_{\text{exc}}} \leqslant \binom{mp^{-4\varepsilon}}{m} \leqslant \left(\frac{emp^{-4\varepsilon}}{m} \right)^m \leqslant p^{-5\varepsilon m}.$$

Moreover, using the inequalities $\sum_{i=0}^{k} \binom{n}{i} \leqslant (en/k)^k$ and $m \geqslant m_{\min} \geqslant n^2 p^{\Delta}/K'$,

$$\sum_{m_{\text{exc}}=0}^{p^{\sigma} \cdot m_{\min}} \binom{n^2}{m_{\text{exc}}} \leqslant \left(\frac{en^2}{p^{\sigma} \cdot m_{\min}} \right)^{p^{\sigma} \cdot m_{\min}} \leqslant \left(\frac{eK'}{p^{\Delta+\sigma}} \right)^{mp^{\sigma}} \leqslant p^{-\varepsilon m}.$$

We may conclude that

$$|\mathcal{I}_m^*| \leqslant e^{2m} \cdot p^{-6\varepsilon m} \leqslant p^{-7\varepsilon m},$$

which completes the proof of (45) (with 7ε instead of $\varepsilon/2$). □

The argument above shows that we do not need the assumption that H is nonbipartite when $p^{\Delta/2} \geqslant n^{-1/2-\xi}$. In the following, we will assume that $p^{\Delta/2} \leqslant n^{-1/2-\xi}$ and that H is not bipartite.

Let γ' be the constant from Claim 7.5. Our assumption on p implies that

$$m_{\max} \leqslant K' \cdot n^2 p^{\Delta} \log(1/p) \leqslant K' \cdot n^{1-2\xi} \log(1/p) < \frac{\gamma' \cdot n}{\big(\log(1/p)\big)^{v_H/2}} - 1,$$

where the last inequality holds because $n \geqslant C$ and C is large. In particular, if $G^* \in \mathcal{I}_m^*$ for some $m_{\min} \leqslant m \leqslant m_{\max}$, then for any two vertices $u, v \in V(G^*)$, we have $\deg_{G^*}(u) + \deg_{G^*}(v) \leqslant m_{\max} + 1 < \gamma' n/(\log(1/p))^{v_H/2}$, so it follows from Claim 7.5 that

$$\deg_{G^*} u \cdot \deg_{G^*} v \geqslant \frac{\gamma'}{\big(\log(1/p)\big)^{v_H}} \cdot m \qquad \text{for every } uv \in E(G^* \setminus G^*_{\mathrm{exc}}) \tag{54}$$

and that every edge uv of $G^* \setminus G^*_{\mathrm{exc}}$ belongs to at least $\gamma' n^{v_H} p^{e_H}/m_{\max}$ copies of H in G^*. Set

$$\beta = \Delta(1 + v_H/2) + 1/2,$$

let G^*_{high} comprise all edges uv of G^* such that

$$\deg_{G^*} u \cdot \deg_{G^*} v \geqslant \big(\log(1/p)\big)^{\beta} \cdot m, \tag{55}$$

and denote by m_{high} the number of edges in G^*_{high}. We claim that

$$m_{\mathrm{high}} \leqslant \frac{8m}{\big(\log(1/p)\big)^{\beta}}. \tag{56}$$

In order to show it, we estimate the number of copies of P_4, the path with four vertices (and three edges), in G^* in two different ways. On the one hand,

$$2N(P_4, G^*) = |\operatorname{Emb}(P_4, G^*)| \leqslant (2m)^2,$$

since every embedding of P_4 into G^* is determined by the images of its two nonincident edges. On the other hand,

$$\begin{aligned}
N(P_4, G^*) &\geqslant \sum_{uv \in E(G^*)} (\deg_{G^*} u - 1) \cdot (\deg_{G^*} v - 2) \\
&= \sum_{uv \in E(G^*)} \deg_{G^*} u \cdot \deg_{G^*} v - 3 \sum_{v \in V(G^*)} (\deg_{G^*} v)^2 + 2m \\
&\geqslant m_{\mathrm{high}} \cdot \big(\log(1/p)\big)^{\beta} \cdot m - 3m \sum_{v \in V(G^*)} \deg_{G^*} v \\
&= m_{\mathrm{high}} \cdot \big(\log(1/p)\big)^{\beta} \cdot m - 6m^2.
\end{aligned}$$

These two lower and upper bounds on $N(P_4, G^*)$ imply (56).

Claim 7.7. *Suppose that φ is an embedding of H into $G^* \setminus (G^*_{\mathrm{exc}} \cup G^*_{\mathrm{high}})$. Then for every $a \in V(H)$,*

$$\deg_{G^*} \varphi(a) \geqslant \frac{m^{1/2}}{\big(\log(1/p)\big)^{\beta \cdot v_H}}.$$

Proof. Define $f\colon V(H) \to \mathbb{R}$ by

$$f(a) = \log\left(\frac{\deg_{G^*}\varphi(a)}{m^{1/2}}\right)$$

and let

$$f^* = \beta \log\log(1/p).$$

It suffices to show that $f(a) \geqslant -v_H f^*$ for every $a \in V(H)$. To this end, note that (54) and our definition of G^*_{high} (see (55)) imply that

$$-f^* \leqslant f(a) + f(b) \leqslant f^* \qquad \text{for every } ab \in E(H), \tag{57}$$

since $\beta \geqslant v_H + 1$ and $p \leqslant C^{-2/\Delta}$ for a large constant C. Since H is not bipartite, it contains an odd cycle. Let Z be one such cycle and suppose that $a_0, \ldots, a_{2\ell}$ are its vertices (listed in an arbitrarily chosen cyclic ordering). It follows from (57), applied to all $2\ell+1$ edges of Z, that

$$2f(a_0) = f(a_0) + f(a_{2\ell}) + \sum_{i=0}^{2\ell-1} (-1)^i\big(f(a_i) + f(a_{i+1})\big) \in \big[-(2\ell+1)f^*, (2\ell+1)f^*\big].$$

Since the particular choice of a_0 among all vertices of Z was arbitrary, we may conclude that

$$-(2\ell+1)f^* \leqslant f(a) \leqslant (2\ell+1)f^* \qquad \text{for every } a \in V(Z),$$

with room to spare. Since $2\ell + 1 = v_Z \leqslant v_H$, this proves the desired inequality for all $a \in V(Z)$. Suppose now that $b \in V(H) \setminus V(Z)$. Since H is connected, it contains a path from b to Z. Let $b_0, b_1, \ldots, b_{\ell'}$, where $b_0 = b$ and $b_{\ell'} \in V(Z)$, be the vertices of a shortest such path (listed in their natural order) and note that $\ell' + 2\ell + 1 \leqslant v_H$. It follows from (57), applied to all ℓ' edges of the path, that

$$f(b) + (-1)^{\ell'-1} f(b_{\ell'}) = \sum_{i=0}^{\ell'-1} (-1)^i\big(f(b_i) + f(b_{i+1})\big) \in \big[-\ell' f^*, \ell' f^*\big]$$

and consequently, as $b_{\ell'} \in V(Z)$, that

$$f(b) \geqslant -\ell' f^* - |f(b'_\ell)| \geqslant -(\ell' + 2\ell + 1)f^* \geqslant -v_H f^*,$$

as claimed. □

Let G^*_{bad} comprise all edges of G^* that do not belong to a copy of H in the graph $G^* \setminus (G^*_{\mathrm{exc}} \cup G^*_{\mathrm{high}})$ and let m_{bad} be the number of such edges; note that $G^*_{\mathrm{bad}} \supseteq G^*_{\mathrm{exc}} \cup G^*_{\mathrm{high}}$. Since each edge of $G^*_{\mathrm{bad}} \setminus G^*_{\mathrm{exc}}$ belongs to at least $\gamma' n^{v_H} p^{e_H} / m_{\max}$ copies of H in G^*, none of which are in $G^* \setminus G^*_{\mathrm{exc}} \cup G^*_{\mathrm{high}}$, we have

$$(m_{\mathrm{bad}} - m_{\mathrm{exc}}) \cdot \frac{\gamma' n^{v_H} p^{e_H}}{m_{\max}} \leqslant \sum_{e \in E(G^*_{\mathrm{exc}} \cup G^*_{\mathrm{high}})} |\operatorname{Emb}(H, G^*; e)|.$$

It follows from Lemma 5.15, Claim 7.4, and inequality (56) that

$$\sum_{e \in E(G^*_{\mathrm{exc}} \cup G^*_{\mathrm{high}})} |\operatorname{Emb}(H, G^*; e)| \leqslant e_H \cdot (2m)^{v_H/2} \cdot \left(\frac{m_{\mathrm{exc}} + m_{\mathrm{high}}}{m}\right)^{1/\Delta} \leqslant 8e_H \cdot \frac{(2m)^{v_H/2}}{\big(\log(1/p)\big)^{\beta/\Delta}}.$$

Consequently,

$$\begin{aligned}
m_{\text{bad}} &\leqslant m_{\text{exc}} + \frac{m_{\max}}{\gamma' n^{v_H} p^{e_H}} \cdot 8e_H \cdot \frac{(2m)^{v_H/2}}{\big(\log(1/p)\big)^{\beta/\Delta}} \\
&\leqslant p^{\sigma} \cdot m_{\min} + \frac{8 \cdot 2^{v_H/2} \cdot e_H}{\gamma'} \cdot \frac{m_{\max}^{v_H/2}}{n^{v_H} p^{e_H} \big(\log(1/p)\big)^{\beta/\Delta}} \cdot m \\
&\leqslant \left(p^{\sigma} + \frac{8 \cdot 2^{v_H/2} \cdot e_H}{\gamma'} \cdot \frac{\big(K' \cdot n^2 p^{\Delta} \log(1/p)\big)^{v_H/2}}{n^{v_H} p^{e_H} \big(\log(1/p)\big)^{\beta/\Delta}} \right) \cdot m \\
&= \left(p^{\sigma} + \frac{8 \cdot e_H \cdot \big(2K'\big)^{v_H/2}}{\gamma' \cdot \big(\log(1/p)\big)^{\beta/\Delta - v_H/2}} \right) \cdot m \leqslant \frac{m}{\log(1/p)},
\end{aligned}$$

since $\beta/\Delta = 1 + v_H/2 + 1/(2\Delta)$ and $p \leqslant C^{-2/\Delta}$ for a large constant C.

Let U be the set of nonisolated vertices of $G^* \setminus G^*_{\text{bad}}$. It follows from Claim 7.7 that

$$|U| \cdot \frac{m^{1/2}}{\big(\log(1/p)\big)^{\beta \cdot v_H}} \leqslant \sum_{v \in U} \deg_{G^*} v \leqslant 2m$$

and thus

$$|U| \leqslant 2m^{1/2} \cdot \big(\log(1/p)\big)^{\beta \cdot v_H} \leqslant m/\log n,$$

as $p^{\Delta/2} \geqslant Cn^{-1}(\log n)^{\Delta v_H^2}$ and C is large enough.

To summarise, we may construct each $G^* \in \mathcal{I}^*_m$ as follows:

(1) Choose $m_{\text{bad}} \leqslant m/\log(1/p)$ and the m_{bad} edges of K_n to form G^*_{bad}.
(2) Choose the vertices of U and the edges of $G^* \setminus G^*_{\text{bad}}$ from the set $\binom{U}{2}$.

Using the above bounds on the size of U, we conclude that

$$|\mathcal{I}^*_m| \leqslant \sum_{m_{\text{bad}}=0}^{m/\log(1/p)} \binom{n^2}{m_{\text{bad}}} \cdot n^{m/\log n} \cdot \binom{4m \cdot \big(\log(1/p)\big)^{2\beta \cdot v_H}}{m - m_{\text{bad}}}.$$

In order to bound the right-hand side from above, we note that, for sufficiently large C,

$$\binom{4m \cdot \big(\log(1/p)\big)^{2\beta \cdot v_H}}{m - m_{\text{bad}}} \leqslant \binom{4m \cdot \big(\log(1/p)\big)^{2\beta \cdot v_H}}{m} \leqslant \left(\frac{4em \cdot \big(\log(1/p)\big)^{2\beta \cdot v_H}}{m} \right)^{m} \leqslant p^{-\varepsilon m/6}$$

and, since $m \geqslant n^2 p^{\Delta}/K'$,

$$\sum_{m_{\text{bad}}=0}^{m/\log(1/p)} \binom{n^2}{m_{\text{bad}}} \leqslant \left(\frac{en^2 \log(1/p)}{m} \right)^{m/\log(1/p)} \leqslant \left(\frac{eK' \cdot \log(1/p)}{p^{\Delta}} \right)^{m/\log(1/p)} \leqslant p^{-\varepsilon m/6}.$$

Since $n^{m/\log n} = e^m \leqslant p^{-\varepsilon m/6}$, we may conclude that $|\mathcal{I}^*_m| \leqslant p^{-\varepsilon m/2}$, which completes the proof of Proposition 7.1.

8. The Poisson regime

Given a nonnegative real μ, we shall denote by $\mathrm{Po}(\mu)$ the Poisson distribution with mean μ. Suppose that $X \sim \mathrm{Po}(\mu)$. A classical result in large deviation theory is that, for every fixed $\delta > 0$,

$$-\log \mathbb{P}\big(X \geqslant (1+\delta)\mu\big) = \big((1+\delta)\log(1+\delta) - \delta\big)\mu + o(\mu),$$

as $\mu \to \infty$. Motivated by this estimate, for any random variable X with positive expectation and any $\delta > 0$, we define

$$\Psi_X(\delta) = \big((1+\delta)\log(1+\delta) - \delta\big)\, \mathbb{E}[X].$$

Theorems 1.4 and 1.6 follow immediately from the following two propositions.

Proposition 8.1. *For every integer $k \geqslant 3$ and all positive real numbers ε and δ, there exists a positive constant C such that the following holds. Suppose that $N \in \mathbb{N}$ and $p \in (0,1)$ satisfy $CN^{-1} \leqslant p^{k/2} \leqslant C^{-1}N^{-1}\log N$. Then $X = X_{N,p}^{k\text{-}AP}$ satisfies*

$$(1-\varepsilon)\Psi_X(\delta) \leqslant -\log \mathbb{P}\big(X \geqslant (1+\delta)\,\mathbb{E}[X]\big) \leqslant (1+\varepsilon)\Psi_X(\delta).$$

Proposition 8.2. *For every $\Delta \geqslant 2$, every connected, Δ-regular graph H, and all positive real numbers ε and δ, there exists a positive constant C such that the following holds. Suppose that $n \in \mathbb{N}$ and $p \in (0,1)$ satisfy $Cn^{-1} \leqslant p^{\Delta/2} \leqslant C^{-1}n^{-1}(\log n)^{\frac{1}{v_H-2}}$. Then $X = X_{n,p}^{H}$ satisfies*

$$(1-\varepsilon)\Psi_X(\delta) \leqslant -\log \mathbb{P}\big(X \geqslant (1+\delta)\,\mathbb{E}[X]\big) \leqslant (1+\varepsilon)\Psi_X(\delta).$$

It is not difficult to show that the requirements on p in these results are optimal up to the choice of the constant C. Indeed, if $X = X_{N,p}^{k\text{-AP}}$, then planting an interval of length $C_\delta N p^{k/2}$ creates $(1+\delta)\,\mathbb{E}[X]$ k-term arithmetic progressions (for a sufficiently large C_δ), which shows that

$$-\log \mathbb{P}\big(X \geqslant (1+\delta)\,\mathbb{E}[X]\big) = O\big(Np^{k/2}\log(1/p)\big).$$

Similarly, if $X = X_{n,p}^{H}$, then planting a clique of size $C_\delta n p^{\Delta/2}$ results in $(1+\delta)\,\mathbb{E}[X]$ copies of H, which proves

$$-\log \mathbb{P}\big(X \geqslant (1+\delta)\,\mathbb{E}[X]\big) = O\big(n^2p^{\Delta}\log(1/p)\big).$$

The upper bounds are $o(\Psi_X(\delta))$, and therefore dominate the Poisson bounds, whenever $p^{k/2} \gg N^{-1}\log N$ or $p^{\Delta/2} \gg n^{-1}(\log n)^{\frac{1}{v_H-2}}$, respectively.

8.1. **Poisson approximation via factorial moments.** For a real number x and a nonnegative integer t, write $m^{\underline{t}} = m(m-1)\cdots(m-t+1)$ for the t-th falling factorial of m. For a random variable X, let $M_t(X) = \mathbb{E}[X^{\underline{t}}]$ be the t-th factorial moment of X. It is straightforward to verify that, if $X \sim \mathrm{Po}(\mu)$, then $M_t(X) = \mu^t$ for all $t \geqslant 0$.

A classical application of the method of moments is that, if (X_n) is a sequence of random variables whose t-th factorial moments converge to μ^t for some fixed μ, then X_n converge in distribution to $\mathrm{Po}(\mu)$. The lemma below can be viewed as an extension of this result to the case when $\mu \to \infty$. It states that, if the t-th factorial moment of some random variable X is approximately μ^t, for each t around $\delta\mu$, then the logarithmic upper tail probability $-\log \mathbb{P}\big(X \geqslant (1+\delta)\,\mathbb{E}[X]\big)$ is well-approximated by $\Psi_X(\delta)$.

Proposition 8.3. *For all positive real numbers ε and δ, there exists a positive constant η such that the following holds. Let X be a nonnegative integer-valued random variable with mean $\mu \geqslant 1/\eta$ such that $|M_t(X) - \mu^t| \leqslant \eta\mu^t$ for every integer t satisfying $(\delta-\varepsilon)\mu \leqslant t \leqslant (\delta+\varepsilon)\mu$. Then*

$$(1-\varepsilon)\Psi_X(\delta) \leqslant -\log \mathbb{P}\big(X \geqslant (1+\delta)\,\mathbb{E}[X]\big) \leqslant (1+\varepsilon)\Psi_X(\delta).$$

Define the continuous function $I\colon [0,\infty) \to [0,\infty)$ by

$$I(\delta) = (1+\delta)\log(1+\delta) - \delta,$$

so that $\Psi_X(\delta) = I(\delta)\cdot\mu$. Note that $I(\delta) > 0$ whenever $\delta > 0$.

Lemma 8.4. *For every nonnegative integer t and every positive real x,*

$$\log\,(x+t)^{\underline{t}} = I(t/x)\cdot x + t\log x + \lambda(x,t),$$

where $0 \leqslant \lambda(x,t) \leqslant (t+1)/x$.

Proof. Observe that $\log\,(x+t)^{\underline{t}} = \sum_{s=1}^{t}\log(x+s)$. Since log is an increasing function, then

$$\int_x^{x+t}\log y\,dy \leqslant \sum_{s=1}^{t}\log(x+s) \leqslant \int_{x+1}^{x+t+1}\log y\,dy.$$

Recalling that $\int \log y \, dy = y(\log y - 1) + C$, we have

$$\int_x^{x+t} \log y \, dy = \big((1+t/x)\log(1+t/x) - t/x\big) \cdot x + t\log x = I(t/x) \cdot x + t \log x.$$

On the other hand,

$$\int_{x+1}^{x+t+1} \log y \, dy - \int_x^{x+t} \log y \, dy \leqslant \log(x+t+1) - \log x = \log\big(1 + (t+1)/x\big) \leqslant (t+1)/x.$$

This proves the claimed estimate. □

Proof of Proposition 8.3. Since, for every positive integer t, the function $x \mapsto x^{\underline{t}}$ is increasing on $[t-1, \infty)$ and nonnegative on $\mathbb{Z}$, Markov's inequality implies that

$$\mathbb{P}\big(X \geqslant (1+\delta)\,\mathbb{E}[X]\big) \leqslant \mathbb{P}\left(X^{\underline{t}} \geqslant \big((1+\delta)\,\mathbb{E}[X]\big)^{\underline{t}}\right) \leqslant \frac{M_t(X)}{\big((1+\delta)\,\mathbb{E}[X]\big)^{\underline{t}}}$$

for every positive integer $t \leqslant (1+\delta)\,\mathbb{E}[X]$. This implies that, for every such t,

$$-\log \mathbb{P}\big(X \geqslant (1+\delta)\,\mathbb{E}[X]\big) \geqslant \log\big((1+\delta)\mu\big)^{\underline{t}} - \log M_t(X). \tag{58}$$

Let $t = \lfloor \delta\mu \rfloor$. Since $(1+\delta)\mu - t \geqslant \mu$, it follows from Lemma 8.4 that

$$\log\big((1+\delta)\mu\big)^{\underline{t}} \geqslant \log(\mu+t)^{\underline{t}} \geqslant I(t/\mu) \cdot \mu + t\log\mu.$$

On the other hand, our assumption implies that

$$\log M_t(X) \leqslant \log\big((1+\eta)\mu^t\big) = t\log\mu + \log(1+\eta) \leqslant t\log\mu + \eta.$$

Finally, since I is continuous, $I(\delta) > 0$, and $|t/\mu - \delta| \leqslant 1/\mu \leqslant \eta$, then substituting the above two inequalities into (58) yields

$$-\log\mathbb{P}\big(X \geqslant (1+\delta)\,\mathbb{E}[X]\big) \geqslant I(\delta)\cdot\mu - \varepsilon\Psi_X(\delta) = (1-\varepsilon)\Psi_X(\delta),$$

provided that η is sufficiently small (as a function of ε and δ).

For the upper bound, we will use the tilting argument, which is a standard trick of large deviation theory. For the sake of brevity, let $m_t = M_t(X)$ for every nonnegative integer t. Let $t = (\delta+\gamma)\mu$, where γ is a small positive constant that depends on ε and δ (but not on η). Since η is allowed to depend on γ and $\mu \geqslant 1/\eta$, we may assume that t is an integer. The idea is to consider a 'tilted' random variable $\tilde{X}$ defined by the relation

$$\mathbb{P}(\tilde{X} = x) = \frac{\mathbb{P}(X = x)\cdot x^{\underline{t}}}{m_t} \qquad \text{for every } x \in \mathbb{Z}.$$

The definition of m_t ensures that $\tilde{X}$ is a well-normalised random variable. In particular,

$$\mathbb{E}[g(\tilde{X})] = \frac{\mathbb{E}[g(X)\cdot X^{\underline{t}}]}{m_t} \qquad \text{for every } g\colon \mathbb{Z} \to \mathbb{Z}. \tag{59}$$

Using the identities

$$x \cdot x^{\underline{t}} = x^{\underline{t+1}} + t \cdot x^{\underline{t}} \qquad \text{and} \qquad x^2 \cdot x^{\underline{t}} = x^{\underline{t+2}} + (2t+1)\cdot x^{\underline{t+1}} + t^2 \cdot x^{\underline{t}},$$

we have

$$\mathbb{E}[\tilde{X}] = \frac{\mathbb{E}[X \cdot X^{\underline{t}}]}{m_t} = \frac{m_{t+1}}{m_t} + t$$

and

$$\mathbb{E}[\tilde{X}^2] = \frac{\mathbb{E}[X^2 \cdot X^{\underline{t}}]}{m_t} = \frac{m_{t+2} + (2t+1)m_{t+1}}{m_t} + t^2,$$

and so

$$\mathrm{Var}[\tilde{X}] = \mathbb{E}[\tilde{X}^2] - \mathbb{E}[\tilde{X}]^2 = \frac{m_{t+2}m_t - m_{t+1}^2 + m_{t+1}m_t}{m_t^2}.$$

Since $t = (\delta+\gamma)\mu$, then $(\delta-\varepsilon)\mu \leqslant t \leqslant t+2 \leqslant (\delta+\varepsilon)\mu$, provided that γ is sufficiently small as a function of δ and ε. Since the assumptions of the proposition imply that m_s is well-approximated by μ^s for each $s \in \{t, t+1, t+2\}$, then a straightforward computation yields

$$(1+\delta+\gamma/2)\mu \leqslant \mathbb{E}[\tilde{X}] \leqslant (1+\delta+3\gamma/2)\mu \qquad \text{and} \qquad \mathrm{Var}[\tilde{X}] \leqslant 10\eta\mu^2 + 2\mu,$$

provided that η is sufficiently small. Therefore, Chebyshev's inequality yields

$$\mathbb{P}\big(|\tilde{X} - (1+\delta+\gamma)\mu| \geqslant \gamma\mu\big) \leqslant \mathbb{P}\big(|\tilde{X} - \mathbb{E}[\tilde{X}]| \geqslant \gamma\mu/2\big) \leqslant \frac{4\cdot(10\eta\mu^2+2\mu)}{\gamma^2\mu^2} \leqslant \varepsilon. \tag{60}$$

Next, using (59) with the function $g(x) = \mathbb{1}\big[|x-(1+\delta+\gamma)\mu| < \gamma\mu\big] \cdot (m_t/x^{\underline{t}})$, we see that

$$\mathbb{P}\big(X \geqslant (1+\delta)\mu\big) \geqslant \mathbb{P}\big(|X-(1+\delta+\gamma)\mu| \leqslant \gamma\mu\big) = \mathbb{E}\left[g(\tilde{X})\right].$$

When $g(\tilde{X})$ is nonzero, then $\tilde{X}$ is bounded from above by $(1+\delta+2\gamma)\mu$, and thus

$$\begin{aligned}\mathbb{E}\left[g(\tilde{X})\right] &\geqslant \mathbb{P}\big(g(\tilde{X}) \neq 0\big) \cdot \frac{m_t}{(1+\delta+2\gamma)^{\underline{t}}} \\ &= \big(1 - \mathbb{P}(|\tilde{X}-(1+\delta+\gamma)\mu| \geqslant \gamma\mu)\big) \cdot \frac{m_t}{(1+\delta+2\gamma)^{\underline{t}}} \overset{(60)}{\geqslant} \frac{(1-\varepsilon)\cdot m_t}{(1+\delta+2\gamma)^{\underline{t}}}.\end{aligned}$$

Using Lemma 8.4 with $x = (1+\gamma)\mu$, we obtain

$$\log{(1+\delta+2\gamma)^{\underline{t}}} \leqslant I\big(t/(1+\gamma)\mu\big)\cdot(1+\gamma)\mu + t\log\big((1+\gamma)\mu\big) + \frac{t+1}{(1+\gamma)\mu}.$$

On the other hand, our assumptions imply that, when η is small,

$$\log m_t \geqslant t\log\mu + \log(1-\eta) \geqslant t\log\mu - 2\eta.$$

Combining the above bounds, we obtain

$$-\log\mathbb{P}\big(X \geqslant (1+\delta)\mu\big) \leqslant I\big(t/(1+\gamma)\mu\big)\cdot(1+\gamma)\mu + t\log(1+\gamma) + \frac{t+1}{(1+\gamma)\mu} + 2\eta - \log(1-\varepsilon).$$

Recalling that $t = (\delta+\gamma)\mu$, the continuity of I and the fact that $I(\delta) > 0$ imply that

$$-\log\mathbb{P}\big(X \geqslant (1+\delta)\mu\big) \leqslant I(\delta)\cdot\mu + \varepsilon\Psi_X(\delta) = (1+\varepsilon)\Psi_X(\delta),$$

provided that γ is sufficiently small and η is sufficiently small. □

8.2. **Cluster analysis.** We will deduce both Propositions 8.1 and 8.2 from Proposition 8.3 and Lemma 8.5, stated below, by analysing the component structure of certain random hypergraphs. Since the proofs turn out to be quite similar, we adopt a general point of view from the start. Suppose that $\mathcal{H}$ is a hypergraph and, given some $p \in (0,1)$, denote by $\mathcal{H}_p$ the random induced subhypergraph of $\mathcal{H}$ obtained by keeping every vertex with probability p, independently. The *dependency graph* $G_{\mathcal{H}}$ is the graph on the vertex set $E(\mathcal{H})$ whose edges are all pairs $\{\sigma_1, \sigma_2\}$ such that $\sigma_1 \cap \sigma_2 \neq \varnothing$. A *cluster* is a set $E' \subseteq E(\mathcal{H})$ that induces a connected subgraph in $G_{\mathcal{H}}$. We write $D_s(\mathcal{H}_p)$ for the number of clusters of size s whose elements are edges in $\mathcal{H}_p$.

Lemma 8.5. *For all positive real numbers c and η, there exists a positive constant K such that the following holds. Let $\mathcal{H}$ be a uniform hypergraph, let $p \in (0,1)$, and define $X = e_{\mathcal{H}_p}$ and $\mu = \mathbb{E}[X]$. Assume that $K \leqslant \mu \leqslant \sqrt{e_{\mathcal{H}}}/K$ and that $\mathbb{E}[D_s(\mathcal{H}_p)] \leqslant \exp(-Ks)$ for every integer s such that $2 \leqslant s \leqslant c\mu$. Then $|M_t(X) - \mu^t| \leqslant \eta\mu^t$ for every integer t such that $1 \leqslant t \leqslant c\mu$.*

Proof. Let $t \leqslant c\mu$ be a positive integer and let $\mathcal{H}(t)$ denote the set of all sequences of t distinct edges of $\mathcal{H}$. For each sequence $\bar{\sigma} = (\sigma_1, \ldots, \sigma_t) \in \mathcal{H}(t)$, let X_σ be the indicator random variable for the event $\sigma_1 \cup \cdots \cup \sigma_t \subseteq E(\mathcal{H}_p)$. Denote the uniformity of $\mathcal{H}$ by k. Our definitions readily imply that $X^{\underline{t}} = \sum_{\bar{\sigma}\in\mathcal{H}(t)} X_{\bar{\sigma}}$, that $|\mathcal{H}(t)| = {e_{\mathcal{H}}}^{\underline{t}}$, and that $\mathbb{E}[X_{\bar{\sigma}}] \geqslant p^{kt}$ for all $\bar{\sigma} \in \mathcal{H}(t)$. Thus

$$M_t(X) = \mathbb{E}[X^{\underline{t}}] = \sum_{\bar{\sigma}\in\mathcal{H}(t)} \mathbb{E}[X_{\bar{\sigma}}] \geqslant {e_{\mathcal{H}}}^{\underline{t}} \cdot p^{kt}.$$

Since, for every $x \geqslant t$,

$$x^{\underline{t}} = x^t \prod_{s=0}^{t-1}\left(1 - \frac{s}{x}\right) \geqslant x^t \left(1 - \sum_{s=0}^{t-1} \frac{s}{x}\right) \geqslant x^t \left(1 - \frac{t^2}{x}\right),$$

and $t \leqslant c\mu \leqslant c\sqrt{e_{\mathcal{H}}}/K \leqslant e_{\mathcal{H}}$ for sufficiently large K, then

$$M_t(X) \geqslant \left(1 - \frac{t^2}{e_{\mathcal{H}}}\right) \cdot {e_{\mathcal{H}}}^t p^{kt} = \left(1 - \frac{t^2}{e_{\mathcal{H}}}\right) \cdot \mu^t \leqslant \left(1 - \frac{c^2}{K^2}\right) \cdot \mu^t \leqslant (1-\eta) \cdot \mu^t,$$

provided that K is sufficiently large.

It remains to prove the upper bound. It will be convenient to partition the set $\mathcal{H}(t)$ of sequences according to the component structure of the subgraph of $G_{\mathcal{H}}$ induced by the elements of the sequence. More precisely, given a nonnegative integer ℓ and integers $s_1, \ldots, s_\ell$ such that $2 \leqslant s_1 \leqslant \cdots \leqslant s_\ell$, let $\mathcal{H}(t; s_1, \ldots, s_\ell)$ be the family of all $\bar{\sigma} = (\sigma_1, \ldots, \sigma_t) \in \mathcal{H}(t)$ such that the set $\{\sigma_1, \ldots, \sigma_t\}$ induces a subgraph in $G_{\mathcal{H}}$ whose ℓ nontrivial connected components (maximal clusters) have sizes $s_1, \ldots, s_\ell$, so that this graph has $t - (s_1 + \cdots + s_\ell)$ isolated vertices.[5] Observe that, for every collection $W_1, \ldots, W_\ell$ of connected subsets of vertices $G_{\mathcal{H}}$ with sizes $s_1, \ldots, s_\ell$, respectively, there are at most $t^{s_1+\cdots+s_\ell} \cdot {e_{\mathcal{H}}}^{t-(s_1+\cdots+s_\ell)}$ sequences $\bar{\sigma} = (\sigma_1, \ldots, \sigma_t) \in \mathcal{H}(t)$ such that the nontrivial connected components of $\{\sigma_1, \ldots, \sigma_t\}$ are exactly $W_1, \ldots, W_\ell$; indeed, there are at most $t^{s_1+\cdots+s_\ell}$ ways to choose the locations of the vertices in $W_1 \cup \cdots \cup W_\ell$ in a sequence of length t and, for each such choice, at most ${e_{\mathcal{H}}}^{t-(s_1+\cdots+s_\ell)}$ choices for the remaining elements of the sequence. We conclude that

$$\sum_{\bar{\sigma} \in \mathcal{H}(t; s_1, \ldots, s_\ell)} \mathbb{E}[X_{\bar{\sigma}}] \leqslant \mu^{t-(s_1+\cdots+s_\ell)} \cdot \prod_{i=1}^{\ell} \mathbb{E}\left[D_{s_i}(\mathcal{H}_p)\right] \cdot t^{s_i}$$

and, consequently, summing over all ℓ and all sequences $s_1, \ldots, s_\ell$ and using the assumed upper bound on the expectation of $D_s(\mathcal{H}_p)$, valid for each $s \leqslant t$,

$$\begin{aligned} M_t(X) &\leqslant \sum_{s=0}^{t} \mu^{t-s} \cdot \sum_{\ell \geqslant 0} \sum_{\substack{s_1+\cdots+s_\ell = s \\ 2 \leqslant s_1 \leqslant \cdots \leqslant s_\ell}} \prod_{i=1}^{\ell} \mathbb{E}\left[D_{s_i}(\mathcal{H}_p)\right] \cdot t^{s_i} \\ &\leqslant \sum_{s=0}^{t} \mu^{t-s} \cdot \sum_{\ell \geqslant 0} \sum_{\substack{s_1+\cdots+s_\ell = s \\ 2 \leqslant s_1 \leqslant \cdots \leqslant s_\ell}} \prod_{i=1}^{\ell} \exp(-Ks_i + s_i \log t) \\ &= \sum_{s=0}^{t} \mu^{t} \cdot \sum_{\ell \geqslant 0} \sum_{\substack{s_1+\cdots+s_\ell = s \\ 2 \leqslant s_1 \leqslant \cdots \leqslant s_\ell}} \exp\big(-Ks + s\log(t/\mu)\big). \end{aligned}$$

Since, for every $s \geqslant 0$, there are at most 2^s sequences $s_1, \ldots, s_\ell$ of positive integers whose sum is s (this includes the case when $s = 0$, when the only such sequence is the empty sequence), we have

$$M_t(X) \leqslant \mu^t \cdot \sum_{s=0}^{t} \exp\big(-Ks + s\log(t/\mu) + s\big).$$

Finally, since $t/\mu \leqslant c$, we may choose $K = K(c, \eta)$ so that

$$M_t(X) \leqslant \mu^t \cdot \sum_{s=0}^{t} \left(\frac{\eta}{1+\eta}\right)^s \leqslant (1+\eta)\mu^t,$$

completing the proof. $\square$

[5] This includes the case $\ell = 0$ in which $\mathcal{H}(t; \varnothing)$ corresponds to induced subgraphs of $G_{\mathcal{H}}$ all of whose connected components are isolated vertices.

8.3. **Proof of Proposition 8.1.** Let $\mathcal{H}$ be the hypergraph on the vertex set $\llbracket N \rrbracket$ whose edges are k-term arithmetic progressions in $\llbracket N \rrbracket$, so that $X = X_{N,p}^{k\text{-AP}} = e_{\mathcal{H}_p}$. Let $\mu = \mathbb{E}[X]$, let $\eta = \eta(\varepsilon, \delta)$ be the constant from the statement of Proposition 8.3, and let $K = K(\varepsilon, \delta, \eta)$ be the constant from the statement of Lemma 8.5.

For any two integers $a, b \in \llbracket N \rrbracket$ with $a < b$, there is at most one k-term arithmetic progression that starts with a and ends with b (and exactly one such progression if $b - a$ is divisible by $k-1$). Therefore, $N^2/(2k) \leqslant e_{\mathcal{H}} \leqslant N^2$ for all large enough N. In particular, since we assume that $CN^{-1} \leqslant p^{k/2} \leqslant C^{-1}N^{-1} \log N$, we find that

$$\frac{C^2}{2k} \leqslant \mu \leqslant \left(\frac{\log N}{C}\right)^2 \tag{61}$$

and thus $\max\{1/\eta, K\} \leqslant \mu \leqslant \sqrt{e_{\mathcal{H}}}/K$ whenever C is large. The claimed estimate on $-\log \mathbb{P}\big(X \geqslant (1+\delta)\,\mathbb{E}[X]\big)$ will follow from Proposition 8.3 and Lemma 8.5 once we verify that $D_s(\mathcal{H}_p)$, the number of clusters of s arithmetic progressions of length k in the set $\llbracket N \rrbracket_p$, satisfies

$$\mathbb{E}[D_s(\mathcal{H}_p)] \leqslant \exp(-Ks)$$

for every s satisfying $2 \leqslant s \leqslant (\delta + \varepsilon)\mu$.

In order to do so, let $\mathcal{D}(s, m)$ be the the set of all clusters $\{\sigma_1, \dots, \sigma_s\}$ of s arithmetic progressions of length k in $\llbracket N \rrbracket$ such that $|\sigma_1 \cup \dots \cup \sigma_s| = m$; we also let $D_{s,m}$ be the number of such clusters whose union is contained in the random set $\llbracket N \rrbracket_p$. When $s \geqslant 2$, the union of any s distinct k-term arithmetic progressions contains between $k+1$ and ks numbers, and therefore $D_{s,m} = 0$ unless $k+1 \leqslant m \leqslant ks$. Thus, we can write

$$D_s(\mathcal{H}_p) = \sum_{m=k+1}^{ks} D_{s,m}.$$

For each integer m, let a_m denote the number of m-element subsets of $\llbracket N \rrbracket$ that are the union of a single, nonempty (but possibly trivial) cluster of k-term arithmetic progressions. Since a progression is uniquely determined by its first and second element, then, for each s,

$$\mathbb{E}[D_{s,m}] \leqslant a_m p^m \binom{m^2}{s}. \tag{62}$$

Claim 8.6. *For every integer $m \geqslant 1$,*

$$a_m \leqslant N^2 \cdot (2kmN)^{\frac{m-k}{k-1}}.$$

Proof. We prove the claimed upper bound on a_m by induction on m. It is vacuously true when $m < k$, since then $a_m = 0$, or when $m = k$, as $a_k = e_{\mathcal{H}} \leqslant N^2$. Assume now that $m \geqslant k+1$ and let A be an arbitrary set counted by a_m. Since $m > k$, then A must be a union of at least two different progressions. Moreover, there are a proper subset $A' \subsetneq A$ that is a union of a (nonempty) cluster of k-term arithmetic progressions and a k-term progression σ that intersects A' such that $A = A' \cup \sigma$; note that the number of σ_i's whose union is A' may be significantly smaller than the number that was used to generate A. By construction, we have that $|A'| = |A| - k + |A' \cap \sigma| = m - k + |A' \cap \sigma|$. Since there are at most $k|A'|N \leqslant kmN$ arithmetic progressions of length k that intersect A' in exactly one element and at most $k^2|A'|^2 \leqslant k^2m^2$ progressions that intersect A' in two or more elements, then

$$a_m \leqslant kmN \cdot a_{m-k+1} + k^2m^2 \cdot (a_{m-k+2} + \dots + a_{m-1}).$$

It follows from our inductive assumption that

$$\begin{aligned} a_m &\leqslant kmN \cdot N^2 \cdot (2kmN)^{\frac{m-2k+1}{k-1}} + k^2m^2 \cdot k \cdot N^2 \cdot (2kmN)^{\frac{m-k-1}{k-1}} \\ &= N^2 \cdot \left((2kmN)^{\frac{m-k}{k-1}}/2 + k^3m^2 \cdot (2kmN)^{\frac{m-k}{k-1} - \frac{1}{k-1}}\right). \end{aligned}$$

Finally, as (61) implies that $m \leqslant ks \leqslant k(\delta+\varepsilon)\mu \leqslant k(\delta+\varepsilon)(C^{-1}\log N)^2$ and $N \geqslant C$, then

$$k^3m^2 \cdot (2kmN)^{-1/(k-1)} \leqslant 1/2,$$

provided that C is sufficiently large. This implies the claimed upper bound on a_m. □

Assume now that $k+1 \leqslant m \leqslant ks$. Since $s \leqslant (\delta+\varepsilon)\mu$, then (61) implies that m is only polylogarithmic in N; on the other hand, $p \leqslant (C^{-1}N^{-1}\log N)^{2/k}$. Since $k \geqslant 3$ and $N \geqslant C$, there is a positive constant γ that depends only on k such that

$$(2kmN)^{1/(k-1)}p \leqslant (2kmN)^{1/(k-1)} \cdot (C^{-1}N^{-1}\log N)^{2/k} \leqslant N^{-2(k+1)\gamma} \leqslant N^{-2m\gamma/(m-k)}.$$

In particular, Claim 8.6 implies that

$$a_m p^m \leqslant N^2p^k \cdot \left((2kmN)^{1/(k-1)}p\right)^{m-k} \leqslant N^2p^k \cdot N^{-2m\gamma} \leqslant N^{-m\gamma},$$

where for the last inequality we use that N^2p^k is at most polylogarithmic in N and $N \geqslant C$. Combining this bound with (62) we conclude that

$$\mathbb{E}[D_{s,m}] \leqslant N^{-m\gamma} \cdot \binom{m^2}{s} \leqslant \exp\left(-m\gamma\log N + s\log\left(\frac{em^2}{s}\right)\right).$$

Let $f\colon (0,\infty) \to (0,\infty)$ be the function defined by

$$f(x) = \exp\left(-x\gamma\log N + s\log\left(\frac{ex^2}{s}\right)\right) = \exp\big(-x\gamma\log N + 2s\log x + s\log(e/s)\big),$$

so that $\mathbb{E}[D_{s,m}] \leqslant f(m)$. Elementary calculus shows that f is maximised at $x = 2s/(\gamma\log N)$. Therefore,

$$\mathbb{E}[D_{s,m}] \leqslant f\left(\frac{2s}{\gamma\log N}\right) = \exp\left(-2s + s\log\left(\frac{4es}{\gamma^2(\log N)^2}\right)\right).$$

Since our assumptions imply that, see (61),

$$\frac{s}{(\log N)^2} \leqslant \frac{(\delta+\varepsilon)\mu}{(\log N)^2} \leqslant \frac{\delta+\varepsilon}{C^2},$$

we may conclude that $\mathbb{E}[D_{s,m}] \leqslant \exp(-(K+k)s)$, provided that C is sufficiently large. Therefore, if C is sufficiently large,

$$\mathbb{E}[D_s(\mathcal{H}_p)] = \sum_{m=k+1}^{ks} \mathbb{E}[D_{s,m}] \leqslant ks \cdot \exp(-Ks - ks) \leqslant \exp(-Ks).$$

This completes the proof.

8.4. **Proof of Proposition 8.2.** Let H be a connected, Δ-regular graph and let $\mathcal{H}$ be the hypergraph on the vertex set $\binom{[\![n]\!]}{2}$ whose edges are copies of H in K_n, so that $X = X_{n,p}^H = e_{\mathcal{H}_p}$. Let $\mu = \mathbb{E}[X]$, let $\eta = \eta(\varepsilon,\delta)$ be the constant from the statement of Proposition 8.3, and let $K = K(\varepsilon,\delta,\eta)$ be the constant from the statement of Lemma 8.5.

Since $(n/v_H)^{v_H} \leqslant \binom{n}{v_H} \leqslant e_{\mathcal{H}} \leqslant n^{v_H}$ for all large enough n, our assumption $Cn^{-1} \leqslant p^{\Delta/2} \leqslant C^{-1}n^{-1}(\log n)^{\frac{1}{v_H-2}}$ and the fact that $2e_H = \Delta v_H$ imply that

$$\left(\frac{C}{v_H}\right)^{v_H} \leqslant \mu \leqslant \frac{(\log n)^{1+\frac{2}{v_H-2}}}{C^{v_H}}, \tag{63}$$

and thus $\max\{1/\eta, K\} \leqslant \mu \leqslant \sqrt{e_{\mathcal{H}}}/K$ whenever C is sufficiently large. The claimed estimate on $-\log\mathbb{P}\big(X \geqslant (1+\delta)\,\mathbb{E}[X]\big)$ will follow from Proposition 8.3 and Lemma 8.5 once we verify that $D_s(\mathcal{H}_p)$, the number of clusters of s copies of H in the random graph $G_{n,p}$, satisfies

$$\mathbb{E}[D_s(\mathcal{H}_p)] \leqslant \exp(-Ks)$$

for every s satisfying $2 \leqslant s \leqslant (\delta+\varepsilon)\mu$.

To this end, for every $s \geqslant 1$, every $k \geqslant 1$, and every $m \geqslant 1$, let $\mathcal{D}(s,k,m)$ denote the set of all clusters $\{\sigma_1, \dots, \sigma_s\}$ of s distinct copies of H in K_n such that the graph $\sigma_1 \cup \dots \cup \sigma_s$ has k vertices (of nonzero degree) and m edges. We further let $D_{s,k,m}$ denote the number of such clusters whose union is contained in $G_{n,p}$. When $s \geqslant 2$, the union of any s distinct copies of H contains between v_H and $v_H s$ vertices and between $e_H + 1$ and $e_H s$ edges, and thus $D_{s,k,m} = 0$ unless $v_H \leqslant k \leqslant v_H s$ and $e_H + 1 \leqslant m \leqslant e_H s$. We can therefore write

$$D_s(\mathcal{H}_p) = \sum_{k=v_H}^{v_H s} \sum_{m=e_H+1}^{e_H s} D_{s,k,m}.$$

Claim 8.7. *There exists a positive constant γ such that, for every $s \geqslant 2$, every $k \geqslant v_H$, and every $m \geqslant e_H + 1$,*

$$\mathbb{E}[D_{s,k,m}] \leqslant n^{-2\gamma m} \binom{k^2}{m} \binom{(2m)^{v_H/2}}{s}.$$

Proof. We first show that, for every $s \geqslant 1$, the set $\mathcal{D}(s,k,m)$ is empty unless

$$m - e_H \geqslant \left(\frac{\Delta}{2} + \frac{1}{2v_H}\right) \cdot (k - v_H). \tag{64}$$

We prove this fact by induction on s. The case $s = 1$ holds vacuously, as the set $\mathcal{D}(1,k,m)$ is nonempty only when $k = v_H$ and $m = e_H$. Assume now that $s \geqslant 2$ and let G be the union of copies of H that form some cluster in $\mathcal{D}(s,k,m)$. By definition, G has k vertices and m edges. Furthermore, for some $s' < s$, $k' \leqslant k$ and $m' < m$, there exist a subgraph $G' \subseteq G$ and a copy σ of H in K_n that intersects (the edge set of) G' such that G' is the union of copies of H that form some cluster in $\mathcal{D}(s',k',m')$, and $G = G' \cup \sigma$. We note that s' may be strictly smaller than $s-1$. Let $J \subseteq H$ be the subgraph of H that is isomorphic to $\sigma \cap G'$, so that $m = m' + e_H - e_J$ and $k = k' + v_H - v_J$. It follows from the inductive assumption that

$$\begin{aligned} m - e_H = m' - e_J &\geqslant \left(\frac{\Delta}{2} + \frac{1}{2v_H}\right) \cdot (k' - v_H) + e_H - e_J \\ &= \left(\frac{\Delta}{2} + \frac{1}{2v_H}\right) \cdot (k - v_H) - \left(\frac{\Delta}{2} + \frac{1}{2v_H}\right) \cdot (v_H - v_J) + e_H - e_J. \end{aligned}$$

We claim that the above inequality implies (64). This is obviously true when $J = H$. If J is a proper subgraph of H, then $2e_J \leqslant \Delta v_J - 1$, since H is connected and Δ-regular, and therefore

$$e_H - e_J \geqslant \frac{\Delta v_H}{2} - \frac{\Delta v_J - 1}{2} = \left(\frac{\Delta}{2} + \frac{1}{2(v_H - v_J)}\right) \cdot (v_H - v_J) \geqslant \left(\frac{\Delta}{2} + \frac{1}{2v_H}\right) \cdot (v_H - v_J),$$

which gives (64).

To complete the proof of the claim, note that

$$\mathbb{E}[D_{s,k,m}] \leqslant n^k p^m \cdot \binom{k^2}{m} \cdot \binom{N(H,k,m)}{s},$$

where $N(H,k,m)$ denotes the largest number of copies of H in a graph with k vertices and m edges. Inequality (64) implies that

$$n^k p^m = n^{v_H} p^{e_H} \cdot n^{k - v_H} p^{m - e_H} \leqslant n^{v_H} p^{e_H} \cdot \left(n p^{\Delta/2 + 1/(2v_H)}\right)^{2v_H(m - e_H)/(\Delta v_H + 1)}.$$

Since $p^{\Delta/2} \leqslant C^{-1} n^{-1} \log n$, then $n^{v_H} p^{e_H}$ is only polylogarithmic in n and $np^{\Delta/2+1/(2v_H)} \leqslant n^{-\gamma'}$ for some positive constant γ'. As $m \geqslant e_H + 1$ and n is large, then there is a positive constant γ such that $n^k p^m \leqslant n^{-2\gamma m}$. The claimed upper bound on $\mathbb{E}[D_{s,k,m}]$ now follows from Theorem 5.7, which implies that $N(H,k,m) \leqslant (2m)^{v_H/2}$. $\square$

Claim 8.7 and the inequality $\binom{a}{b} \leqslant (ea/b)^b$ impy that

$$\begin{aligned}\mathbb{E}[D_{s,k,m}] &\leqslant \exp\left(-2\gamma m \log n + m\log(ek^2/m) + s\log\left(e(2m)^{v_H/2}/s\right)\right)\\ &\leqslant \exp\left(-\gamma m \log n + s\log\left(e(2m)^{v_H/2}/s\right)\right),\end{aligned}$$

where the second inequality as $k \leqslant v_H s \leqslant v_H(\delta+\varepsilon)\mu$ and μ is at most polylogarithmic in n, see (63). Let $f\colon (0,\infty) \to (0,\infty)$ be the function defined by

$$f(x) = \exp\left(-\gamma x \log n + s\log\left(e(2x)^{v_H/2}/s\right)\right),$$

so that $\mathbb{E}[D_{s,k,m}] \leqslant f(m)$. Elementary calculus shows that f is maximised at $x = sv_H/(2\gamma\log n)$. Therefore,

$$\mathbb{E}[D_{s,k,m}] \leqslant f\left(\frac{sv_H}{2\gamma\log n}\right) = \exp\left(-\frac{sv_H}{2} + \frac{sv_H}{2}\log\left(\frac{e^{2/v_H}v_H s^{1-2/v_H}}{\gamma\log n}\right)\right).$$

Since our assumptions imply that, see (63),

$$\frac{s^{1-2/v_H}}{\log n} \leqslant \frac{\big((\delta+\eta)\mu\big)^{1-2/v_H}}{\log n} \leqslant \frac{(\delta+\eta)^{1-2/v_H}}{C^{v_H-2}},$$

we may conclude that $\mathbb{E}[D_{s,k,m}] \leqslant \exp(-(K+v_He_H)s)$, provided that C is sufficiently large. Therefore, if C is sufficiently large,

$$\mathbb{E}[D_s(\mathcal{H}_p)] = \sum_{k=v_H}^{v_Hs}\sum_{m=e_H+1}^{e_Hs}\mathbb{E}[D_{s,k,m}] \leqslant s^2v_He_H\cdot\exp(-Ks - v_He_Hs) \leqslant \exp(-Ks).$$

This completes the proof.

9. Beyond polynomials with nonnegative coefficients

Although Theorem 3.1 applies only to the case where $X = X(Y)$ is a polynomial with nonnegative coefficients, the proof can be adapted to yield a similar result for all nonnegative functions $X\colon \{0,1\}^N \to \mathbb{R}_{\geqslant 0}$. In this case, the degree assumption in Theorem 3.1 has to be replaced by a more general assumption on the 'complexity' of X.

Given an $I \subseteq [\![N]\!]$ and a $z \in \{0,1\}^N$, we let

$$F(I,z) = \{y \in \{0,1\}^N : y_i = z_i \text{ for all } i \in I\};$$

we call sets of this from *subcubes*. If F is a subcube, then there is a unique set I such that $F = F(I,z)$, for some z. We can thus define the *codimension* of F by $\operatorname{codim} F = |I|$. Given a nonnegative function X on the hypercube, we define the *complexity* of X to be the smallest integer d for which it is possible to represent X as a linear combination with nonnegative coefficients of indicator functions of subcubes with codimension at most d. The complexity of X is well-defined, and at most N, since $X = \sum_{z\in\{0,1\}^N} X(z)\mathbb{1}_{F([\![N]\!],z)}$ is such a linear combination. Note also that the complexity of every polynomial with nonnegative coefficients and total degree d is at most d.

Assume now that Y is a random variable taking values in $\{0,1\}^N$ and that $X = X(Y)$. Given a subcube $F \subseteq \{0,1\}^N$, we write $\mathbb{E}_F[X] = \mathbb{E}[X \mid Y \in F]$ for the expectation of X conditioned on $Y \in F$. We further define $\Phi_X\colon \mathbb{R}_{\geqslant 0} \to \mathbb{R}_{\geqslant 0} \cup \{\infty\}$ by

$$\Phi_X(\delta) = \min\big\{-\log\mathbb{P}(Y\in F) : F \subseteq \{0,1\}^N \text{ is a subcube with } \mathbb{E}_F[X] \geqslant (1+\delta)\,\mathbb{E}[X]\big\}. \tag{65}$$

If X is an increasing function of Y, this definition coincides with our earlier definition of $\Phi_X(\delta)$, because then the minimum is achieved at a subcube of the form $F(I,\mathbf{1})$, where $\mathbf{1}$ is the N-dimensional all-ones vector. It is not difficult to adapt the proof of Theorem 3.1 to show the following.

Theorem 9.1. *For every positive integer d and all positive real numbers ε and δ with $\varepsilon < 1/2$, there is a positive $K = K(d, \varepsilon, \delta)$ such that the following holds. Let Y be a sequence of N independent* $\mathrm{Ber}(p)$ *random variables for some $p \in (0, 1-\varepsilon]$ and assume that $X = X(Y)$ has complexity at most d and satisfies $\Phi_X(\delta - \varepsilon) \geqslant K \log(1/p)$. Denote by $\mathcal{F}^*$ the collection of all subcubes $F \subseteq \{0,1\}^N$ satisfying*

(F1) $\mathbb{E}_F[X] \geqslant (1 + \delta - \varepsilon)\,\mathbb{E}[X]$,
(F2) $\operatorname{codim} F \leqslant K \cdot \Phi_X(\delta + \varepsilon)$*, and*
(F3) $\mathbb{E}_F[X] - \mathbb{E}_{F'}[X] \geqslant \mathbb{E}[X]/\big(K \cdot \Phi_X(\delta + \varepsilon)\big)$ *for every subcube F' containing F,*

and assume that for every $m \in \mathbb{N}$, there are at most $(1/p)^{\varepsilon m/2}$ subcubes of codimension m in $\mathcal{F}^*$. Then

$$(1 - \varepsilon)\Phi_X(\delta - \varepsilon) \leqslant -\log \mathbb{P}\big(X \geqslant (1 + \delta)\,\mathbb{E}[X]\big) \leqslant (1 + \varepsilon)\Phi_X(\delta + \varepsilon) \tag{66}$$

and, writing $\mathcal{E}^*$ for the collection of those $F \in \mathcal{E}^*$ with $-\log \mathbb{P}(Y \in F) \leqslant (1 + \varepsilon)\Phi_X(\delta + \varepsilon)$,

$$\mathbb{P}\big(X \geqslant (1 + \delta)\,\mathbb{E}[X] \text{ and } Y \notin F \text{ for all } F \in \mathcal{E}^*\big) \leqslant \varepsilon \mathbb{P}\big(X \geqslant (1 + \delta)\,\mathbb{E}[X]\big). \tag{67}$$

Theorem 9.1 can be used to study the upper tail problem for induced subgraph counts. Suppose that H is a fixed graph and $X = X_{n,p}^{H\text{-ind}}$ is the number of induced copies of H in the random graph $G_{n,p}$. Let $N = \binom{n}{2}$ and, for an arbitrary bijection $\sigma_n \colon \binom{[\![n]\!]}{2} \to [\![N]\!]$, let Y_i be the indicator random variable of the event that $\sigma_n^{-1}(i)$ is an edge in $G_{n,p}$. Then we can write

$$X = \sum_{\substack{H' \subseteq K_n \\ H' \cong H}} \prod_{e \in E(H)} Y_{\sigma_n(e)} \prod_{e \in \binom{V(H)}{2} \setminus E(H)} (1 - Y_{\sigma_n(e)}).$$

In particular, the complexity of X is bounded by e_H and Theorem 9.1 applies. On the other hand, it is clear that X is generally not monotone, so one cannot use Theorem 3.1.

10. Concluding remarks

Optimality of Theorem 1.5. The assumptions of Theorem 1.5 are not optimal in several ways. For one, we believe that the theorem should hold for all $n^{-1}(\log n)^{\frac{1}{v_H - 2}} \ll p^{\Delta/2} \ll 1$. Proving this would likely require a more precise description of the structure of cores at lower densities.

More fundamentally, the assumption that H is nonbipartite is likely unnecessary. Unfortunately, if H is bipartite and there exists a positive constant ξ such that

$$n^{-1}(\log n)^{\frac{1}{v_H - 2}} \ll p^{\Delta/2} \ll n^{-1/2 - \xi},$$

then the entropic stability condition is not satisfied when the constant ε appearing in the definition of this condition is sufficiently small.

For example, when $H = C_4$, then every copy of the complete bipartite graph K_{2,cn^2p^2} in K_n is a core, provided that c is sufficiently large (note that $n^2p^2 \ll n$ by our assumption on p). There are $\binom{n}{cn^2p^2}$ such copies in K_n and $\binom{n}{cn^2p^2}$ is larger than $(1/p)^{\Omega(n^2p^2)}$ under our assumption on p. Therefore, the entropic stability condition does not hold and we cannot use Theorem 3.1 to analyse the upper tail probability for this graph. Having said that, we do believe that the general approach underlying the proof of Theorem 3.1 can still be used to handle this case. More precisely, the upper tail probability is always bounded from above by the expected number of cores and we believe that this expectation is dominated by the probability of appearance of the smallest core.

Nonregular graphs. It is an open problem to extend Theorems 1.5 and 1.6 to nonregular graphs. It is straightforward to extend Theorem 1.6 to the more general case of strictly balanced graphs; however, note that Šileikis and Warnke [56] constructed graphs for which the conclusion of Theorem 1.6 does not hold. In the localised regime, very little is known for non-regular graphs. Recently, Šileikis and Warnke [57] determined the order of the logarithmic upper tail probability for the number of copies of the star graph $K_{1,s}$ in $G_{n,p}$.

The phase transition between the Poisson and the localised regimes. We believe that the logarithmic upper tail probabilities of the random variables considered in this paper are always determined by either the Poisson behaviour, the localised behaviour, or the coexistence of the two (in the regime where they are commensurate). More precisely, we believe that for both $X = X_{n,p}^H$ (for a connected, Δ-regular H) and $X = X_{N,p}^{k\text{-AP}}$,

$$-\log \mathbb{P}\big(X \geqslant (1+\delta)\,\mathbb{E}[X]\big) = \big(1 \pm o(1)\big) \cdot \min_{0 \leqslant \delta' \leqslant \delta} \big(\Phi_X(\delta') + \Psi_X(\delta - \delta')\big), \tag{68}$$

as long as $\mathbb{E}[X] \to \infty$ and $p \to 0$. Let $p^* = p^*(\delta, n)$ be such that $\Phi_X(\delta) = \Psi_X(\delta)$. Note that if $p \ll p^*$, then $\Psi_X(\delta) \ll \Phi_X(\delta)$ and we recover Theorems 1.4 and 1.6, whereas if $p \gg p^*$, we have $\Psi_X(\delta) \gg \Phi_X(\delta)$ and (68) implies (in some cases a stronger version of) Theorems 1.3, 1.5, and 1.7. If $p = \Theta(p^*)$, then both terms are of the same order and the conjecture allows for the upper tail to be dominated by configurations exhibiting features of both the Poisson and localised regimes.

Structural theorems for non-complete graphs. In the case when $X = X_{n,p}^H$ for a connected, Δ-regular graph H, we have neither determined the asymptotics of $\Phi_X(\delta)$ in the range $np^\Delta \to c \in (0, \infty)$ nor given a structural description of the upper tail event $\{X \geqslant (1+\delta)\,\mathbb{E}[X]\}$ for any density p. Doing the former would yield the logarithmic upper tail probability of X, via Proposition 7.1; it is likely that the value of $\Phi_X(\delta)$ is given by a mixture of the clique construction and a 'hub-like' construction in which a constant number of vertices have degrees linear in n. As for the latter, the method used to prove Theorem 5.8 can be generalised to yield an analogous statement in which K_r is replaced with an arbitrary Δ-regular graph H. Armed with such a 'stability' statement, it is relatively straightforward to show that, when $np^\Delta \to 0$, the random graph $G_{n,p}$ conditioned on the upper tail event $\{X \geqslant (1+\delta)\,\mathbb{E}[X]\}$ contains an 'almost-clique' of the 'right' size, as was the case when $H = K_r$. We were not able to prove such a structural statement in the complementary range $np^\Delta = \Omega(1)$.

Stability results for arithmetic progressions. An interesting problem is to characterise the near-minimisers of the optimisation problem for $\Phi_X(\delta)$ when $X = X_{N,p}^{k\text{-AP}}$. More precisely, we ask for a description of all subsets $I \subseteq [\![N]\!]$ that satisfy $\mathbb{E}_I[X] \geqslant (1+\delta)\,\mathbb{E}[X]$ and $|I| \leqslant (1+\varepsilon)\Phi_X(\delta+\varepsilon)$. As a consequence of Theorem 3.1 and the entropic stability of $X_{N,p}^{k\text{-AP}}$, which we established in the proof of Proposition 4.3, such a result would imply a structural characterisation of the upper tail event. Since the dominant contribution to the difference $\mathbb{E}_I[X] - \mathbb{E}[X]$ comes from k-term arithmetic progressions contained in I, this problem is equivalent to understanding the structure of sets $I \subseteq \mathbb{Z}$ that are near-maximisers of the number of k-term arithmetic progressions (among subsets of a given size). The structure of true maximisers was described by Green and Sisask [33] in the case for $k = 3$.

Decomposing the upper tail measure. Let $\bar{Y}$ be the random variable obtained by conditioning Y on the upper tail event $\{X(Y) \geqslant (1+\delta)\,\mathbb{E}[X]\}$ and let $\tilde{Y}$ be the random variable obtained by first choosing a uniformly random solution I of the optimisation problem for $\Phi_X(\delta)$ and then conditioning Y on $\prod_{i \in I} Y_i = 1$. It would be very interesting to determine necessary and sufficient conditions so that $\bar{Y}$ and $\tilde{Y}$ are close in some metric. In particular, are the assumptions of Theorem 3.1 sufficiently strong to imply this? This question is closely related to the more general problem of decomposing a Gibbs measure into a mixture of product measures. The work of Eldan and Gross [28], and, more recently, of Austin [4], gives general conditions for the existence of such a decomposition.

Moderate deviations. Throughout this paper, we have assumed that δ is a fixed, positive constant. It is interesting and natural to study the probability of $\{X \geqslant (1+\delta)\,\mathbb{E}[X]\}$ when δ is allowed to depend on N and p. In the case where $\delta\,\mathbb{E}[X]$ is of the same order as $\sqrt{\mathrm{Var}(X)}$, one can often prove a Central Limit Theorem, see [6, 7, 54]. The regime in which $\delta\,\mathbb{E}[X] \gg \sqrt{\mathrm{Var}(X)}$

but $\delta \to 0$ is often referred to as the moderate deviation regime. One expects that, under reasonable assumptions, the logarithmic upper tail probability $-\log \mathbb{P}(X \geqslant (1+\delta)\,\mathbb{E}[X])$ is of order $\min\{(\delta\,\mathbb{E}[X])^2/\operatorname{Var}(X),\, \Phi_X(\delta)\}$; this has been verified in certain cases, see [9, 32, 61]. Our methods can be adapted to the moderate deviation regime. In an upcoming work [34], we calculate the logarithmic upper tail probability for $X = X_{N,p}^{k\text{-AP}}$ for nearly all pairs (p, δ) for which localisation is believed to occur—that is, when $\Phi_X(\delta) \ll (\delta\,\mathbb{E}[X])^2/\operatorname{Var}(X)$.

Other random graph models. Upper tails for subgraph counts have been considered in random graph models other than $G_{n,p}$, such as exponential random graphs [18], random geometric graphs [19], random regular graphs [59], and (dense) uniform random graphs [24]. The framework developed here can be generalised to other (non-product) measures on the hypercube, providing a possible approach to such questions. It is likely that this requires adapting the notions of cores and entropic stability to the model.

References

[1] N. Alon. On the number of subgraphs of prescribed type of graphs with a given number of edges. *Israel J. Math.*, 38(1-2):116–130, 1981.

[2] F. Augeri. Nonlinear large deviation bounds with applications to traces of Wigner matrices and cycles counts in Erdős–Rényi graphs. `arXiv:1810.01558`.

[3] F. Augeri. A transportation approach to the mean-field approximation. `arXiv:1903.08021`.

[4] T. Austin. The structure of low-complexity Gibbs measures on product spaces. `arXiv:1810.07278`.

[5] J. Balogh, R. Morris, W. Samotij, and L. Warnke. The typical structure of sparse K_{r+1}-free graphs. *Trans. Amer. Math. Soc.*, 368(9):6439–6485, 2016.

[6] A. D. Barbour, M. Karoński, and A. Ruciński. A central limit theorem for decomposable random variables with applications to random graphs. *J. Combin. Theory Ser. B*, 47(2):125–145, 1989.

[7] Y. Barhoumi-Andréani, C. Koch, and H. Liu. Bivariate fluctuations for the number of arithmetic progressions in random sets. `arXiv:1902.04176`.

[8] B. B. Bhattacharya, S. Ganguly, E. Lubetzky, and Y. Zhao. Upper tails and independence polynomials in random graphs. *Adv. Math.*, 319:313–347, 2017.

[9] B. B. Bhattacharya, S. Ganguly, X. Shao, and Y. Zhao. Upper tails for arithmetic progressions in a random set. `arXiv:1605.02994`.

[10] B. Bollobás. Threshold functions for small subgraphs. *Math. Proc. Camb. Philos. Soc.*, 90(2):197–206, 1981.

[11] B. Bollobás. *Random graphs*. Cambridge University Press, Cambridge, 2nd edition, 2001.

[12] J. A. Bondy and U. S. R. Murty. *Graph theory*, volume 244 of *Graduate Texts in Mathematics*. Springer, New York, 2008.

[13] S. Boucheron, G. Lugosi, and P. Massart. *Concentration inequalities*. Oxford University Press, Oxford, 2013.

[14] J. Briët and S. Gopi. Gaussian width bounds with applications to arithmetic progressions in random settings. `arXiv:1711.05624`.

[15] S. Chatterjee. The missing log in large deviations for triangle counts. *Random Structures Algorithms*, 40(4):437–451, 2012.

[16] S. Chatterjee. *Large deviations for random graphs*, volume 2197 of *Lecture Notes in Mathematics*. Springer, Cham, 2017. Lecture notes from the 45th Probability Summer School held in Saint-Flour (École d'Été de Probabilités de Saint-Flour), June 2015.

[17] S. Chatterjee and A. Dembo. Nonlinear large deviations. *Adv. Math.*, 299:396–450, 2016.

[18] S. Chatterjee and P. Diaconis. Estimating and understanding exponential random graph models. *Ann. Statist.*, 41(5):2428–2461, 2013.

[19] S. Chatterjee and M. Harel. Localization in random geometric graphs with too many edges. `arXiv:1401.7577`.

[20] S. Chatterjee and S. R. S. Varadhan. The large deviation principle for the Erdős–Rényi random graph. *European J. Combin.*, 32(7):1000–1017, 2011.

[21] F. R. K. Chung, R. L. Graham, P. Frankl, and J. B. Shearer. Some intersection theorems for ordered sets and graphs. *J. Combin. Theory Ser. A*, 43:23–37, 1986.

[22] N. A. Cook and A. Dembo. Large deviations of subgraph counts for sparse Erdős–Rényi graphs. `arXiv:1809.11148`.

[23] R. DeMarco and J. Kahn. Tight upper tail bounds for cliques. *Random Structures Algorithms*, 41(4):469–487, 2012.

[24] A. Dembo, E. Lubetzky, et al. A large deviation principle for the Erdős–Rényi uniform random graph. *Electron. Commun. Probab.*, 23, 2018.

[25] A. Dembo and O. Zeitouni. *Large deviations techniques and applications*, volume 38 of *Applications of Mathematics*. Springer, New York, 2nd edition, 1998.

[26] R. Diestel. *Graph theory*, volume 173 of *Graduate Texts in Mathematics*. Springer, Berlin, 5th edition, 2017.

[27] R. Eldan. Gaussian-width gradient complexity, reverse log-Sobolev inequalities and nonlinear large deviations. `arXiv:1612.04346`.

[28] R. Eldan and R. Gross. Decomposition of mean-field Gibbs distributions into product measures. *Electron. J. Probab.*, 23, 2018.

[29] P. Erdős. On the number of complete subgraphs contained in certain graphs. *Magyar Tud. Akad. Mat. Kutató Int. Közl.*, 7:459–464, 1962.

[30] E. Friedgut and J. Kahn. On the number of copies of one hypergraph in another. *Israel J. Math.*, 105:251–256, 1998.

[31] D. Galvin. Three tutorial lectures on entropy and counting. `arXiv:1406.7872`.

[32] C. Goldschmidt, S. Griffiths, and A. Scott. Moderate deviations of subgraph counts in the Erdős–Rényi random graphs $G(n,m)$ and $G(n,p)$. `arXiv:1902.06830`.

[33] B. Green and O. Sisask. On the maximal number of 3-term arithmetic progressions in subsets of $\mathbb{Z}/p\mathbb{Z}$. *Bulletin London Math. Soc.*, 40(6):945–955, 2008.

[34] M. Harel, F. Mousset, and W. Samotij. Upper tails for arithmetic progressions in the moderate deviation regime. Manuscript.

[35] T. E. Harris. A lower bound for the critical probability in a certain percolation process. *Proc. Cambridge Philos. Soc.*, 56:13–20, 1960.

[36] W. Hoeffding. Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.*, 58(301):13–30, 1963.

[37] S. Janson. Poisson approximation for large deviations. *Random Structures Algorithms*, 1(2):221–229, 1990.

[38] S. Janson, K. Oleszkiewicz, and A. Ruciński. Upper tails for subgraph counts in random graphs. *Israel J. Math.*, 142:61–92, 2004.

[39] S. Janson and A. Ruciński. The infamous upper tail. *Random Structures Algorithms*, 20(3):317–342, 2002.

[40] S. Janson and A. Ruciński. The deletion method for upper tail estimates. *Combinatorica*, 24(4):615–640, 2004.

[41] S. Janson and A. Ruciński. Upper tails for counting objects in randomly induced subhypergraphs and rooted random graphs. *Ark. Mat.*, 49(1):79–96, 2011.

[42] S. Janson and L. Warnke. The lower tail: Poisson approximation revisited. *Random Structures Algorithms*, 48(2):219–246, 2016.

[43] M. Karoński and A. Ruciński. On the number of strictly balanced subgraphs of a random graph. In *Graph Theory, Lágow 1981*, volume 1018 of *Lecture Notes in Mathematics*, pages 79–83. Springer, 1983.

[44] G. Katona. A theorem of finite sets. In *Proc. Colloq. Theory of Graphs, Tihany 1966*, pages 187–207. Academic Press, New York, 1966.

[45] P. Keevash. Shadows and intersections: Stability and new proofs. *Adv. Math.*, 218:1685–1703, 2008.

[46] J. H. Kim and V. H. Vu. Concentration of multivariate polynomials and its applications. *Combinatorica*, 20(3):417–434, 2000.

[47] J. H. Kim and V. H. Vu. Divide and conquer martingales and the number of triangles in a random graph. *Random Structures Algorithms*, 24(2):166–174, 2004.

[48] G. Kozma and W. Samotij. Lower tails via relative entropy. Manuscript.

[49] J. B. Kruskal. The number of simplices in a complex. In *Mathematical Optimization Techniques*, volume 10, pages 251–278. Univ. of California Press, Berkeley, CA, 1963.

[50] E. Lubetzky and Y. Zhao. On replica symmetry of large deviations in random graphs. *Random Structures Algorithms*, 47(1):109–146, 2015.

[51] E. Lubetzky and Y. Zhao. On the variational problem for upper tails in sparse random graphs. *Random Structures Algorithms*, 50(3):420–436, 2017.

[52] F. Mousset, A. Noever, K. Panagiotou, and W. Samotij. On the probability of nonexistence in binomial subsets. `arXiv:1711.06216`.

[53] I. Rivin. Counting cycles and finite dimensional L^p norms. *Adv. in Appl. Math.*, 29(4):647–662, 2002.

[54] A. Ruciński. When are small subgraphs of a random graph normally distributed? *Probab. Theory Related Fields*, 78(1):1–10, 1988.

[55] M. Šileikis. On the upper tail of counts of strictly balanced subgraphs. *Electron. J. Combin.*, 19(1):4, 2012.

[56] M. Šileikis and L. Warnke. A counterexample to the DeMarco–Kahn upper tail conjecture. `arXiv:1809.09595`.

[57] M. Šileikis and L. Warnke. Upper tail bounds for stars. `arXiv:1901.10637`.

[58] M. Talagrand. Concentration of measure and isoperimetric inequalities in product spaces. *Publ. Math. Inst. Hautes Études Sci.*, 81(1):73–205, 1995.

[59] P. van der Hoorn, G. Lippner, and E. Mossel. Regular graphs with linearly many triangles. `arXiv:1904.02212`.

[60] V. H. Vu. Concentration of non-Lipschitz functions and applications. *Random Structures Algorithms*, 20(3):262–316, 2002.

[61] L. Warnke. Upper tails for arithmetic progressions in random subsets. *Israel J. Math.*, 221(1):317–365, 2017.

[62] Y. Zhao. On the lower tail variational problem for random graphs. *Combin. Probab. Comput.*, 26(2):301–320, 2017.

Matan Harel, School of Mathematical Sciences, Tel Aviv University, Tel Aviv 6997801, Israel
Email address: `matanharel8@tauex.tau.ac.il`

Frank Mousset, School of Mathematical Sciences, Tel Aviv University, Tel Aviv 6997801, Israel
Email address: `moussetfrank@gmail.com`

Wojciech Samotij, School of Mathematical Sciences, Tel Aviv University, Tel Aviv 6997801, Israel
Email address: `samotij@tauex.tau.ac.il`