Upper Tail Bounds for Stars

Matas Šileikis[*]

Institute of Computer Science of the Czech Academy of Sciences
182 07 Prague, Czech Republic

matas.sileikis@gmail.com

Lutz Warnke[†]

School of Mathematics
Georgia Institute of Technology
Atlanta GA 30332, U.S.A.

warnke@math.gatech.edu

Abstract

For $r \geqslant 2$, let X be the number of r-armed stars $K_{1,r}$ in the binomial random graph $G_{n,p}$. We study the upper tail $\mathbb{P}(X \geqslant (1+\varepsilon)\mathbb{E}X)$, and establish exponential bounds which are best possible up to constant factors in the exponent (for the special case of stars $K_{1,r}$ this solves a problem of Janson and Ruciński, and confirms a conjecture by DeMarco and Kahn). In contrast to the widely accepted standard for the upper tail problem, we do not restrict our attention to constant ε, but also allow for $\varepsilon \geqslant n^{-\alpha}$ deviations.

Mathematics Subject Classifications: 05C80, 60C05, 60F10

1 Introduction

The study of (the distribution of) small subgraphs in the binomial random graph $G_{n,p}$ is one of the most fundamental and influential problems in the theory of random graphs. Starting with the seminal work of Erdős and Rényi [11] from 1960, the early results for the number X_H of copies of H in $G_{n,p}$ concerned the threshold of appearance (i.e.,

when $\mathbb{P}(X_H > 0) \to 1$) and the range of edge-probabilities p for which X_H is asymptotically normal; these basic features were eventually resolved in the 1980s by Bollobás [2] and Ruciński [25]. Later the focus changed to finer details of the distribution of X_H, and the *lower tail* $\mathbb{P}(X_H \leqslant (1 - \varepsilon)\mathbb{E}X_H)$ was studied intensively in the late 1980s (often for the special case $\varepsilon = 1$). This led to the discovery of Janson's inequality [13, 14, 24], which gives exponential bounds for $\mathbb{P}(X_H \leqslant (1 - \varepsilon)\mathbb{E}X_H)$ that are best possible up to constant factors in the exponent (cf. the recent work of Janson and Warnke [20]).

Since the early 1990s the 'infamous' *upper tail* $\mathbb{P}(X_H \geqslant (1 + \varepsilon)\mathbb{E}X_H)$ has remained an important challenge, providing a well-known testbed for concentration inequalities (see, e.g., [16]). After polynomial bounds around 1990 by Spencer [29] and exponential bounds in the late 1990s via the Kim–Vu polynomial concentration method [21, 30], in 2002 Janson, Oleszkiewicz and Ruciński [17] obtained a breakthrough: via a moment based method they obtained exponential estimates for $\mathbb{P}(X_H \geqslant (1 + \varepsilon)\mathbb{E}X_H)$ that, for constant ε, are best possible up to logarithmic factors in the exponent (see also [9, 19] for extensions to random hypergraphs, and arithmetic progressions in random subsets of integers). The *upper tail problem* of closing the aforementioned logarithmic gap has remained open during the last decade, and only recently this has been settled for cliques K_r by DeMarco and Kahn [7, 8] (see also Chatterjee [3] for $r = 3$), and for arithmetic progressions by Warnke [31]. Modern large deviation theory also gives partial results [4, 23, 1, 6] for large edge-probabilities of form $p = \Theta(1)$ or $p \geqslant n^{-\delta_H}$.

In this paper we solve the upper tail problem for r-armed stars $K_{1,r}$, and as a conceptual novelty we will also allow ε to depend on n (i.e., do not restrict our attention to constant ε, as usual). The casual reader might suspect that tail estimates for r-armed stars are essentially trivial, but this is only true for $r = 1$ (where $X_{K_{1,1}} = |E(G_{n,p})|$ since $K_{1,1} = K_2$). To put this into context, Janson, Oleszkiewicz and Ruciński [17] proved that for r-regular graphs H, such as cliques K_{r+1}, the upper tail satisfies

$$p^{O_{H,\varepsilon}(n^2 p^r)} \leqslant \mathbb{P}(X_H \geqslant (1 + \varepsilon)\mathbb{E}X_H) \leqslant e^{-\Omega_{H,\varepsilon}(n^2 p^r)},$$

where the subscripts in $O_{H,\varepsilon}$ and $\Omega_{H,\varepsilon}$ indicate that the implicit constants may depend on H and ε. They also highlighted $K_{1,r}$ (with $r \geqslant 2$) as key example where the form of the exponent is more complicated, i.e, has different expressions for different ranges of p. This surprising intricacy is further manifested by the history of the infamous upper tail problem. Namely, Vu [30] argued in 2000 that his general results were essentially unimprovable due to r-armed stars, for which he obtained bounds of the form

$$p^{O_{r,\varepsilon}(n^{1+1/r} p)} \leqslant \mathbb{P}(X_{K_{1,r}} \geqslant (1 + \varepsilon)\mathbb{E}X_{K_{1,r}}) \leqslant e^{-\Omega_{r,\varepsilon}(n^{1+1/r} p)}. \tag{1}$$

However, Janson, Oleszkiewicz and Ruciński [17] later discovered that the upper tail behaviour of $K_{1,r}$ is more delicate (the lower bound in (1) is not always correct), and obtained bounds of the more involved form

$$p^{O_{r,\varepsilon}(\max\{n^{1+1/r} p, n^2 p^r\})} \leqslant \mathbb{P}(X_{K_{1,r}} \geqslant (1 + \varepsilon)\mathbb{E}X_{K_{1,r}}) \leqslant e^{-\Omega_{r,\varepsilon}(\max\{n^{1+1/r} p, n^2 p^r\})}. \tag{2}$$

In words, for stars the form of the upper tail changes around $p \approx n^{-1/r}$, which is an intriguing phenomenon (that does not occur for cliques). To lend some intuition, the lower

bound in (1) simply comes from the probability that $G_{n,p}$ contains a large star $K_{1,m}$ that itself enforces at least $\binom{m}{r} \geqslant (1+\varepsilon)\mathbb{E}X_{K_{1,r}}$ copies of $K_{1,r}$: for suitable $m = \Theta_{r,\varepsilon}(n^{1+1/r}p)$ this gives a lower bound of form $p^m \geqslant p^{O_{r,\varepsilon}(n^{1+1/r}p)}$, which tacitly requires $p = O_\varepsilon(n^{-1/r})$ to ensure that $m + 1 \leqslant n$. For larger edge-probabilities p, the refined lower bound in (2) comes from the probability that $G_{n,p}$ contains a complete bipartite graph $K_{y,z}$ that itself enforces at least $y\binom{z}{r} \geqslant (1+\varepsilon)\mathbb{E}X_{K_{1,r}}$ copies of $K_{1,r}$: for suitable $y = \Theta_\varepsilon(np^r)$ and $z = n - y$ this gives a lower bound of form $p^{yz} \geqslant p^{O_{r,\varepsilon}(n^2 p^r)}$, cf. Lemma 14. In fact, a recent conjecture by DeMarco and Kahn [8, 28] for general H asserts that in (2) the lower bound is sharp, except that the form of the upper tail changes once more for sufficiently small p: motivated by Poisson approximation heuristics, then the 'correct' exponent becomes $\mu := \mathbb{E}X_{K_{1,r}} = \Theta_r(n^{r+1}p^r)$, see (3). Here the corresponding lower bound of form $e^{-\Theta_{r,\varepsilon}(\mu)}$ comes from the event that $G_{n,p}$ contains $(1+\varepsilon)\mathbb{E}X_{K_{1,r}}$ disjoint copies of $K_{1,r}$, cf. Lemma 15. However, despite some partial results [26, 27, 33] for stars, the quest for matching bounds in (1)–(2) remained open.

1.1 Main results

Our first basic result settles the upper tail problem of r-armed stars for constant ε, by closing the existing $\log(1/p)$ gap in the exponent of (1) and (2) for *all* $p \in (0, 1]$. In particular, (3) below confirms[1] Conjecture 10.1 of DeMarco and Kahn [8] in the special case $H = K_{1,r}$. For subgraph counts this is the first example of a sharp upper tail estimate where, for constant ε, the form of $-\log \mathbb{P}(X \geqslant (1+\varepsilon)\mu)$ undergoes multiple changes, i.e., has more than two different expressions for different ranges of p.

Theorem 1 (Upper tail problem for constant ε). *Given $r \geqslant 2$, let $X = X_{r,n,p}$ be the number of copies of $K_{1,r}$ in $G_{n,p}$. Set $\mu := \mathbb{E}X$. For $p \in (0,1]$ and $\varepsilon > 0$ satisfying $1 \leqslant (1+\varepsilon)\mu \leqslant X_{r,n,1}$ we have*

$$-\log \mathbb{P}(X \geqslant (1+\varepsilon)\mu) = \Theta_{r,\varepsilon}(\Phi) \quad with \quad \Phi := \min\Big\{\mu, \max\big\{\mu^{1/r}, \mu/n^{r-1}\big\}\log(1/p)\Big\}. \quad (3)$$

Note that the assumption $(1+\varepsilon)\mu \leqslant X_{r,n,1}$ is necessary (since $X > X_{r,n,1}$ is impossible), and that the assumption $(1+\varepsilon)\mu \geqslant 1$ is natural (since otherwise $\mathbb{P}(X \geqslant (1+\varepsilon)\mu) = \mathbb{P}(X \geqslant 1) = 1 - \mathbb{P}(X = 0)$ holds). As discussed above, the form of the exponent Φ in (3) suggests that two different mechanisms explain the upper tail behaviour for constant ε. Namely, for small p we have $\Phi = \mu$, in which case the upper tail is controlled by Poisson behaviour. For larger p we have $\Phi = \max\big\{\mu^{1/r}, \mu/n^{r-1}\big\}\log(1/p)$, in which case the upper tail is governed by 'clustered' behaviour, i.e., that $G_{n,p}$ contains $\Theta_{r,\varepsilon}(\max\big\{\mu^{1/r}, \mu/n^{r-1}\big\})$ appropriately clustered edges that create $(1+\varepsilon)\mu$ copies of $K_{1,r}$ (via containment of a suitable complete bipartite graph).

---

[1]Using Corollary 1.8 in [17] and the discussion of Remark 8.3 in [17] it is not difficult to check that the special case $H = K_{1,r}$ of Conjecture 10.1 in [8] indeed reduces to (3) with $\varepsilon = 1$; see also equation (4.27) in [27] and Remark 2 in [26].

Our second result determines the correct dependence of the stars upper tail on ε, up to constant factors in the exponent (unlike in Theorem 1 above, where the dependence on ε may differ between upper and lower bounds). In particular, (4) below solves Problem 6.1 of Janson and Ruciński [18] in the special case $H = K_{1,r}$. For subgraph counts this is the first example where, for p bounded away from one, the *asymptotic order* of $-\log \mathbb{P}(X \geqslant (1 + \varepsilon)\mu)$ is determined for non-constant $\varepsilon = \varepsilon(n)$ of form $\varepsilon \geqslant n^{-\alpha}$ (the assumption $\Phi(\varepsilon) \geqslant 1$ is natural, since it ensures that we are dealing with exponentially small probabilities).

Theorem 2 (Upper tail problem for $\varepsilon = \varepsilon(n) \geqslant n^{-\alpha}$). *Given $r \geqslant 2$, let $X = X_{r,n,p}$ be the number of copies of $K_{1,r}$ in $G_{n,p}$. Set $\mu := \mathbb{E}X$, $\sigma^2 := \operatorname{Var}X$, and $\varphi(x) := (1 + x)\log(1 + x) - x$. Given $\xi \in (0, 1)$ there is $\alpha = \alpha(r) > 0$ such that, for $p \in (0, 1 - \xi]$ and $\varepsilon \geqslant n^{-\alpha}$ satisfying $\Phi(\varepsilon) \geqslant 1$ and $1 \leqslant (1 + \varepsilon)\mu \leqslant X_{r,n,1}$, we have*

$$-\log \mathbb{P}(X \geqslant (1 + \varepsilon)\mu) = \Theta_{r,\xi}\big(\Phi(\varepsilon)\big), \tag{4}$$

with

$$\Phi(\varepsilon) := \min\Big\{\varphi(\varepsilon)\mu^2/\sigma^2, \ \max\big\{(\varepsilon\mu)^{1/r}, (\varepsilon\mu)/n^{r-1}\big\}\log(e/p)\Big\}. \tag{5}$$

Remark 3. The variance satisfies $\sigma^2 = \Theta_r((1 - p)\mu(1 + (np)^{r-1}))$; see, e.g., Lemma 3.5 in [15]. Furthermore, if $\mu^{1-1/r} \geqslant \log n$ holds, then in (5) we can replace $\varphi(\varepsilon)\mu^2/\sigma^2$ by $(\varepsilon\mu)^2/\sigma^2$; see Lemma 12.

Conjecture 4 (Correct upper tail behaviour). Theorem 2 remains valid without the assumption $\varepsilon \geqslant n^{-\alpha}$.

We now lend some intuition for the form of the exponent $\Phi(\varepsilon)$. The second term in (5) comes from *local* 'clustered' behaviour, and it refines the term $\max\{\mu^{1/r}, \mu/n^{r-1}\}\log(1/p)$ in (3). The crucial difference is that here not all $(1 + \varepsilon)\mu$ stars are enforced via a bipartite construction, but only the 'excess' number $\Theta(\varepsilon\mu)$ of stars (relying on the intuitive fact that, say, at least $(1 - \varepsilon)\mu$ copies of $K_{1,r}$ are typically present in $G_{n,p}$), cf. Lemma 16. It turns out that the first term $\varphi(\varepsilon)\mu^2/\sigma^2$ in (5) comes from two related *global* mechanisms. For small p this term is of order $\varphi(\varepsilon)\mu$, refining the 'disjoint copies' based term μ in (3), cf. Lemma 17 (the well-known function φ is the large deviation rate function of the Poisson distribution). For larger p, normal approximation heuristics[2] suggest that $\mathbb{P}(X \geqslant (1 + \varepsilon)\mu) = \mathbb{P}((X - \mu)/\sigma \geqslant \varepsilon\mu/\sigma) \approx e^{-\Theta_r((\varepsilon\mu)^2/\sigma^2)}$ for very small ε, and this sub-Gaussian tail is consistent with the $\varphi(\varepsilon)\mu^2/\sigma^2$ term in (5) since $\varphi(\varepsilon) = \Theta(\varepsilon^2)$ as $\varepsilon \to 0$ (larger values of ε do not concern us here, since then the term $\varphi(\varepsilon)\mu^2/\sigma^2$ does not attain the minimum in (5) for the relevant p). Here the corresponding lower bound of form $e^{-\Theta_r(\varphi(\varepsilon)\mu^2/\sigma^2)}$ comes from the event that $G_{n,p}$ contains at least $(1 + \varepsilon)\binom{n}{2}p$ edges, cf. Lemma 18. Finally, Conjecture 4 intuitively predicts that the form of the upper tail is indeed determined by the more likely of the above-described different 'local' and 'global' mechanisms, and Theorem 2 confirms this prediction for $\varepsilon = \varepsilon(n) \geqslant n^{-\alpha}$.

---

[2]The same normal heuristic suggests that in (3) we should perhaps have used μ^2/σ^2 instead of μ, but it turns out that then the μ^2/σ^2 term would only matter for Φ (i.e., determine the minimum) in a range of p where $\mu^2/\sigma^2 = \Theta_r(\mu)$ holds.

Our third result approaches the upper tail problem from a conceptually slightly different perspective, studying $\mathbb{P}(X \geqslant \mu + t)$ for general deviations t (this contrasts Theorem 1 and 2 above, where we focus on the large deviations range $t = \varepsilon\mu$ and then put restrictions on ε). For subgraph counts, inequality (6) below is the first example where, for moderately large edge-probabilities p, the order of $-\log\mathbb{P}(X \geqslant \mu + t)$ is *completely resolved* for all deviations exceeding the standard deviation $\sigma = \sqrt{\operatorname{Var} X}$. We complement this result with inequality (7) below, which is the first example where the asymptotic order of $-\log\mathbb{P}(X \geqslant \mu + t)$ is resolved for nearly all deviations t where the 'clustered' behaviour determines the exponent (in which case $t^2/\sigma^2 \geqslant M(t)\log(e/p)$ is the natural target assumption for $\mu^{1-1/r} \geqslant \log n$ by Remark 3 and Conjecture 4).

Theorem 5 (General upper tail bounds: general deviations and clustered regime). *Given* $r \geqslant 2$, *let* $X = X_{r,n,p}$ *be the number of copies of* $K_{1,r}$ *in* $G_{n,p}$. *Set* $\mu := \mathbb{E}X$ *and* $\sigma^2 := \operatorname{Var} X$, *as well as*

$$M(t) := \max\{t^{1/r}, t/n^{r-1}\} \quad and \quad \Psi(t) := \min\left\{t^2/\sigma^2,\ M(t)\log(e/p)\right\}.$$

Given $\xi \in (0,1)$, *then the following holds whenever* $p \in (0, 1-\xi]$ *and* $1 \leqslant \mu + t \leqslant X_{r,n,1}$.

(i) *If* $p \geqslant (\log n)/n$ *and* $t \geqslant \sigma$, *then*

$$-\log\mathbb{P}(X \geqslant \mu + t) = \Theta_{r,\xi}\big(\Psi(t)\big). \tag{6}$$

(ii) *If* $\mu \geqslant \xi$ *and* $t > 0$ *satisfies* $t^2/\sigma^2 \geqslant M(t)\log(e/p) \cdot (\log n)^{2r}$, *then*

$$-\log\mathbb{P}(X \geqslant \mu + t) = \Theta_{r,\xi}\big(M(t)\log(e/p)\big). \tag{7}$$

By Remark 3, inequalities (6)–(7) provide further evidence for Conjecture 4, verifying it for $p \geqslant (\log n)/n$ among other things.

1.2 Some comments

The main focus of this paper are upper bounds on the upper tail $\mathbb{P}(X \geqslant (1+\varepsilon)\mu)$ of the number of r-armed stars in $G_{n,p}$. Developing [31, 33], here our high-level proof strategy is to introduce a combinatorial event $\mathcal{T}$ with the following property: when $\mathcal{T}$ holds, then we can find a subgraph $G_0 \subseteq G_{n,p}$ with approximately the same number of stars $K_{1,r}$ as $G_{n,p}$ and 'not too large' maximum degree. Heuristically speaking, the upper tail event $X \geqslant (1+\varepsilon)\mu$ then implies that either (i) the event $\mathcal{T}$ fails, or (ii) the aforementioned well-behaved subgraph $G_0 \subseteq G_{n,p}$ contains least $(1+\varepsilon/2)\mu$ many r-stars, say. As we shall see, the probability of having such a subgraph $G_0 \subseteq G_{n,p}$ as in (ii) can easily be controlled via a Chernoff-like tail bound of Warnke [31], and $\mathbb{P}(\neg\mathcal{T})$ can be controlled via somewhat technical union bound arguments. This together eventually gives the desired upper bounds on the upper tail $\mathbb{P}(X \geqslant (1+\varepsilon)\mu)$; see Section 2 for more details.

Finally, let us briefly compare our upper tail results for stars with very recent results from the large deviation theory literature, which are spearheaded by Chatterjee, Dembo, Lubetzky, Varadhan, Zhao, and many others (see, e.g., [5, 22, 4, 23, 10, 1, 6]). For general H, these aim at determining the asymptotics of $-\log \mathbb{P}(X_H \geqslant (1+\varepsilon)\mathbb{E}X_H)$ for constant ε, but in most cases this level of precision is achieved at the cost of an extra assumption of form $p = \Theta(1)$ or $p \geqslant n^{-\delta_H}$. For stars $H = K_{1,r}$, inequality (4) from Theorem 2 is weaker than the usual large deviations goal in the sense that it only determines $-\log \mathbb{P}(X_{K_{1,r}} \geqslant (1+\varepsilon)\mathbb{E}X_{K_{1,r}})$ up to constant factors, but it is stronger in the sense that it covers a much wider range of the parameters, in particular all $p = p(n)$ of interest and also non-constant $\varepsilon = \varepsilon(n) \geqslant n^{-\alpha}$ (so that it also applies to some moderate deviations). Obtaining such tail estimates with increased ranges of applicability is useful for combinatorial applications, where one is usually 'willing to give up a little bit on the tail', in particular on the 'inessential numerical constants' in the exponent (see [30, 18]). Furthermore, estimates of form (6)–(7) from Theorem 5 are also quite satisfactory from a concentration inequality perspective. Overall, we hope that our results will stimulate more research into such estimates for other graphs H.

1.3 Organization

In Section 2 we prove the upper bounds on the upper tail from Theorem 1, 2, and 5, and discuss a simple extension. The corresponding (fairly routine) lower bounds are then established in Appendix A.

2 Upper bounds on the upper tail

In this section we establish the upper bounds on the upper tail $\mathbb{P}(X \geqslant (1+\varepsilon)\mu)$ from Theorem 1, 2, and 5. Our core argument has two strands. In the first *combinatorial part* we iteratively decrease the maximum degree of the random graph $G_{n,p} = G_J \supseteq \cdots \supseteq G_0$ by edge-deletion (the idea is to remove large stars K_{1,D_j} with $D_j \gg r$ from G_j) until the final graph G_0 has sufficiently low maximum degree, say at most D. This degree bound allows us to estimate the number of stars $K_{1,r}$ in G_0 via a well-behaved auxiliary random variable X_D. Taking into account the number of stars K_r which are removed when passing from $G_{n,p} = G_J$ to G_0, this allows us to approximate the number $X = X_{r,n,p}$ of copies of $K_{1,r}$ in $G_{n,p}$ using X_D and several further auxiliary random variables N_{D_j} (which intuitively bound the number of K_{1,D_j} in $G_{n,p}$). In the second *probabilistic part* we then estimate the upper tails of these auxiliary variables using a concentration inequality of Warnke [31] and ad-hoc union bound arguments (exploiting the careful definitions of the variables X_D and N_{D_j} given in Section 2.1). Putting things together, the core argument then proceeds roughly as follows: by the combinatorial part $X \geqslant (1+\varepsilon)\mu$ can only happen if at least one of the auxiliary variables X_D or N_{D_j} is large, and by the probabilistic part the probability of this event is at most the desired upper tail probability (for suitable choices of the degree constraint D and other parameters).

In Section 2.1 we first implement this argument in the simpler setup of Theorem 1,

and in Section 2.2 we then extend the argument to the more precise tail estimates of Theorem 2 and 5. Finally, in Section 2.3 we also briefly discuss a straightforward extension (to a certain sum of iid random variables).

2.1 Core argument for Theorem 1

We start by introducing the main random variables and events for Theorem 1 (as we shall see, their careful definitions will facilitate the interplay between the combinatorial and probabilistic parts of our argument). For $x \geqslant 0$, let X_x denote the maximum number of copies of $K_{1,r}$ in any subgraph $H \subseteq G_{n,p}$ with maximum degree at most $\lfloor x \rfloor$. For $y > 0$, let N_y denote the maximum size of any collection of edge-disjoint $K_{1,\lceil y \rceil}$ in $G_{n,p}$. For $\beta, D, t > 0$ let $\mathcal{T} = \mathcal{T}(\beta, D, t)$ denote the event that, intuitively, prevents creation of an excess number of r-stars by a few large stars, more precisely the event that

$$N_{D_j} < \frac{\beta M}{D_j} \qquad \text{for all } j \in \mathbb{N} = \{0, 1, \dots\}, \tag{8}$$

where we tacitly used the following convenient parametrization:

$$\begin{aligned} M = M(t) &:= \max\{t^{1/r},\ t/n^{r-1}\}, \\ D_j = D_j(D) &:= 2^j D. \end{aligned} \tag{9}$$

(In this subsection we shall only use $t = \varepsilon\mu$; working with general t is convenient for the later extensions.)

We now further motivate the definition of the event $\mathcal{T}$, by relating the upper bound (8) with the 'best' bipartite constructions that enforce the excess t copies of $K_{1,r}$, giving the 'clustered' lower bounds on the upper tail (see Lemma 16). In the case $t \ll n^r$, the best lower bound creates t many r-stars via a large star $K_{1,m}$ with $m = \Theta_r(t^{1/r})$. Considering $j \in \mathbb{N}$ with $m/2 < D_j \leqslant m$, after noting $M = t^{1/r}$ it then follows from (8) that there are no vertices of degree m (for β small enough), i.e., the event $\mathcal{T}$ prevents containment of the graph $K_{1,m}$. In the complementary case $t \gg n^r$, the best lower bound creates t many r-stars via a complete bipartite graph $K_{y,z}$ with suitable $y = \Theta_r(t/n^r)$ and $z = \Theta_r(n)$. Considering $j \in \mathbb{N}$ such that $z/2 < D_j \leqslant z$, after noting $M = t/n^{r-1}$ it then follows from (8) that there are at most $O(\beta M/z) < y$ edge-disjoint stars $K_{1,z}$ (for β small enough), i.e., the event $\mathcal{T}$ prevents containment of the graph $K_{y,z}$. Overall, these two considerations demonstrate that the form of (8) is natural for large j.

The following combinatorial lemma is at the heart of our argument, and it intuitively states that $X \approx X_D$ whenever the event $\mathcal{T} = \mathcal{T}(\beta, D, t)$ holds. Its proof is inspired by ideas developed in [31, 33], but contains several new ideas. For example, instead of iteratively sparsifying an auxiliary hypergraph (which encodes the edge-sets of all stars $K_{1,r}$ in $G_{n,p}$) we here iteratively sparsify the random graph $G_{n,p}$ itself. Furthermore, in order to obtain the correct tail behaviour, in inequality (8) we need to work with $M = \max\{t^{1/r}, t/n^{r-1}\}$ instead of the simpler choice $M = t^{1/r}$ suggested by [31] (we achieve this by adding an extra degree bound to the argument, bounding the initial maximum degree by $\overline{M} = \min\{M, n\}$ instead of just M).

Lemma 6. *Given $\beta \in (0, 1/32]$ and $D, t > 0$, the event $\mathcal{T}(\beta, D, t)$ implies $X_D \leqslant X \leqslant X_D + t/2$.*

The lower bound $X \geqslant X_D$ of Lemma 6 is trivial. For the upper bound the idea is to iteratively decrease the maximum degree of $G_{n,p}$, yielding $G_{n,p} = G_J \supseteq \cdots \supseteq G_0$. By bounding the number of $K_{1,r}$ which are removed when passing from G_{j+1} to G_j, this eventually allows us to estimate the total number of $K_{1,r}$.

Proof of Lemma 6. Define $\overline{M} := \min\{M, n\}$. Recalling the discussion following (9), note that $\overline{M} = t^{1/r}$ whenever $M = t^{1/r}$, and $\overline{M} = n$ otherwise. Let J be the smallest integer $J \geqslant 0$ with $D_J \geqslant \overline{M}$. We set $G_J = G_{n,p}$, and inductively construct $G_J \supseteq \cdots \supseteq G_0$ as follows. Given G_{j+1} with $0 \leqslant j \leqslant J - 1$, let $\mathcal{C}_{j+1}$ be a maximal set of edge-disjoint collection of stars $K_{1,\lceil D_j \rceil}$. We remove all edges from G_{j+1} which are incident to a centre vertex of some star in $\mathcal{C}_{j+1}$, and denote the resulting graph by G_j.

Writing $\Delta_j = \Delta(G_j)$ for the maximum degree of G_j, we claim that $\Delta_j \leqslant \lfloor D_j \rfloor$ for all $0 \leqslant j \leqslant J$. For $G_J = G_{n,p}$ we use a case distinction. If $M \geqslant n$, then trivially $\Delta_J \leqslant n = \lfloor \overline{M} \rfloor \leqslant \lfloor D_J \rfloor$. Otherwise $D_J \geqslant \overline{M} = M$, in which case (8) entails $N_{D_J} < \beta < 1$, so $G_{n,p} = G_J$ contains no $K_{1,\lceil D_J \rceil}$, and $\Delta_J \leqslant \lceil D_J \rceil - 1 \leqslant \lfloor D_J \rfloor$ follows. Further considering G_{j+1} with $0 \leqslant j \leqslant J - 1$, we note that $\Delta_j \leqslant \lceil D_j \rceil - 1 \leqslant \lfloor D_j \rfloor$ by construction, because otherwise we could add another $K_{1,\lceil D_j \rceil}$ to $\mathcal{C}_{j+1}$ (contradicting maximality).

With $G_J \supseteq \cdots \supseteq G_0$ in hand, we now count the total number of copies of $K_{1,r}$ in $G_{n,p} = G_J$. Note that, given an edge $e = \{v_1, v_2\}$ of G_{j+1} with $0 \leqslant j < J$, we can construct any $K_{1,r}$ in G_{j+1} containing e by first selecting a centre vertex $v_c \in \{v_1, v_2\}$ and then $r - 1$ additional neighbours of v_c. Hence in G_{j+1} any edge is contained in at most $2\binom{\Delta_{j+1}}{r-1} \leqslant 2^r D_j^{r-1}/(r-1)! \leqslant 4D_j^{r-1}$ copies of $K_{1,r}$. Recalling the definition of N_{D_j}, note that when, passing from G_{j+1} to G_j, we remove at most $N_{D_j} \Delta_{j+1} \leqslant 2N_{D_j} D_j$ edges. So, since G_0 contains at most $X_{D_0} = X_D$ copies of $K_{1,r}$, using (8) and $\max_{0 \leqslant j < J} D_j \leqslant \overline{M}$ it follows that

$$X \leqslant X_D + \sum_{0 \leqslant j < J} \left(2N_{D_j} D_j \cdot 4D_j^{r-1}\right) \leqslant X_D + 8\beta M \cdot \sum_{j \in \mathbb{N}: D_j \leqslant \overline{M}} D_j^{r-1}. \tag{10}$$

Recalling $D_j = 2^j D$ and $r \geqslant 2$, using $\overline{M} = \min\{M, n\}$, $M = \max\{t^{1/r}, t/n^{r-1}\}$ and $\beta \leqslant 1/32$ we infer

$$X - X_D \leqslant 8\beta M \cdot 2\overline{M}^{r-1} \leqslant 16\beta \cdot \min\{M^r, Mn^{r-1}\} \leqslant t/2, \tag{11}$$

which completes the proof. $\qquad \square$

Applying Lemma 6 with $t = \varepsilon\mu$, in the probabilistic part of the argument it remains to estimate $\mathbb{P}(X_D \geqslant \mu + \varepsilon\mu/2)$ and $\mathbb{P}(\neg\mathcal{T}(\beta, D, \varepsilon\mu))$. We shall exploit the maximum degree constraint of X_D via the following upper tail inequality of Warnke [31], which extends classical Chernoff bounds to random variables with well-behaved dependencies (and allows us to go beyond the method of typical bounded differences [32]).

Theorem 7 (Corollary of [31, Theorem 9]). *Let $(\xi_i)_{i \in \mathfrak{S}}$ be a finite family of independent random variables with $\xi_i \in \{0, 1\}$. Given a family $\mathcal{I}$ of subsets of $\mathfrak{S}$, consider random variables $Y_\alpha := \prod_{i \in \alpha} \xi_i$ with $\alpha \in \mathcal{I}$, and suppose $\sum_{\alpha \in \mathcal{I}} \mathbb{E} Y_\alpha \leqslant \mu$. Define $Z_C := \max \sum_{\alpha \in \mathcal{J}} Y_\alpha$, where the maximum is taken over all $\mathcal{J} \subseteq \mathcal{I}$ with $\max_{\beta \in \mathcal{J}} |\{\alpha \in \mathcal{J} : \alpha \cap \beta \neq \varnothing\}| \leqslant C$. Set $\varphi(x) := (1 + x) \log(1 + x) - x$. Then, for all $C, t > 0$,*

$$\mathbb{P}(Z_C \geqslant \mu + t) \leqslant \exp\left(-\frac{\varphi(t/\mu)\mu}{C}\right) \leqslant \exp\left(-\frac{t^2}{2C(\mu + t)}\right). \tag{12}$$

The main observation is that, in every subgraph $H \subseteq G_{n,p}$ with maximum degree at most D, any star $K_{1,r}$ shares edges with $O(D^{r-1})$ other stars. For X_D this allows us to *routinely* apply Theorem 7 with Lipschitz-like parameter $C = O(D^{r-1})$, making inequality (13) plausible. For Theorem 1 the crux is that our choice of D will ensure $\mu/D^{r-1} = \Theta_r(\Phi)$, so (13) suggests that $X_D \leqslant \mu + \varepsilon\mu/2$ fails with probability at most $e^{-\Omega_{r,\varepsilon}(\Phi)}$.

Corollary 8. *For all $n \geqslant 1$, $p \in (0, 1]$ and $D, t > 0$ we have*

$$\mathbb{P}(X_D \geqslant \mu + t/2) \leqslant \exp\left(-\frac{\varphi(t/\mu)\mu}{16 D^{r-1}}\right) \leqslant \exp\left(-\frac{\min\{t, t^2/\mu\}}{48 D^{r-1}}\right). \tag{13}$$

Proof. Let $\mathcal{K}_{1,r}(G)$ contain all edge-subsets of G that are isomorphic to $K_{1,r}$. Writing $Y_\alpha := \mathbb{1}_{\{\alpha \subseteq E(G_{n,p})\}}$, there is a subgraph $H \subseteq G_{n,p}$ with maximum degree at most $\lfloor D \rfloor$ such that $X_D = \sum_{\alpha \in \mathcal{J}} Y_\alpha$ for $\mathcal{J} := \mathcal{K}_{1,r}(H)$. Given $\beta \in \mathcal{J}$, we construct all edge-intersecting stars $\alpha \in \mathcal{J}$ as in the proof of Lemma 6, and infer

$$\max_{\beta \in \mathcal{J}} |\{\alpha \in \mathcal{J} : \alpha \cap \beta \neq \varnothing\}| \leqslant r \cdot 2 \binom{\lfloor D \rfloor}{r-1} \leqslant \frac{2r D^{r-1}}{(r-1)!} \leqslant 4 D^{r-1} =: C. \tag{14}$$

It follows that $X_D \leqslant Z_C$, where Z_C is defined as in Theorem 7 with $\mathcal{I} = \mathcal{K}_{1,r}(K_n)$. It is well-known (and easy to check by calculus) that for $x \geqslant 0$ we have

$$\varphi(x/2) \geqslant \varphi(x)/4 \qquad \text{and} \qquad x^2 \geqslant \varphi(x) \geqslant \min\{x, x^2\}/3. \tag{15}$$

Putting things together, using Theorem 7 and (15) it follows that

$$\mathbb{P}(X_D \geqslant \mu + t/2) \leqslant \mathbb{P}(Z_C \geqslant \mu + t/2) \leqslant \exp\left(-\frac{\varphi(t/\mu)\mu}{4C}\right) \leqslant \exp\left(-\frac{\min\{t, t^2/\mu\}}{12C}\right), \tag{16}$$

which completes the proof of (13) by choice of C (see (14) above). $\square$

We shall estimate $\mathbb{P}(\neg\mathcal{T}(\beta, D, \varepsilon\mu))$ via a union bound argument and the following upper tail estimate for N_{D_j}. The technical assumption (17) intuitively ensures that vertices with degree at least D are unlikely. For Theorem 1 the crux is that our choice of D will also ensure $np/(eD_j) \leqslant p^{\Omega(1)}$, so applications of inequality (18) with $x = \beta M/D_j$ suggest that $\mathcal{T}$ and thus (8) fails with probability at most $n \cdot n^{-3} p^{\Omega(M)} \leqslant n^{-2} e^{-\Omega_\varepsilon(\Phi)}$.

Lemma 9. *For all $n \geqslant 1$, $p \in (0, 1]$, and $D > 0$ satisfying*

$$\left(e^3 np/D\right)^D \leqslant n^{-8} \tag{17}$$

the following holds. For all $x > 0$ we have

$$\mathbb{P}(N_{D_j} \geqslant x) \leqslant \frac{1}{n^3} \left(\frac{np}{e\lceil D_j \rceil}\right)^{xD_j/2} \mathbb{1}_{\{D_j \leqslant n\}}. \tag{18}$$

Proof. As $\binom{m}{z} \leqslant (me/z)^z$ for all integers $m, z \geqslant 1$, by exploiting the disjointness condition of N_{D_j} we infer

$$\mathbb{P}(N_{D_j} \geqslant x) \leqslant n^{\lceil x \rceil} \binom{n}{\lceil D_j \rceil}^{\lceil x \rceil} p^{\lceil x \rceil \lceil D_j \rceil} \leqslant \left(n\left(\frac{enp}{\lceil D_j \rceil}\right)^{\lceil D_j \rceil}\right)^{\lceil x \rceil}. \tag{19}$$

As the function $x \mapsto (e^3 np/x)^x$ is decreasing for $x \geqslant e^2 np$, and (17) implies $\lceil D_j \rceil \geqslant D \geqslant e^3 np$, we deduce

$$\left(\frac{enp}{\lceil D_j \rceil}\right)^{\lceil D_j \rceil} = \left(\frac{e^3 np}{\lceil D_j \rceil}\right)^{\frac{\lceil D_j \rceil}{2}} \left(\frac{np}{e\lceil D_j \rceil}\right)^{\frac{\lceil D_j \rceil}{2}}$$

$$\leqslant \left(\frac{e^3 np}{D}\right)^{\frac{D}{2}} \left(\frac{np}{e\lceil D_j \rceil}\right)^{\frac{\lceil D_j \rceil}{2}} \leqslant n^{-4} \left(\frac{np}{e\lceil D_j \rceil}\right)^{\frac{\lceil D_j \rceil}{2}}.$$

Plugging this into (19) readily establishes inequality (18), since trivially $N_{D_j} = 0$ when $D_j > n$. $\qquad\square$

For the proof of the upper bound of Theorem 1 it remains to pick suitable D, i.e., which satisfies the technical assumption (17) and yields the 'correct' exponent in (13) and suitable applications of (18).

Proof of the upper bound in (3) of Theorem 1. For concreteness, define $\beta := 1/32$ and $\gamma := 1/(16r)$, as well as

$$A := \max\left\{e^4, \, 8/\gamma\right\}, \quad s := \log(e/p^\gamma), \quad \text{and} \quad D := A \cdot \max\left\{1, \, \frac{\min\{\mu^{1/r}, n\}}{s^{1/(r-1)}}\right\}.$$

For later reference, we record that there is a constant $d = d(r) > 0$ such that, for $n \geqslant n_0(r)$,

$$dn^{r+1}p^r \leqslant \mu \leqslant n^{r+1}p^r. \tag{20}$$

By Lemma 6, the upper tail of the number $X = X_{r,n,p}$ of $K_{1,r}$-copies satisfies

$$\mathbb{P}(X \geqslant (1+\varepsilon)\mu) \leqslant \mathbb{P}(X_D \geqslant \mu + \varepsilon\mu/2) + \mathbb{P}(\neg\mathcal{T}(\beta, D, \varepsilon\mu)). \tag{21}$$

Gearing up to bound $\mathbb{P}(\neg\mathcal{T}(\beta, D, \varepsilon\mu))$ via Lemma 9, using $e = p^\gamma e^s$ and inequality (20) together with the bound $s^{1/(r-1)} \leqslant s = 1 + \log p^{-\gamma} \leqslant p^{-\gamma}$ (as $1 + x \leqslant e^x$) it follows that

$$\frac{np}{eD} = \frac{np^{1-\gamma}e^{-s}}{D} \leqslant \mathbb{1}_{\{p\leqslant n^{-1/(1-\gamma)}\}}\frac{e^{-s}}{A} + \mathbb{1}_{\{p>n^{-1/(1-\gamma)}\}}\frac{e^{-s}}{A\min\{d^{1/r}n^{1/r}p^{2\gamma},\ p^{2\gamma-1}\}} \tag{22}$$
$$\leqslant \frac{e^{-s}}{A} \leqslant e^{-s},$$

where here and below we shall always tacitly assume $n \geqslant n_0(r,d)$ whenever necessary. Since the above calculation also gives $D \geqslant Anp^{1-\gamma}$, together with $D \geqslant A$ it follows that

$$\left(e^3 np/D\right)^D \leqslant (p^\gamma/e)^D \leqslant \mathbb{1}_{\{p\leqslant n^{-1}\}}n^{-A\gamma} + \mathbb{1}_{\{p>n^{-1}\}}e^{-Anp^{1-\gamma}} \leqslant n^{-8}.$$

Applying a union bound argument, using estimates (18), $D_j = 2^j D \geqslant D$, and (22) it follows that

$$\mathbb{P}(\neg\mathcal{T}(\beta, D, \varepsilon\mu)) \leqslant \sum_{j\in\mathbb{N}}\mathbb{P}(N_{D_j} \geqslant \beta M/D_j) \leqslant n\cdot\frac{1}{n^3}\left(\frac{np}{eD}\right)^{\beta M/2} \leqslant \frac{1}{n^2}\cdot e^{-\beta Ms/2}. \tag{23}$$

Recalling (21) and the definition of $M = M(\varepsilon\mu)$, by applying Corollary 8 with $t := \varepsilon\mu$ it follows that there is a constant $c = c(\beta, A, \gamma, r) > 0$ and suitable parameters $\zeta, \Pi > 0$ such that

$$\mathbb{P}(X \geqslant (1+\varepsilon)\mu) \leqslant \exp\left(-\frac{\min\{\varepsilon, \varepsilon^2\}\mu}{48D^{r-1}}\right) + \frac{1}{n^2}\exp\left(-\frac{\beta Ms}{2}\right)$$
$$\leqslant (1+n^{-2})\cdot\exp\left(-c\underbrace{\min\{\varepsilon,\ \varepsilon^2,\ \varepsilon^{1/r}\}}_{=:\zeta}\underbrace{\min\{\mu,\ \max\{\mu^{1/r},\ \mu/n^{r-1}\}s\}}_{=:\Pi}\right). \tag{24}$$

We find the above upper tail estimate very satisfactory, but in the literature it has become standard to suppress multiplicative factors such as $1+n^{-2}$ in (24), which is straightforward when $c\zeta\Pi \geqslant 1$ holds (rescaling the exponent $c\zeta\Pi$ by a factor of $1/2$, say). In the remaining case $1 > c\zeta\Pi$ Markov's inequality gives

$$\mathbb{P}(X \geqslant (1+\varepsilon)\mu) \leqslant \frac{1}{1+\varepsilon} = 1 - \frac{\varepsilon}{1+\varepsilon} \leqslant \exp\left(-\frac{\varepsilon}{1+\varepsilon}\right) \leqslant \exp\left(-\tfrac{c}{2}\min\{\varepsilon, 1\}\zeta\Pi\right).$$

Finally, noting $s = \log(e/p^\gamma) \geqslant \log(1/p^\gamma) = \gamma\log(1/p)$ then establishes the upper bound in (3). $\qquad\square$

2.2 Extension of the argument to Theorem 2 and 5

We now extend the arguments from Section 2.1 to the upper bounds of Theorem 2 and 5. To obtain sub-Gaussian decay $\varphi(\varepsilon)\mu^2/\sigma^2$ in the exponent of tail-inequality (13) for X_D, in view of the well-known variance estimate $\sigma^2 = \Theta_{r,\xi}((1+(np)^{r-1})\mu)$ from Remark 3 we here

would like to pick $D = \Theta_{r,\xi}(1 + np)$ for some range of $t = \varepsilon\mu$. However this choice causes a major problem:[3] in the key estimate (22) we can no longer win an extra log-factor (via $np/(eD) \leqslant e^{-s}$) when we bound the N_{D_j} variables using (18) from Lemma 9. Our strategy for overcoming this obstacle is to refine the event $\mathcal{T} = \mathcal{T}(\beta, D, t)$, by enforcing different upper bounds on N_{D_j} when $D_j = 2^j D$ is small (so that in the probabilistic arguments we automatically win an extra logarithmic factor, without destroying the combinatorial counting arguments from Lemma 6).

Turning to the details, for $\gamma, \beta, D, t > 0$ let $\mathcal{T}^+ = \mathcal{T}^+(\gamma, \beta, D, t)$ denote the 'refined' event that

$$N_{D_j} < \frac{\beta M s}{D_j} \qquad \text{for all } j \in \mathbb{N} \text{ with } D_j < \min\{M, n\}/s^{1/(r-1)}, \text{ and} \qquad (25)$$

$$N_{D_j} < \frac{\beta M}{D_j} \qquad \text{for all } j \in \mathbb{N} \text{ with } D_j \geqslant \min\{M, n\}/s^{1/(r-1)}, \qquad (26)$$

where, in addition to the parameters $M = \max\{t^{1/r}, t/n^{r-1}\}$ and $D_j = 2^j D$ from (9), we tacitly used

$$s = s(\gamma) := \log(e/p^\gamma). \qquad (27)$$

Lemma 10. *Given* $\beta \in (0, 1/64]$ *and* $\gamma, D, t > 0$, *the event* $\mathcal{T}^+(\gamma, \beta, D, t)$ *implies* $X_D \leqslant X \leqslant X_D + t/2$.

Proof. The proof of Lemma 6 carries over, except for the final inequalities (10)–(11) that bound X from above. Recalling that $\overline{M} = \min\{M, n\}$, by mimicking the argument leading to (10) we here obtain

$$X - X_D \leqslant \sum_{0 \leqslant j < J} \left(2 N_{D_j} D_j \cdot 4 D_j^{r-1}\right)$$

$$\leqslant 8\beta M \cdot \left(s \sum_{j \in \mathbb{N}: D_j < \overline{M}/s^{1/(r-1)}} D_j^{r-1} + \sum_{j \in \mathbb{N}: \overline{M}/s^{1/(r-1)} \leqslant D_j \leqslant \overline{M}} D_j^{r-1}\right).$$

Recalling $D_j = 2^j D$ and $r \geqslant 2$, using $\beta \leqslant 1/64$ it then follows similarly to (10)–(11) that

$$X - X_D \leqslant 8\beta M \cdot 4 \overline{M}^{r-1} \leqslant 32\beta \cdot \min\{M^r, Mn^{r-1}\} \leqslant t/2,$$

which completes the proof. $\square$

We are now ready to prove the following slightly more general upper tail estimate for the number $X = X_{r,n,p}$ of $K_{1,r}$-copies in $G_{n,p}$, which (as we shall see) implies the upper bounds in Theorems 2 and 5.

---

[3]For $D = \Theta_{r,\xi}(1 + np)$ another problem is that the technical assumption (17) of Lemma 9 then breaks when np is close to one, which partially explains why in the upcoming Theorem 11 we shall exclude fairly small t when $np \in (n^{-\gamma}, \gamma \log n)$.

Theorem 11 (Upper tail bounds: technical result). *Given $r \geqslant 2$, let $X = X_{r,n,p}$. Set $\mu := \mathbb{E}X$, $\Lambda := \mu(1 + (np)^{r-1})$, and $\varphi(x) := (1 + x)\log(1 + x) - x$. Given $\gamma > 0$, suppose that either*

$$np \notin (n^{-\gamma}, \gamma \log n) \quad \text{or} \quad t^2/\mu \geqslant \mathbb{1}_{\{t \leqslant \min\{\mu, n^r\}\}} \gamma \min\{t^{1/r}(\log n)^r, \, Ms(\log n)^{r-1}\}$$

holds, where the parameters M and s are defined in (9) and (27). Then we have

$$\mathbb{P}(X \geqslant \mu + t) \leqslant (1 + n^{-1}) \cdot \exp\left(-\Omega_{r,\gamma}\left(\min\left\{\varphi(t/\mu)\mu^2/\Lambda, \, M\log(e/p)\right\}\right)\right). \qquad (28)$$

Proof. Let $\beta := 1/64$. We distinguish the following three cases: (i) $np \geqslant \gamma \log n$, (ii) $np \leqslant n^{-\gamma}$, and (iii) $t^2/\mu \geqslant \mathbb{1}_{\{t \leqslant \min\{\mu, n^r\}\}} \gamma \min\{t^{1/r}(\log n)^r, Ms(\log n)^{r-1}\}$. Note that in all three cases we may assume $\gamma \leqslant 1/(16r)$, since decreasing γ yields a less restrictive assumption. Furthermore, in case (iii) we may also assume that $n^{-\gamma} \leqslant np \leqslant \log n$ holds (otherwise case (i) or (ii) apply). For concreteness, define

$$A := \max\left\{e^4, \, 8 \cdot (3/\gamma)^{1/(r-1)}, \, 8/\gamma\right\} \quad \text{and} \quad D := A \cdot \max\left\{1 + np, \, \left(\frac{\varphi(t/\mu)\mu}{Ms}\right)^{1/(r-1)}\right\}.$$

(We remark that in cases (i)–(ii) the simpler choice $D = A(1 + np)$ suffices.) We defer the somewhat technical proofs of the following claims regarding Lemma 9: (a) assumption (17) holds, and (b) inequality (18) implies

$$\mathbb{P}\left(\neg\mathcal{T}^{+}(\beta, \gamma, D, t)\right) \leqslant \frac{1}{n}\max\left\{e^{-\beta Ms/2}, \, e^{-\Psi}\right\} \qquad \text{with} \qquad \Psi := \varphi(t/\mu)\mu^2/\Lambda, \qquad (29)$$

where here and below we shall again tacitly assume $n \geqslant n_0(r)$. Analogously to inequalities (21) and (24), by first applying Lemma 10 and Corollary 8, and then using $(1 + np)^{r-1} = \Theta_r(\Lambda/\mu)$, it follows that

$$\mathbb{P}(X \geqslant \mu + t) \leqslant \exp\left(-\frac{\varphi(t/\mu)\mu}{16D^{r-1}}\right) + \frac{1}{n}\exp\left(-\frac{\beta}{2}\min\{\Psi, \, Ms\}\right)$$

$$\leqslant (1 + n^{-1}) \cdot \exp\left(-\Omega_{r,\gamma}\left(\min\{\Psi, \, Ms\}\right)\right).$$

Since $s = \log(e/p^\gamma) \geqslant \gamma \log(e/p)$, this establishes inequality (28).

It remains to verify claims (a) and (b) above, and start with claim (a), i.e., that the assumption (17) of Lemma 9 holds. Note that $D \geqslant A(1 + np) \geqslant e^4 np$. Furthermore, in case (i) we have $D \geqslant A\gamma \log n$, and in case (ii) we have $np \leqslant n^{-\gamma}$ and $D \geqslant A$. So, in both cases, using $A \geqslant 8/\gamma$ we infer

$$\left(e^3 np/D\right)^D \leqslant \min\{e^{-D}, (np)^D\} \leqslant n^{-A\gamma} \leqslant n^{-8}. \qquad (30)$$

Proceeding analogously, in the cumbersome case (iii) it suffices to show $D \geqslant 8\log n$. Using $\gamma \leqslant 1/(16r)$, $p \geqslant n^{-1-\gamma}$ and (20), it is routine to see that $s \leqslant \log n$ and $\mu \geqslant n^{1/2}$.

Assuming $t \geqslant \mu$, by first using (15) and then distinguishing the cases $t \geqslant n^r$ (where $M = t/n^{r-1}$) and $t \leqslant n^r$ (where $M = t^{1/r}$), it follows that

$$D^{r-1} \geqslant \frac{A^{r-1}\varphi(t/\mu)\mu}{Ms} \geqslant \frac{A^{r-1}t}{3Ms} \geqslant \frac{A^{r-1}\min\{n^{r-1},\ \mu^{1-1/r}\}}{3\log n} \geqslant (A\log n)^{r-1}. \qquad (31)$$

Assuming $t \leqslant \mu$, we note that assumption $p \leqslant (\log n)/n$ implies $\mu \leqslant n^r$ (hence $t \leqslant n^r$ and thus $M = t^{1/r}$, as noted above). Hence, by first using (15) and then the assumed lower bound on t from case (iii), we infer

$$\begin{aligned} D^{r-1} &\geqslant \frac{A^{r-1}\varphi(t/\mu)\mu}{Ms} \geqslant \frac{A^{r-1}t^2}{3\mu Ms} \\ &\geqslant \frac{\gamma A^{r-1}\min\{t^{1/r}(\log n)^r,\ Ms(\log n)^{r-1}\}}{3Ms} = \gamma/3 \cdot (A\log n)^{r-1}. \end{aligned}$$

Each time $D \geqslant 8\log n$ follows readily by definition of A, establishing claim (a), as discussed above.

Finally, we verify claim (b), i.e., that inequality (18) implies estimate (29). We start by observing that if $\mathcal{T}^+(\beta,\gamma,D,t)$ fails then a fortiori $N_{D_0} \geqslant 1$. Hence, using (18) with $x = 1$ and $D_0 = D \geqslant e^3 np$, we deduce

$$\mathbb{P}(\neg\mathcal{T}^+(\beta,\gamma,D,t)) \leqslant \mathbb{P}(N_{D_0} \geqslant 1) \leqslant \frac{1}{n^3} \cdot e^{-D}. \qquad (32)$$

Analogously to (23), using inequality (18) and $D_j = 2^j D \geqslant e^3 np$ it also follows that

$$\begin{aligned} \mathbb{P}(\neg\mathcal{T}^+(\beta,\gamma,D,\varepsilon\mu)) &\leqslant \sum_{\substack{j\in\mathbb{N}:\\ D_j\leqslant\overline{M}/s^{1/(r-1)}}} \mathbb{P}\left(N_{D_j} \geqslant \frac{\beta Ms}{D_j}\right) + \sum_{\substack{j\in\mathbb{N}:\\ D_j\geqslant\overline{M}/s^{1/(r-1)}}} \mathbb{P}\left(N_{D_j} \geqslant \frac{\beta M}{D_j}\right) \\ &\leqslant \frac{1}{n^2} \cdot e^{-\beta Ms} + \frac{1}{n^2} \cdot \left(\frac{np}{e\lceil\overline{M}/s^{1/(r-1)}\rceil}\right)^{\beta M/2}. \end{aligned}$$
$$(33)$$

We now use a fairly technical case distinction to verify that the two estimates (32)–(33) together imply (29). Assuming $\overline{M} \geqslant np^{1-2\gamma}$, analogously to the proof of (22) we have $nps/\overline{M} \leqslant p^{2\gamma}s \leqslant p^\gamma = e^{1-s}$, so that

$$\left(\frac{np}{e\lceil\overline{M}/s^{1/(r-1)}\rceil}\right)^{\beta M/2} \leqslant \left(\frac{nps}{e\overline{M}}\right)^{\beta M/2} \leqslant e^{-\beta Ms/2} \qquad \text{when } \overline{M} \geqslant np^{1-2\gamma}. \qquad (34)$$

Next we assume $p \leqslant n^{-1/(1-\gamma)}$, which implies $np/e \leqslant p^\gamma/e = e^{-s}$, so that

$$\left(\frac{np}{e\lceil\overline{M}/s^{1/(r-1)}\rceil}\right)^{\beta M/2} \leqslant \left(\frac{np}{e}\right)^{\beta M/2} \leqslant e^{-\beta Ms/2} \qquad \text{when } p \leqslant n^{-1/(1-\gamma)}. \qquad (35)$$

In the remaining case $\overline{M} < np^{1-2\gamma}$ and $p \geqslant n^{-1/(1-\gamma)}$ hold. Since $\overline{M} < n$ implies $\overline{M} = M$, we infer $t \leqslant M^r = (\overline{M})^r \leqslant n^r p^{r-2r\gamma}$. So, recalling that $\Psi \leqslant t^2/\Lambda \leqslant t^2/[(np)^{r-1}\mu]$ by (15) and that $\mu \geqslant dn^{r+1}p^r$ by (20), using $D \geqslant np$, $p \geqslant n^{-1/(1-\gamma)}$ and $\gamma \leqslant 1/(16r)$ we deduce that

$$\frac{\Psi}{D} \leqslant \frac{t^2}{(np)^r\mu} \leqslant \frac{n^r p^{r-4r\gamma}}{\mu} \leqslant \frac{1}{dnp^{4r\gamma}} \leqslant \frac{1}{dn^{1-4r\gamma/(1-\gamma)}} \leqslant \frac{1}{dn^{1/2}} \leqslant 1,$$

establishing $D \geqslant \Psi$. It follows that

$$e^{-D} \leqslant e^{-\Psi} \qquad \text{when } \overline{M} < np^{1-2\gamma} \text{ and } p \geqslant n^{-1/(1-\gamma)}, \tag{36}$$

which together with inequalities (32)–(35) implies the claimed estimate (29). $\qquad\square$

We now deduce the upper bounds of Theorem 2 and 5 from the upper tail inequality (28).

Proof of the upper bound in (4) of Theorem 2. Let $\gamma := 1/(9r)$. For $t := \varepsilon\mu \geqslant n^{-\alpha}\mu$ and $n \geqslant n_0(r)$ it is routine to check that $t^{2-1/r}/\mu \geqslant \mathbb{1}_{\{np \geqslant n^{-\gamma}\}}\gamma(\log n)^r$ holds for $\alpha = \alpha(r) > 0$ sufficiently small. Hence Theorem 11 applies with $t = \varepsilon\mu$, where $\Lambda = \Theta_{r,\xi}(\sigma^2)$ by Remark 3. Using $\Phi(\varepsilon) \geqslant 1$ it follows that

$$\mathbb{P}(X \geqslant (1+\varepsilon)\mu) \leqslant (1+n^{-1}) \cdot e^{-\Omega_{r,\xi}(\Phi(\varepsilon))} \leqslant e^{-\Omega_{r,\xi}(\Phi(\varepsilon))}, \tag{37}$$

establishing the upper bound in (4). $\qquad\square$

For Theorem 5 we shall simplify the form of the exponent in (28) via the following auxiliary result, writing $a_n \asymp b_n$ instead of $a_n = \Theta(b_n)$ for typographic reasons (the assumption $p \geqslant n^{-9}$ in (ii) is ad-hoc).

Lemma 12. *Given $\xi \in (0,1)$, the following holds whenever $p \in (0, 1-\xi]$.*
(i) *If $t \leqslant \mu$, then*
$$t^2/\sigma^2 \asymp \varphi(t/\mu)\mu^2/\sigma^2 \asymp_{r,\xi} \varphi(t/\mu)\mu^2/\Lambda. \tag{38}$$
(ii) *If $t \geqslant \mu$ and $t^{1-1/r} \geqslant (\log n)\mathbb{1}_{\{p<1/n\}}$, then $p \geqslant n^{-9}$ implies*
$$t^2/\sigma^2 \geqslant \varphi(t/\mu)\mu^2/\sigma^2 \asymp_{r,\xi} \varphi(t/\mu)\mu^2/\Lambda = \Omega_{r,\xi}\big(M\log(e/p)\big). \tag{39}$$
(iii) *If $t^2/\sigma^2 \geqslant \min\{M,1\}$ and $\mu + t \geqslant 1$, then $t = \Omega_{r,\xi}(1)$.*

Proof. Inequality (38) and the first two estimates of equation (39) follow immediately from (15) and $\Lambda = \Theta_{r,\xi}(\sigma^2)$, see Remark 3. We now turn to the final inequality of equation (39). By combining (15) and $\Lambda/\mu = 1 + (np)^{r-1}$ with $M = \max\{t^{1/r}, t/n^{r-1}\}$ and $t^{1-1/r} \geqslant (\log n)\mathbb{1}_{\{p<1/n\}} + \mu^{1-1/r}\mathbb{1}_{\{p \geqslant 1/n\}}$, using $p \geqslant n^{-9}$ and $\mu^{1-1/r} = \Omega_r(n^{1/r}(np)^{r-1})$, see (20), it follows similarly to (31) that

$$\frac{\varphi(t/\mu)\mu^2/\Lambda}{M} \geqslant \frac{t\mu}{3\Lambda M} \geqslant \frac{\min\{t^{1-1/r}, n^{r-1}\}}{6\max\{1, (np)^{r-1}\}}$$

$$\geqslant \frac{1}{6}\min\left\{\log n, \frac{\mu^{1-1/r}}{(np)^{r-1}}, n^{r-1}, \frac{1}{p^{r-1}}\right\} = \Omega_r\big(\log(e/p)\big),$$

where we exploited that calculus gives $p^{r-1} \log(e/p) = O_r(1)$; this completes the proof of claims (i)–(ii).

For claim (iii) we may of course assume $t \leqslant 1/2$ (otherwise there is nothing to show). Hence $t^2/\sigma^2 \geqslant \min\{M, 1\} \geqslant \min\{t^{1/r}, 1\} = t^{1/r}$ implies $t^{2-1/r} \geqslant \sigma^2 = \Omega_{r,\xi}(\mu)$ by Remark 3, which in turn gives $t = \Omega_{r,\xi}(1)$, because $t + \mu \geqslant 1$ and $t \leqslant 1/2$ together imply $\mu \geqslant 1 - t \geqslant 1/2$, completing the proof. $\qquad\square$

Proof of the upper bound in (6) *of Theorem 5.* Applying Theorem 11 (with $\gamma = 1$), using (i)–(ii) of Lemma 12 it follows that inequality (28) holds with $\Omega_{r,\xi}(\Psi(t))$ in the exponent, where $\Psi(t) \geqslant \min\{1, t^{1/r}\} = \Omega_{r,\xi}(1)$ by (iii) of Lemma 12. Absorbing the $1 + n^{-1}$ factor similar to (37) then establishes the upper bound in (6). $\qquad\square$

Proof of the upper bound in (7) *of Theorem 5.* Since $\sigma^2 = \Omega_{r,\xi}(\mu)$ by Remark 3, note that the assumption

$$t^2/\sigma^2 \geqslant M \log(e/p) \cdot (\log n)^{2r} \tag{40}$$

implies $t^2/\mu \geqslant M \log(e/p) \cdot (\log n)^{r-1}$, so that Theorem 11 (with $\gamma = 1$) applies. Using (40), by (iii) of Lemma 12 we also infer that $M \geqslant t^{1/r} = \Omega_{r,\xi}(1)$. Absorbing the $1 + n^{-1}$ factor as before, it remains to show that the exponent of inequality (28) is $\Omega_{r,\xi}(M \log(e/p))$. For $t \leqslant \mu$ this follows from (38) of Lemma 12 and (40). For $t \geqslant \mu$ this follows from (39) of Lemma 12, since (40) and $p < n^{-1}$ imply $t^2/(\log n)^{2r+1} \geqslant \sigma^2 M = \Omega_{r,\xi}(\mu) = \Omega_{r,\xi}(1)$ and thus $t^{1-1/r} \geqslant (\log n) \mathbb{1}_{\{p<1/n\}}$, as required. $\qquad\square$

2.3 Straightforward extension to a certain sum of iid random variables

We close this section by recording that minor (and in fact simpler) variants of our proofs also apply to the following sum of independent random variables:

$$X := \sum_{i \in [n]} \binom{Y_i}{r} \qquad \text{with independent } Y_i \sim \text{Bin}(n, p). \tag{41}$$

Indeed, in view of the structural similarities to the number of r-armed stars in $G_{n,p}$ (which satisfies $X_{n,r,p} = \sum_{v \in [n]} \binom{d_v}{r}$, writing d_v for the degree of v), here we set $X_x := \sum_{i \in [n]: Y_i \leqslant \lfloor x \rfloor} \binom{Y_i}{r}$, and define N_x as the number of $i \in [n]$ with $Y_i \geqslant \lceil x \rceil$. Now the proofs of Lemma 6 and 10 carry over with minor changes: exploiting that there are no dependencies between the Y_i, using a simple dyadic decomposition we here obtain

$$X \leqslant X_D + \sum_{0 \leqslant j < J} \left[N_{D_j} \cdot \binom{\lfloor D_{j+1} \rfloor}{r} \right] \leqslant X_D + 2 \sum_{0 \leqslant j < J} N_{D_j} D_j^r \leqslant \cdots \leqslant X_D + t/2.$$

For the proof of Corollary 8 it suffices to show that $X_D \leqslant Z_C$ holds in the present setting. Since Y_i is a sum of n independent indicators $\xi_{i,j}$, we may write each $\binom{Y_i}{r}$ as a sum of $\binom{n}{r}$ dependent indicators (which each are products of some r distinct independent variables $\xi_{i,j}$). Using the constraint $Y_i \leqslant \lfloor D \rfloor$ the analogous left hand side of (14) is thus bounded by $r \cdot \binom{\lfloor D \rfloor}{r-1} \leqslant 2D^{r-1}$, which in turn implies $X_D \leqslant Z_C$, as desired. Since the proof of Lemma 9 also remains valid (as inequality (19) carries over), we thus arrive at the following result.

Theorem 13 (Upper tail bounds: an extension)**.** *The upper bounds on the upper tail* $\mathbb{P}(X \geqslant (1 + \varepsilon)\mu)$ *from Theorems 1, 2, 5, and 11 remain valid for the random variable X defined in* (41)*.*

Perhaps surprisingly, we are not aware of any standard method or inequality (for sums of iid variables) which can routinely recover the upper tail bounds from Theorem 13, and we leave it as an intriguing open problem to find or develop one. Here one technical difficulty seems to be that each summand $\binom{Y_i}{r}$ has an upper tail that decays slower than exponentially (for $r \geqslant 2$), which presumably is closely linked to the somewhat non-standard $\log(1/p)$ term in the exponent.

Acknowledgements.

References

[1] B. Bhattacharya, S. Ganguly, E. Lubetzky, and Y. Zhao. Upper tails and independence polynomials in random graphs. *Adv. Math.* **319** (2017), 313–347.

[2] B. Bollobás. Threshold functions for small subgraphs. *Math. Proc. Cambridge Philos. Soc.* **90** (1981), 197–206.

[3] S. Chatterjee. The missing log in large deviations for triangle counts. *Random Struct. Alg.* **40** (2012), 437–451.

[4] S. Chatterjee and A. Dembo. Nonlinear large deviations. *Adv. Math.* **299** (2016), 396–450.

[5] S. Chatterjee and S.R.S. Varadhan. The large deviation principle for the Erdős-Rényi random graph. *European J. Combin.* **32** (2011), 1000–1017.

[6] N. Cook and A. Dembo. Large deviations of subgraph counts for sparse Erdős–Rényi graphs. Preprint (2018). arXiv:1809.11148.

[7] B. DeMarco and J. Kahn. Upper tails for triangles. *Random Struct. Alg.* **40** (2012), 452–459.

[8] B. DeMarco and J. Kahn. Tight upper tail bounds for cliques. *Random Struct. Alg.* **41** (2012), 469–487.

[9] A. Dudek, J. Polcyn, and A. Ruciński. Subhypergraph counts in extremal and random hypergraphs and the fractional q-independence. *J. Comb. Optim.* **19** (2010), 184–199.

[10] R. Eldan. Gaussian-width gradient complexity, reverse log-Sobolev inequalities and nonlinear large deviations. *Geom. Funct. Anal.* **28** (2018), 1548–1596.

[11] P. Erdős and A. Rényi. On the evolution of random graphs. *Magyar Tud. Akad. Mat. Kutató Int. Közl.* **5** (1960), 17–61.

[12] T.E. Harris. A lower bound for the critical probability in a certain percolation process. *Proc. Cambridge Philos. Soc.* **56** (1960), 13–20.

[13] S. Janson. Poisson approximation for large deviations. *Random Struct. Alg.* **1** (1990), 221–229.

[14] S. Janson, T. Łuczak, and A. Ruciński. An exponential bound for the probability of nonexistence of a specified subgraph in a random graph. In *Random graphs '87 (Poznań, 1987)*, pp. 73–87, Wiley, Chichester (1990).

[15] S. Janson, T. Łuczak, and A. Ruciński. *Random graphs. Wiley-Interscience Series in Discrete Mathematics and Optimization.* Wiley-Interscience, New York (2000).

[16] S. Janson and A. Ruciński. The infamous upper tail. *Random Struct. Alg.* **20** (2002), 317–342.

[17] S. Janson, K. Oleszkiewicz, and A. Ruciński. Upper tails for subgraph counts in random graphs. *Israel J. Math.* **142** (2004), 61–92.

[18] S. Janson and A. Ruciński. The deletion method for upper tail estimates. *Combinatorica* **24** (2004), 615–640.

[19] S. Janson and A. Ruciński. Upper tails for counting objects in randomly induced subhypergraphs and rooted random graphs. *Ark. Mat.* **49** (2011), 79–96.

[20] S. Janson and L. Warnke. The lower tail: Poisson approximation revisited. *Random Struct. Alg.* **48** (2016), 219–246.

[21] J.H. Kim and V.H. Vu. Concentration of multivariate polynomials and its applications. *Combinatorica* **20** (2000), 417–434.

[22] E. Lubetzky and Y. Zhao. On replica symmetry of large deviations in random graphs. *Random Struct. Alg.* **47** (2015), 109–146.

[23] E. Lubetzky and Y. Zhao. On the variational problem for upper tails in sparse random graphs. *Random Struct. Alg.* **50** (2017), 420–436.

[24] O. Riordan and L. Warnke. The Janson inequalities for general up-sets. *Random Struct. Alg.* **46** (2015), 391–395.

[25] A. Ruciński. When are small subgraphs of a random graph normally distributed? *Probab. Theory Related Fields* **78** (1988), 1–10.

[26] M. Šileikis. On the upper tail of counts of strictly balanced subgraphs. *Electron. J. Combin.* **19** (1) #P4 (2012).

[27] M. Šileikis. *Inequalities for Sums of Random Variables: a combinatorial perspective.* PhD thesis, AMU Poznań (2012). Available from http://hdl.handle.net/10593/2870.

[28] M. Šileikis and L. Warnke. A counterexample to the DeMarco-Kahn Upper Tail Conjecture. *Random Struct. Alg.* **55** (2019), 775–794.

[29] J. Spencer. Counting extensions. *J. Combin. Theory Ser. A* **55** (1990). 247–255.

[30] V.H. Vu. A large deviation result on the number of small subgraphs of a random graph. *Combin. Probab. Comput.* **10** (2001), 79–94.

[31] L. Warnke. Upper tails for arithmetic progressions in random subsets. *Israel J. Math.* **221** (2017), 317–365.

[32] L. Warnke. On the method of typical bounded differences. *Combin. Probab. Comput.* **25** (2016), 269–299.

[33] L. Warnke. On the missing log in upper tail estimates. *J. Combin. Theory Ser. B* **140** (2020), 98–146.

A Appendix: Lower bounds on the upper tail

In this appendix we establish fairly routine lower bounds on the upper tail $\mathbb{P}(X \geqslant (1 + \varepsilon)\mu)$ from Theorem 1, 2, and 5 (omitting some straightforward details). Following [31] we obtain our lower bounds via the following three events: that many copies of $K_{1,r}$ 'cluster' on few edges (see Lemma 14 and 16), that most copies of $K_{1,r}$ arise disjointly (see Lemma 15 and 17), and that $G_{n,p}$ contains more edges than expected (see Lemma 18).

A.1 Basic argument for Theorem 1

For Theorem 1 we shall use two different lower bounds, and the first one is based on the idea that relatively few edges (which cluster in an appropriate way) can create fairly many stars $K_{1,r}$. This is formalized by the following result, which implies $\mathbb{P}(X_{r,n,p} \geqslant x) \geqslant \mathbb{P}(F \subseteq G_{n,p}) = p^{|E(F)|}$ since $F \subseteq G_{n,p}$ enforces $X_{r,n,p} \geqslant x$.

Lemma 14 (Clustering)**.** *For every $r \geqslant 1$ there is $D = D(r) > 0$ so that for all $n \geqslant 1$ and $0 < x \leqslant X_{r,n,1}$ there is $F \subseteq K_n$ with $|E(F)| \leqslant D \max\{x^{1/r}, x/n^{r-1}, 1\}$ edges such that F contains at least x copies of $K_{1,r}$.*

Inspired by the proofs of Theorem 1.3 and 1.5 in [17], the idea is to use a complete bipartite graph $F = K_{y,z}$ with $z = \Theta_r(\min\{x^{1/r}, n\})$ and $y = \Theta_r(x/z^r)$, which contains $yz = \Theta_r(x/z^{r-1}) = O_r(\max\{x^{1/r}, x/n^{r-1}\})$ edges and at least $y\binom{z}{r} = \Theta_r(yz^r) = \Omega_r(x)$ copies of $K_{1,r}$ (certain border cases require minor care).

Proof of Lemma 14. Let $x_0 := 2(4r)^r$, $n_0 := (r + 1)x_0$, and $D := n_0^2$. If (i) $x_0 \leqslant x \leqslant n^{r+1}/D$ and $n \geqslant n_0$, then we let $F := K_{y,z}$, with $z := \lceil \min\{x^{1/r}, n\}/4 \rceil$ and $y := \lceil r^r x/z^r \rceil$. Note that $F \subseteq K_n$ exists, since it is easy to check that $1 < y < n/2$ and $1 < z < n/2$, say (we leave the details to the reader). Furthermore, F contains at least $y\binom{z}{r} \geqslant y(z/r)^r \geqslant x$ many $K_{1,r}$, and $|E(F)| = yz \leqslant 2r^r x/z^{r-1} \leqslant D \max\{x^{1/r}, x/n^{r-1}\}$ edges.

If either (ii) $1 \leqslant n < n_0$ or (iii) $x > n^{r+1}/D$ and $n \geqslant n_0$, then we let $F := K_n$, which trivially contains $X_{r,n,1} \geqslant x$ copies of $K_{1,r}$, and $|E(F)| < n^2 < \max\{n_0^2, Dx/n^{r-1}\} = D \max\{1, x/n^{r-1}\}$ edges.

Finally, if (iv) $x < x_0$ and $n \geqslant n_0$, then we let $F := K_{n_0}$, which contains at least $n_0/(r + 1) = x_0 > x$ vertex disjoint copies of $K_{1,r}$ and $|E(F)| < n_0^2 = D$ edges, completing the proof. $\square$

The second lower bound is inspired by the fact that $X = X_{n,r,p}$ is approximately Poisson for small p, in which case most $K_{1,r}$ arise disjointly. Indeed, the following standard result bounds $\mathbb{P}(X = m)$ from below by the probability that there are exactly m vertex-disjoint copies of $K_{1,r}$ (see [8, 26, 31] for similar arguments), which for $m = (1+\varepsilon)\mu$ will imply $\mathbb{P}(X \geqslant m) \geqslant e^{-\Theta_{r,\varepsilon}(m)}$; the precise form of (42) will be useful later on.

Lemma 15 (Disjoint approximation). *Given $r \geqslant 2$ there are $n_0, b > 0$ (depending only on r) such that, for all $n \geqslant n_0$, $0 < p \leqslant n^{-1-1/(r+1)}$ and integers $m \in \mathbb{N}$ satisfying $0 \leqslant m \leqslant 99 \max\{\mu, n^{1/(r+1)}\}$, we have*

$$\mathbb{P}(X = m) \geqslant e^{-b} \cdot \binom{X_{r,n,1}}{m} p^{rm}(1-p^r)^{X_{n,r,1}-m}. \tag{42}$$

Proof. Let $\mathcal{K}$ contain all copies of $K_{1,r}$ in K_n. Define $\mathfrak{S}_m$ as the collection of all m-element subsets of $\mathcal{K}$ in which all stars $K_{1,r}$ are vertex disjoint. Given $\mathcal{C} \subseteq \mathfrak{S}_m$, define $\mathcal{I}_\mathcal{C}$ as the event that all stars $K_{1,r}$ of $\mathcal{C}$ are present, and define $\mathcal{D}_\mathcal{C}$ as the event that none of the stars $K_{1,r}$ in $\mathcal{K} \setminus \mathcal{C}$ are present. Note that

$$\mathbb{P}(X = m) \geqslant \sum_{\mathcal{C} \in \mathfrak{S}_m} \mathbb{P}(\mathcal{I}_\mathcal{C} \text{ and } \mathcal{D}_\mathcal{C}) = \sum_{\mathcal{C} \in \mathfrak{S}_m} \mathbb{P}(\mathcal{I}_\mathcal{C})\mathbb{P}(\mathcal{D}_\mathcal{C} \mid \mathcal{I}_\mathcal{C}) \geqslant |\mathfrak{S}_m| p^{rm} \cdot \min_{\mathcal{C} \in \mathfrak{S}_m} \mathbb{P}(\mathcal{D}_\mathcal{C} \mid \mathcal{I}_\mathcal{C}).$$

Distinguishing the number of edges in which each star $\alpha \in \mathcal{K} \setminus \mathcal{C}$ overlaps with some star $K_{1,r}$ from the vertex-disjoint collection $\mathcal{C} \in \mathfrak{S}_m$, using Harris inequality [12] and $np = o(1)$ we routinely obtain

$$\mathbb{P}(\mathcal{D}_\mathcal{C} \mid \mathcal{I}_\mathcal{C}) \geqslant (1-p^r)^{X_{n,r,1}-m} \prod_{1 \leqslant j < r} (1-p^{r-j})^{O_r(mn^{r-j})} \geqslant (1-p^r)^{X_{n,r,1}-m} e^{-O_r(mnp)},$$

where $mnp = O(\max\{n^{r+2}p^{r+1}, n^{1+1/(r+1)}p\}) = O(1)$. Furthermore, with $(z-y)^y/y! \leqslant \binom{z}{y} \leqslant z^y/y!$, $1 - x \geqslant e^{-2x}$ and $X_{n,r,1} = n\binom{n-1}{r}$ in mind, basic counting (and a short calculation) gives

$$|\mathfrak{S}_m| \geqslant \left((n-m)\binom{n-(r+1)m}{r}\right)^m \Big/ m! \geqslant \binom{X_{n,r,1}}{m} e^{-O_r(m^2/n)}.$$

This completes the proof of (42) since $m^2 = O(\max\{n^{2(r+1)}p^{2r}, n^{2/(r+2)}\}) = O(n^{2/(r+2)}) = o(n)$. $\qquad \square$

Combining the above two results, we now prove the lower bound of Theorem 1.

Proof of the lower bound in (3) of Theorem 1. We will tacitly assume $n \geqslant n_0(r, \varepsilon)$ whenever necessary. Applying Lemma 14 with $x := (1+\varepsilon)\mu$, there is $F \subseteq K_n$ with $|E(F)| \leqslant O_{r,\varepsilon}(\max\{\mu^{1/r}, \mu/n^{r-1}\})$ edges that contains at least $(1+\varepsilon)\mu$ copies of $K_{1,r}$. If $\Phi = \max\{\mu^{1/r}, \mu/n^{r-1}\}\log(1/p)$, then it follows that

$$\mathbb{P}(X \geqslant (1+\varepsilon)\mu) \geqslant \mathbb{P}(F \subseteq G_{n,p}) = p^{|E(F)|} \geqslant e^{-O_{r,\varepsilon}(\Phi)}. \tag{43}$$

Otherwise $\Phi = \mu$, which by a short calculation implies $\mu \leqslant (\log n)^3$, say (since assuming for contradiction that $\mu > (\log n)^3$ would imply $p = \Omega_r(n^{-1-1/r}) \geqslant n^{-2}$ and thus $\max\{\mu^{1/r}, \mu/n^{r-1}\} \log(1/p) = O_r(\max\{\mu^{1/r}, \mu/n^{r-1}\} \log n) < \mu$). Applying Lemma 15 with $m := \lceil (1+\varepsilon)\mu \rceil < n^{1/(r+1)}$, using $\binom{z}{y} \geqslant (z/y)^y$, $\mu = X_{n,r,1}p^r$, $1 - x \geqslant e^{-2x}$ and $m \geqslant (1+\varepsilon)\mu \geqslant 1$ it follows that

$$\mathbb{P}(X \geqslant (1+\varepsilon)\mu) \geqslant \mathbb{P}(X = m) \geqslant e^{-O_r(1)} \cdot \left(\frac{\mu}{m}\right)^m e^{-2\mu} \geqslant e^{-\Theta_{r,\varepsilon}(m)} \geqslant e^{-O_{r,\varepsilon}(\Phi)}, \qquad (44)$$

establishing the lower bound in (3). $\qquad\square$

A.2 Refined arguments for Theorem 2 and 5

For Theorem 2 and 5 we shall refine the previous two lower bounds, and also introduce a new third lower bound. Each time some care is needed to obtain the 'correct' dependence on $t = \varepsilon\mu$ in the exponent, and we start by refining the 'clustering' based lower bound from Lemma 14 and (43).

Lemma 16 (Refined clustering bound). *Given $r \geqslant 1$ and $\xi \in (0,1)$ there are $n_0, c > 0$ (depending only on r, ξ) such that, for all $n \geqslant n_0$, $p \in (0, 1-\xi]$ and $t \geqslant \sigma$ satisfying $1 \leqslant \mu + t \leqslant X_{r,n,1}$, we have*

$$\mathbb{P}(X \geqslant \mu + t) \geqslant \exp\left(-c \max\{t^{1/r}, t/n^{r-1}\} \log(1/p)\right). \qquad (45)$$

In case of $p = o(1)$ the basic proof idea is to obtain $\mu + t$ copies of $K_{1,r}$ as follows: (i) we first use the clustering construction from Lemma 14 to plant $2t$ copies of $K_{1,r}$, and (ii) then use Harris' inequality and a one-sided Chebychev's inequality to show that typically $\geqslant \mu - t$ of the *remaining* $\tilde{X}_{n,r,1} := X_{n,r,1} - 2t$ other copies of $K_{1,r}$ are present (the crux is that the *expected* number of such copies is $\tilde{X}_{n,r,1}p^r = \mu - o(t)$, so having $\geqslant \mu - t$ of them intuitively seems likely). For the resulting lower bound step (i) with probability $p^{O_r(\max\{t^{1/r}, t/n^{r-1}\})}$ thus ought to give the main contribution, making (45) plausible. For technical reasons, in the actual argument we have to plant $\min\{(\beta+1)t, \lceil \mu + t \rceil\}$ copies of $K_{1,r}$ for carefully chosen $\beta > 0$. By mimicking the proof of Theorem 21 in [31] we then easily arrive at (45) above; we leave the details to the reader.

We next refine the 'disjoint approximation' based lower bound used in Lemma 15 and (44) for small p. The idea is that inequality (42) intuitively relates $X = X_{n,r,p}$ to a binomial random variable with mean $\mu = X_{n,r,1} \cdot p^r$, which makes the following Chernoff-type bound for the upper tail plausible.

Lemma 17 (Disjoint approximation: Chernoff-type lower bound). *Given $r \geqslant 2$ there are $n_0, c, d > 0$ (depending only on r) such that, for all $n \geqslant n_0$, $0 < p \leqslant n^{-1-1/(r+1)}$ and $t > 0$ satisfying $1 \leqslant \mu + t \leqslant 9 \max\{\mu, n^{1/(r+1)}\}$, we have*

$$\mathbb{P}(X \geqslant \mu + t) \geqslant d \exp\left(-c\varphi(t/\mu)\mu\right). \qquad (46)$$

Noting the binomial-like form of inequality (42), it is routine to check that Lemma 15 indeed implies (46) above (e.g., by summing (42) as in the proof of Theorem 22 in [31]); we leave the details to the reader.

Our third lower bound for moderately large p it is based on the idea that a deviation in the number of edges should typically entail a deviation in the number of $K_{1,r}$ copies (in concrete words: if $G_{n,p}$ has substantially more than $\binom{n}{2}p$ edges, then we expect to have more $K_{1,r}$ copies than on average).

Lemma 18 (Deviation in number of edges: sub-Gaussian type lower bound). *Given $r \geqslant 2$ and $\xi \in (0,1)$ there are $n_0, \beta, c > 0$ (depending only on r, ξ) such that, setting $\Lambda := \mu(1 + (np)^{r-1})$, for all $n \geqslant n_0$, $\xi n^{-1} \leqslant p \leqslant 1 - \xi$ and $\sigma \leqslant t \leqslant \beta\mu$ we have*

$$\mathbb{P}(X \geqslant \mu + t) \geqslant \exp\left(-ct^2/\Lambda\right). \tag{47}$$

Remark 19. By Remark 3, in inequality (47) we have $\Lambda = \Theta_{r,\xi}(\sigma^2)$, where $\sigma^2 = \operatorname{Var} X$.

Setting $\varepsilon := t/\mu$, the basic proof idea is to (i) condition on having $|E(G_{n,p})| \geqslant (1+\varepsilon)\binom{n}{2}p$ edges, and (ii) then show that this conditioning converts $X \geqslant \mu + t = (1+\varepsilon)\mu$ into a typical event (the crux is that this conditioning drives up the expected value of $X = X_{n,r,p}$; to see this it might help to think of the uniform random graph $G_{n,m}$ with $m = (1+\varepsilon)\binom{n}{2}p$ edges). For the resulting lower bound the conditioning thus ought to give the main contribution, which by folklore results satisfies $\mathbb{P}(|E(G_{n,p})| \geqslant (1+\varepsilon)\binom{n}{2}p) = \exp\left(-\Theta_{\xi}(\varepsilon^2\binom{n}{2}p)\right)$. This makes inequality (47) plausible, since $\varepsilon^2\binom{n}{2}p = \varepsilon^2 \cdot \Theta_{r,\xi}(\mu^2/\Lambda) = \Theta_{r,\xi}(t^2/\Lambda)$ for the considered range of p. A simple modification of the proof of Theorem 24 in [31] makes this idea rigorous and establishes (47) above; we leave the details to the reader (we mention in passing that a tilting argument also works here).

Stitching the above three results together, we now prove the lower bounds of Theorem 2 and 5.

Proof of the lower bound in (7) *of Theorem 5.* By (iii) of Lemma 12 we infer that $M \geqslant t^{1/r} = \Omega_{r,\xi}(1)$, which in turn implies $t^2/\sigma^2 \geqslant M \log(e/p) \cdot (\log n)^{2r} \geqslant 1$ and thus $t \geqslant \sigma$. Hence an application of Lemma 16 (see inequality (45)) establishes the lower bound in (7). $\qquad\square$

Proof of the lower bound in (6) *of Theorem 5.* We shall only assume $p \geqslant n^{-1}$ instead of $p \geqslant n^{-1}\log n$. Applying Lemmas 16 and 18, and using Remark 19, it follows that there is $\beta = \beta(r,\xi) > 0$ such that

$$\mathbb{P}(X \geqslant \mu + t) \geqslant \max\left\{e^{-\Theta_{r,\xi}(M\log(e/p))},\ \mathbb{1}_{\{t \leqslant \beta\mu\}}e^{-\Theta_{r,\xi}(t^2/\sigma^2)}\right\}. \tag{48}$$

By a virtually identical calculation as in the proof of (39) from Lemma 12, for $t \geqslant \beta\mu$ it follows that $t^2/\sigma^2 \geqslant \Omega_{r,\xi}(M\log(e/p))$ holds. After adjusting the implicit constants, it follows that we can remove the indicator in inequality (48), which in view of $\Psi(t) = \min\{t^2/\sigma^2, M\log(e/p)\}$ establishes the lower bound in (6). $\qquad\square$

Proof of the lower bound in (4) *of Theorem 2.* Set $t := \varepsilon\mu$ and $M := \max\{t^{1/r}, t/n^{r-1}\}$, as usual. Using (15) we have $(\varepsilon\mu)^2/\sigma^2 \geqslant \varphi(\varepsilon)\mu^2/\sigma^2 \geqslant \Phi(\varepsilon) \geqslant 1$ by assumption, so $t \geqslant \sigma$ follows. In the following we shall distinguish the three cases (i) $n^{-1} \leqslant p \leqslant 1 - \xi$, (ii) $n^{-1-1/(r+1)} \leqslant p < n^{-1}$, and (iii) $0 < p < n^{-1-1/(1+r)}$.

In cases (i)–(ii) note that, say, $\mu^{1-1/r} = \Omega_r(n^{1/3r}) > \log n$ holds. Using (i)–(ii) of Lemma 12, it thus suffices to prove the lower bound of (4) with exponent $\Phi(\varepsilon)$ replaced by $\Psi(t)$ defined in Theorem 5. In case (i) this bound follows from the above proof (valid for $n^{-1} \leqslant p \leqslant 1 - \xi$) of the lower bound in (6), and in case (ii) we shall now argue that this bound follows from inequality (45) of Lemma 16, by establishing that $t^2/\sigma^2 = \Omega_{r,\xi}(M \log(e/p))$ holds. Indeed, since $p < n^{-1}$ and Remark 3 imply $\sigma^2 = \Theta_r(\mu)$, after recalling $\mu^{1-1/r} = \Omega_r(n^{1/3r})$ and $t = \varepsilon\mu \geqslant n^{-\alpha}\mu$ it then follows for $\alpha = \alpha(r) > 0$ sufficiently small (say, $\alpha < 1/6r$) that

$$\frac{t^2/\sigma^2}{M \log(e/p)} = \frac{\min\left\{t^{2-1/r},\ tn^{r-1}\right\}}{\sigma^2 \log(e/p)} \geqslant \frac{\min\left\{\mu^{1-1/r}n^{-2\alpha},\ n^{r-1-\alpha}\right\}}{\Theta_{r,\xi}(\log n)} \geqslant 1, \qquad (49)$$

completing the proof in cases (i)–(ii).

In the remaining case (iii) Lemmas 16 and 17 imply that, for some constant $d = d(r) \in (0, 1]$, we have

$$\mathbb{P}(X \geqslant \mu + t) \geqslant d \cdot \max\left\{e^{-\Theta_{r,\xi}(M \log(e/p))},\ \mathbb{1}_{\{\mu+t \leqslant 9 \max\{\mu, n^{1/(r+1)}\}\}}e^{-\Theta_{r,\xi}(\varphi(t/\mu)\mu)}\right\}. \qquad (50)$$

We claim that for $\mu + t > 9 \max\left\{\mu, n^{1/(r+1)}\right\}$ we have $\varphi(t/\mu)\mu = \Omega_r\left(M \log(e/p)\right)$. Indeed, noting that $\varphi(x) \geqslant x(\log x)/2$ for $x \geqslant e^2 \approx 7.4$ (which is easy to check by calculus), it follows that

$$\frac{\varphi(t/\mu)\mu}{M \log(e/p)} \geqslant \frac{\min\{t^{(r-1)/r},\ n^{r-1}\}\log(t/\mu)}{2\log(e/p)} = \Omega_r\left(n^{(r-1)/(r^2+r)} \cdot \frac{\log(t/\mu)}{\log(e/p)}\right).$$

Moreover, $\log(t/\mu)/\log(e/p) = \Omega_r(1)$ when $\mu \leqslant p$, and $\log(t/\mu)/\log(e/p) = \Omega_r((\log n)^{-1})$ when $\mu > p$. In each case the claimed inequality holds, which allows omitting the indicator in (50). Since $\mu = \Theta_r(\mu^2/\sigma^2)$ by Remark 3, now $\mathbb{P}(X \geqslant (1+\varepsilon)\mu) \geqslant d \cdot e^{-\Theta_{r,\xi}(\Phi(\varepsilon))}$ follows, which in view of $\Phi(\varepsilon) \geqslant 1$ completes the proof. $\qquad\square$