On the missing log in upper tail estimates ☆

Lutz Warnke

School of Mathematics, Georgia Institute of Technology, Atlanta GA 30332, USA

A R T I C L E   I N F O

A B S T R A C T

In the late 1990s, Kim and Vu pioneered an inductive method for showing concentration of certain random variables X. Shortly afterwards, Janson and Ruciński developed an alternative inductive approach, which often gives comparable results for the upper tail $\mathbb{P}(X \geq (1 + \varepsilon)\mathbb{E}X)$. In some cases, both methods yield upper tail estimates which are best possible up to a logarithmic factor in the exponent, but closing this narrow gap has remained a technical challenge. In this paper we present a BK-inequality based combinatorial sparsification idea that can recover this missing logarithmic term in the upper tail.

As an illustration, we consider random subsets of the integers $\{1, \ldots, n\}$, and prove sharp upper tail estimates for various objects of interest in additive combinatorics. Examples include the number of arithmetic progressions, Schur triples, additive quadruples, and (r, s)-sums.

© 2019 Elsevier Inc. All rights reserved.

1. Introduction

Concentration inequalities are of great importance in discrete mathematics, theoretical computer science, and related fields. They intuitively quantify random fluctuations of a given random variable X, by bounding the probability that X differs substantially from its expected value $\mu = \mathbb{E}X$. In combinatorial applications, X often counts certain objects (e.g., the number of subgraphs or arithmetic progressions), in which case the random variable X can usually be written as a low-degree polynomial of many independent random variables. In this context concentration inequalities with exponentially small estimates are vital (e.g., to make union bound arguments amenable), and here Kim and Vu [19,30,32] achieved a breakthrough in the late 1990s. Their powerful concentration inequalities have since then, e.g., been successfully applied to many combinatorial problems, been included in standard textbooks, and earned Vu the George Pólya Prize in 2008.

In probabilistic combinatorics, the exponential rate of decay of the *lower tail* $\mathbb{P}(X \leq \mu - t)$ and *upper tail* $\mathbb{P}(X \geq \mu + t)$ have received considerable attention, since they are of great importance in applications (of course, this is also an interesting problem in concentration of measure). The behaviour of the lower tail is nowadays well-understood due to the celebrated Janson- and Suen-inequalities [10,21,17,11,13]. By contrast, the behaviour of the 'infamous' upper tail has remained a well-known technical challenge (see also [14,12]). Here the inductive method of Kim and Vu [19,32] from around 1998 often yields inequalities of the form

$$\mathbb{P}(X \geq (1+\varepsilon)\mu) \leq \exp\big(-c(\varepsilon)\mu^{1/q}\big), \tag{1}$$

where $q \geq 1$ is some constant. In 2000, Janson and Ruciński [15] developed an alternative inductive approach, which often gives comparable results for the upper tail, i.e., which recovers (1) up to the usually irrelevant numerical value of the parameter c. Studying the sharpness of the tail inequality (1) is an important problem according to Vu (see Section 4.8 in [32]). In fact, one main aim of the paper [15] was 'to stimulate more research into these methods' since 'neither of [them] seems yet to be fully developed'. In other words, Janson and Ruciński were asking for further improvements of the aforementioned fundamental proof techniques (the papers [15,32] already contained several tweaking options for decreasing q).

In this paper we address this technical challenge in cases where the inductive methods of Kim–Vu and Janson–Ruciński are nearly sharp. The crux is that, for several interesting classes of examples (naturally arising, e.g., in additive combinatorics), the upper tail inequality (1) is best possible up to a logarithmic factor in the exponent. Closing such narrow gaps has recently become an active area of research in combinatorial probability (see, e.g., [14,12,16,6,7,34,35]). The goal of this paper is to present a new idea that can add such missing logarithmic terms to the upper tail. From a conceptual perspective, this paper thus makes a new effect amenable to the rich toolbox of the Kim–Vu and

Janson–Ruciński methods (we believe that our techniques will be useful elsewhere). For example, under certain somewhat natural technical assumptions, our methods allow us to improve the classical upper tail inequality (1) to estimates of the form

$$\mathbb{P}(X \geq (1+\varepsilon)\mu) \leq \exp\Big(-c(\varepsilon)\min\{\mu,\, \mu^{1/q}s\}\Big) \quad \text{with} \quad s \in \{\log n,\, \log(1/p)\}, \quad (2)$$

where the reader may wish to tentatively think of the parameters $n = \omega(1)$ and $p = o(1)$ as those in the binomial random graph $G_{n,p}$ (here some extra assumptions are necessary, since there are examples where (1) is sharp, see Sections 1.1 and 6.1). This seemingly small improvement of (1) is conceptually important, since in several interesting applications the resulting inequality is best possible up to the value of c. Indeed, as we shall see, sharp examples with $\mathbb{P}(X \geq (1+\varepsilon)\mu) = \exp\big(-\Theta(\min\{\mu, \mu^{1/q}\log(1/p))\big)$ for $\varepsilon = \Theta(1)$ naturally arise when X counts various objects of great interest in additive combinatorics, such as the number of arithmetic progressions (of given length) or additive quadruples in random subsets of the integers $[n] = \{1, \ldots, n\}$.

In the remainder of this introduction we illustrate our methods with some applications, outline our high-level proof strategy, and discuss the structure of this paper. Noteworthily, our proof techniques do not solely rely on induction, but a blend of combinatorial and probabilistic arguments.

1.1. Flavour of the results

We now illustrate the main flavour of our upper tail results with some concrete examples. Many important counting problems can be rephrased as the number of edges induced by the random induced subhypergraph $\mathcal{H}_p = \mathcal{H}[V_p(\mathcal{H})]$ (see, e.g., [14,22,16,35, 37]), where $V_p(\mathcal{H})$ denotes the binomial random subset where each vertex $v \in V(\mathcal{H})$ is included independently with probability p. Our methods yield the following upper tail inequality for $\mathcal{H}_p$, which extends one of the main results from [35] for the special case $q = 2$, and sharpens one of the principle results of Janson and Ruciński [16] by a logarithmic factor in the exponent.

Theorem 1 *(Counting edges of random induced subhypergraphs). Let $1 \leq q < k$ and $\gamma, D > 0$. Assume that $\mathcal{H}$ is a k-uniform hypergraph with $v(\mathcal{H}) \leq N$ vertices and $e(\mathcal{H}) \geq \gamma N^q$ edges. Suppose that $\Delta_q(\mathcal{H}) \leq D$, where $\Delta_q(\mathcal{H})$ denotes the maximum number of edges of $\mathcal{H}$ that contain q given vertices. Let $X = e(\mathcal{H}_p)$ and $\mu = \mathbb{E}X$. Then for any $\varepsilon > 0$ there is $c = c(\varepsilon, k, \gamma, D) > 0$ such that for all $p \in (0, 1]$ we have*

$$\mathbb{P}(X \geq (1+\varepsilon)\mu) \leq \exp\Big(-c\min\{\mu,\, \mu^{1/q}\log(e/p)\}\Big). \quad (3)$$

This upper tail inequality is conceptually best possible in several ways. First, the restriction to $q < k$ is necessary (see Section 6.1 for a counterexample when $q = k$),

Second, in several important applications (3) is sharp (yields the correct exponential rate of decay), i.e., there is a matching lower bound of form

$$\mathbb{P}(X \geq (1+\varepsilon)\mu) \geq \mathbb{1}_{\{1 \leq (1+\varepsilon)\mu \leq e(\mathcal{H})\}} \exp\left(-C(\varepsilon) \min\{\mu, \ \mu^{1/q} \log(e/p)\}\right), \qquad (4)$$

where the restriction $1 \leq (1+\varepsilon)\mu \leq e(\mathcal{H})$ is natural.[1] In particular, letting the edges of the hypergraph $\mathcal{H}$ with vertex-set $V(H) = [n]$ encode classical objects from additive combinatorics and Ramsey Theory, sharp examples of type (3)–(4) include the number of k-term arithmetic progressions, Schur triples $x + y = 2z$, additive quadruples $x_1 + x_2 = y_1 + y_2$, and (r, s)-sums $x_1 + \cdots + x_r = y_1 + \cdots + y_s$ in the binomial random subset $[n]_p = V_p(\mathcal{H})$ of the integers; see Section 1.1.1 and 6.1 for more details/concrete examples.

The two expressions in the exponent of the upper tail (3)–(4) correspond to different phenomena.[2] Namely, in some range we expect that $X = e(\mathcal{H}_p)$ is approximately Poisson, in which case $\mathbb{P}(X \geq 2\mu)$ decays roughly like $\exp(-c\mu)$. Similarly, the $\exp(-c\mu^{1/q} \log(1/p)) = p^{c\mu^{1/q}}$ term intuitively corresponds to 'clustered' behaviour (see also [35,27,12]), where few vertices $U \subseteq V_p(\mathcal{H})$ induce many edges in $\mathcal{H}_p = \mathcal{H}[V_p(\mathcal{H})]$: e.g., in each of the above-mentioned examples there always is such a set with $|U| = c\mu^{1/q}$ and $e(\mathcal{H}[U]) \geq 2\mu$, which readily implies $\mathbb{P}(X \geq 2\mu) \geq \mathbb{P}(U \subseteq V_p(\mathcal{H})) = p^{c\mu^{1/q}}$. Note that classical tail inequalities of form (1) fail to handle these phenomena properly (lacking Poisson behaviour and the extra $\log(1/p)$ term).

1.1.1. Upper tail examples from additive combinatorics and Ramsey theory

In the following exemplary upper tail bounds (5)–(8) we tacitly allow the implicit constants to depend on ε.

Example 2. *Arithmetic progressions* (APs) are central objects in additive combinatorics. Given $k \geq 3$, let $X = X_{n,k,p}$ denote the number of arithmetic progressions of length k in the binomial random subset $[n]_p$ of the integers (to clarify: we count k-subsets $\{x_1, \ldots, x_k\} \subseteq [n]_p$ forming APs); note that $\mu = \mathbb{E}X = \Theta(n^2 p^k)$. Then, for any $\varepsilon > 0$ and $p = p(n) \in (0, 1]$ satisfying $1 \leq (1+\varepsilon)\mu \leq X_{n,k,1}$, we have

$$\mathbb{P}(X \geq (1+\varepsilon)\mu) = \exp\left(-\Theta\left(\min\{\mu, \ \mu^{1/2} \log(1/p)\}\right)\right). \qquad (5)$$

Example 3. *Schur triples* $\{x, y, z\} \subseteq [n]$ with $x + y = z$ (where $x \neq y$) are classical objects in Number theory and Ramsey theory (see, e.g., [9] and [8,24]). Let $X = X_{n,p}$ denote the number of Schur triples in $[n]_p$; note that $\mu = \mathbb{E}X = \Theta(n^2 p^3)$. Then, for any $\varepsilon > 0$ and $p = p(n) \in (0, 1]$ satisfying $1 \leq (1+\varepsilon)\mu \leq X_{n,1}$, we have

---

[1] Note that $\mathbb{P}(X \geq (1+\varepsilon)\mu) = 0$ when $(1+\varepsilon)\mu > e(H)$, and that $\mathbb{P}(X \geq (1+\varepsilon)\mu) = 1 - \mathbb{P}(X = 0)$ when $(1+\varepsilon)\mu < 1$.

[2] A phenomenon not relevant for the qualitative accuracy of (3)–(4) is that $|V_p(\mathcal{H})|$ can also be somewhat 'bigger' than $\mathbb{E}|V_p(\mathcal{H})|$, which in some range yields sub-Gaussian type tail behaviour, see also [35,27].

$$\mathbb{P}(X \geq (1+\varepsilon)\mu) = \exp\left(-\Theta\left(\min\{\mu, \ \mu^{1/2}\log(1/p)\}\right)\right). \tag{6}$$

The same tail bound also holds for ℓ-sums (studied, e.g., in [1]), where the 3-element subsets satisfy $x + y = \ell z$.

Example 4. *Additive quadruples* are 4-subsets $\{x_1, x_2, y_1, y_2\} \subseteq [n]$ satisfying $x_1 + x_2 = y_1 + y_2$. The number of these quadruples is also called *additive energy*, which is an important quantity in additive combinatorics (see, e.g., [2,5]). Let $X = X_{n,p}$ denote the number of additive quadruples in $[n]_p$; note that $\mu = \mathbb{E}X = \Theta(n^3 p^4)$. Then, for any $\varepsilon > 0$ and $p = p(n) \in (0,1]$ satisfying $1 \leq (1+\varepsilon)\mu \leq X_{n,1}$, we have

$$\mathbb{P}(X \geq (1+\varepsilon)\mu) = \exp\left(-\Theta\left(\min\{\mu, \ \mu^{1/3}\log(1/p)\}\right)\right). \tag{7}$$

Example 5. (r, s)-*sums* are $(r + s)$-subsets $\{x_1, \ldots, x_r, y_1, \ldots, y_2\} \subseteq [n]$ satisfying $x_1 + \cdots + x_r = y_1 + \cdots + y_s$. In the special case $r = s$ the number of these sets is called $2r$-fold additive energy, which is useful in the context of Roth's theorem (see, e.g., [5]). Given $r, s \geq 1$ satisfying $r + s \geq 3$, let $X = X_{n,r,s,p}$ denote the number of (r, s)-sums in $[n]_p$; note that $\mu = \mathbb{E}X = \Theta(n^{r+s-1}p^{r+s})$. Then, for any $\varepsilon > 0$ and $p = p(n) \in (0,1]$ satisfying $1 \leq (1+\varepsilon)\mu \leq X_{n,r,s,1}$, we have

$$\mathbb{P}(X \geq (1+\varepsilon)\mu) = \exp\left(-\Theta\left(\min\{\mu, \ \mu^{1/(r+s-1)}\log(1/p)\}\right)\right). \tag{8}$$

Similar tail bounds also hold for integer solutions of linear homogeneous systems, see Section 6.1 for the details.

1.1.2. Subgraph counts in random graphs: sub-Gaussian type upper tail bounds

As a side-product, our proof techniques also yield new results with a slightly different flavour. To illustrate this with subgraph counts in the binomial random graph $G_{n,p}$, let $X = X_H$ denote the number of copies of H in $G_{n,p}$. Set $\mu = \mathbb{E}X$. Here *sub-Gaussian type* upper tail estimates[3] of the form

$$\mathbb{P}(X \geq \mu + t) \leq C \exp(-ct^2/\operatorname{Var}X) \tag{9}$$

have been extensively studied [23,30,15,25,18,36,37] during the last decades, usually with emphasis on small deviations of form $\sqrt{\operatorname{Var}X} \leq t = o(\mu)$, say (differing from the large deviations regime $t = \Theta(\mu)$ considered in the classical upper tail problem for subgraph counts). In particular, for so-called 'strictly balanced' graphs H three different approaches [30,15,25] have been developed during the years 2000–2012, which each establish a form of inequality (9) for $t \leq \mu = O(\log n)$. Our methods allow us to break

---

[3] For subgraph counts lower tail estimates of sub-Gaussian type follow from Janson's inequality (see, e.g., [17]).

this logarithmic barrier slightly, answering a question of Janson and Ruciński [13]; see Section 6.2.1 for more details.

Theorem 6 *(Subgraph counts: sub-Gaussian type upper tail bounds). For any strictly balanced graph H there are $n_0, c, C, \xi > 0$ such that inequality* (9) *holds whenever $n \geq n_0$ and $0 < t \leq \mu \leq (\log n)^{1+\xi}$.*

1.2. Glimpse of the proof strategy

In contrast to most of the previous work, in this paper we take a more combinatorial perspective to concentration of measure (and avoid induction via a more iterative point of view). Our high-level proof strategy proceeds roughly as follows. In the *deterministic part* of the argument, we define several 'good' events $\mathcal{E}_i = \mathcal{E}_i(\mathcal{H}, \varepsilon)$, and show that the following implication holds:

$$\text{all } \mathcal{E}_i \text{ hold} \quad \Longrightarrow \quad X < (1 + \varepsilon)\mathbb{E}X. \tag{10}$$

In the *probabilistic part* of the argument, we show that for some suitable parameter Ψ we have

$$\mathbb{P}(\text{some } \mathcal{E}_i \text{ fails}) \leq \exp(-\Psi). \tag{11}$$

Combining both parts then readily yields an exponential upper tail estimate of the form

$$\mathbb{P}(X \geq (1 + \varepsilon)\mathbb{E}X) \leq \mathbb{P}(\text{some } \mathcal{E}_i \text{ fails}) \leq \exp(-\Psi).$$

In this paper we illustrate the above approach by implementing (10)–(11) in a general Kim–Vu/Janson–Ruciński type setup. To communicate our ideas more clearly, our below informal discussion again uses the simpler random induced subhypergraph setup (a more detailed sketch is given in Sections 3.1.2–3.1.3).

For the deterministic part (10), we shall crucially exploit a good event $\mathcal{E}_{Q,\varepsilon}$ of the following form: all subhypergraphs with 'small' maximum degree have 'not too many' edges, i.e., that $e(\mathcal{J}) < (1 + \varepsilon/2)\mathbb{E}X$ holds for all $\mathcal{J} \subseteq \mathcal{H}_p$ with $\Delta_1(\mathcal{J}) \leq Q$, say. Our *sparsification idea* proceeds roughly as follows. First, using combinatorial arguments (and further good events) we find a nested sequence of subhypergraphs

$$\mathcal{H}_p = \mathcal{J}_q \supseteq \mathcal{J}_{q-1} \supseteq \cdots \supseteq \mathcal{J}_2 \supseteq \mathcal{J}_1, \tag{12}$$

which gradually decreases the maximum degree down to $\Delta_1(\mathcal{J}_1) \leq Q$. The crux is that $\mathcal{E}_{Q,\varepsilon}$ then implies $e(\mathcal{J}_1) < (1 + \varepsilon/2)\mathbb{E}X$. In the second step we exploit various good events (and properties of the constructed sequence) to show that we obtained $\mathcal{J}_1$ by removing relatively few edges from $\mathcal{H}_p$, such that

$$X = e(\mathcal{H}_p) = e(\mathcal{J}_1) + \sum_{1 \le j < q} e(\mathcal{J}_{j+1} \setminus \mathcal{J}_j) < (1+\varepsilon/2)\mathbb{E}X + (\varepsilon/2)\mathbb{E}X = (1+\varepsilon)\mathbb{E}X. \quad (13)$$

In fact, the combinatorial arguments leading to (12)–(13) develop a 'maximal matching' based sparsification idea from [35], which is key for handling some vertices of $\mathcal{H}_p$ with exceptionally high degrees, say.

The probabilistic part (11) works hand in hand with the above deterministic arguments. Similar to $\mathcal{E}_{Q,\varepsilon}$, we shall throughout work with 'relative estimates', i.e., which are valid for all subhypergraphs of $\mathcal{H}_p$ satisfying some extra properties (e.g., that $\Delta_j(\mathcal{J}) \le R_j$ holds for all $J \subseteq \mathcal{H}_p$ with $\Delta_{j+1}(\mathcal{J}) \le R_{j+1}$). These estimates are crucial for bringing combinatorial arguments of type (12)–(13) into play (instead of relying solely on inductive reasoning), and they hinge on a concentration inequality from [35]. Perhaps surprisingly, this inequality allows us to estimate $\mathbb{P}(\neg\mathcal{E}_{Q,\varepsilon})$ and similar 'relative' events *without* taking a union bound over all subhypergraphs. For the matching based sparsification idea briefly mentioned above, we exploit the fact that the relevant 'matchings' guarantee the 'disjoint occurrence' of suitably defined events. This observation allows us to estimate the probability of certain 'bad' events via BK-inequality based moment arguments.

Finally, in our probabilistic estimates the logarithmic terms in (2)–(3) arise in a fairly delicate way (which comes as no surprise, since there are examples where (1) is sharp). We now illustrate the underlying technical idea for binomial random variables $X \sim \mathrm{Bin}(n, p)$ with $\mu = np$, where for $x \ge e(e/p)^\alpha \mu$ we have

$$\mathbb{P}(X \ge x) \le \binom{n}{x}p^x \le \left(\frac{e\mu}{x}\right)^x \le \left(\frac{p}{e}\right)^{\alpha x} = \exp\left(-\alpha x \log(e/p)\right).$$

Our proofs apply this 'overshooting the expectation yields extra terms in the exponent' idea to a set of carefully chosen auxiliary random variables. As the reader can guess, the technical details are, e.g., complicated by the fact that the edges of $\mathcal{H}_p$ are *not* independent, and that we may *not* assume $x \gg \mu$.

1.3. Guide to the paper

In Section 2 we introduce our key probabilistic tools. In Section 3 we give a fairly detailed proof outline, and present our main combinatorial and probabilistic arguments in the random induced subhypergraphs setup. In Section 4 we then extend the discussed arguments to a more general setup. In Section 5 we derive some concrete upper tail inequalities, which in Section 6 are then applied to several pivotal examples.

The reader interested in our *proof techniques* may wish to focus on Section 3, which contains our core ideas and arguments. The reader interested in *applications* may wish to skip to Section 6, where the 'easy-to-apply' concentration inequalities of Section 5.1 are used in several different examples. Finally, the reader interested in *comparing* our

results *with the literature* may wish to focus on the general setup of Section 4.1 and the concentration inequalities in Section 5.2.

2. Probabilistic preliminaries

2.1. A Chernoff-type upper tail inequality

In this subsection we state a powerful Chernoff-type upper tail inequality from [35]. It might be instructive to check that, for sums $X = \sum_{i \in \mathcal{A}} \xi_i$ of independent random variables $\xi_i \in [0,1]$, inequality (14) below reduces to the classical Chernoff bound (writing $i \sim j$ if $i = j$, for $Y_i = \xi_i$, $\mathcal{I} = \mathcal{A}$ and $C = 1$ we have $X = Z_C$). We think of $\sim$ as a 'dependency relation': $\alpha \nsim \beta$ implies that the random variables Y_α and Y_β are independent. For indicator random variables $Y_\alpha \in \{0,1\}$ the condition $\max_{\beta \in \mathcal{J}} \sum_{\alpha \in \mathcal{J}:\alpha \sim \beta} Y_\alpha \leq C$ essentially ensures that each variable Y_β with $\beta \in \mathcal{J}$ 'depends' on at most C variables Y_α with $\alpha \in \mathcal{J}$. Intuitively, Z_C defined below thus corresponds to an approximation of $X = \sum_{\alpha \in \mathcal{I}} Y_\alpha$ with 'bounded dependencies'.

Theorem 7. *Given a family of non-negative random variables $(Y_\alpha)_{\alpha \in \mathcal{I}}$ with $\sum_{\alpha \in \mathcal{I}} \mathbb{E}Y_\alpha \leq \mu$, assume that $\sim$ is a symmetric relation on $\mathcal{I}$ such that each Y_α with $\alpha \in \mathcal{I}$ is independent of $\{Y_\beta : \beta \in \mathcal{I} \text{ and } \beta \nsim \alpha\}$. Let $Z_C = \max \sum_{\alpha \in \mathcal{J}} Y_\alpha$, where the maximum is taken over all $\mathcal{J} \subseteq \mathcal{I}$ with $\max_{\beta \in \mathcal{J}} \sum_{\alpha \in \mathcal{J}:\alpha \sim \beta} Y_\alpha \leq C$. Set $\varphi(x) = (1+x)\log(1+x) - x$. Then for all $C, t > 0$ we have*

$$\mathbb{P}(Z_C \geq \mu + t) \leq \exp\left(-\frac{\varphi(t/\mu)\mu}{C}\right) = e^{-\mu/C} \cdot \left(\frac{e\mu}{\mu+t}\right)^{(\mu+t)/C}$$

$$\leq \min\left\{\exp\left(-\frac{t^2}{2C(\mu+t/3)}\right), \left(1+\frac{t}{2\mu}\right)^{-t/(2C)}\right\} \leq \left(1+\frac{t}{\mu}\right)^{-t/(4C)}.$$

(14)

Remark 8. In applications there often is a family of independent random variables $(\xi_\sigma)_{\sigma \in \mathcal{A}}$ such that each Y_α is a function of $(\xi_\sigma)_{\sigma \in \alpha}$. Then it suffices to define $\alpha \sim \beta$ if $\alpha \cap \beta \neq \emptyset$ (as $\alpha \nsim \beta$ implies that Y_α and Y_β depend on disjoint sets of variables ξ_σ).

Remark 9. Theorem 7 remains valid after weakening the independence assumption to a form of negative correlation: it suffices if $\mathbb{E}(\prod_{i \in [s]} Y_{\alpha_i}) \leq \prod_{i \in [s]} \mathbb{E}Y_{\alpha_i}$ for all $(\alpha_1, \ldots, \alpha_s) \in \mathcal{I}^s$ satisfying $\alpha_i \nsim \alpha_j$ for $i \neq j$. For example, writing $\alpha \sim \beta$ if $\alpha \cap \beta \neq \emptyset$, it is not hard to check that this weaker condition holds for variables of form $Y_\alpha = w_\alpha \mathbb{1}_{\{\alpha \in \mathcal{H}_m\}}$, where the uniform model $\mathcal{H}_m = \mathcal{H}[V_m(\mathcal{H})]$ is defined as in Section 3.5.

Remark 10. Replacing the assumption $\sum_{\alpha \in \mathcal{I}} \mathbb{E}Y_\alpha \leq \mu$ of Theorem 7 with $\sum_{\alpha \in \mathcal{I}} \lambda_\alpha \leq \mu$ and $\min_{\alpha \in \mathcal{I}} \lambda_\alpha \geq 0$, the correlation condition of Remark 9 can be further weakened to $\mathbb{E}(\prod_{i \in [s]} Y_{\alpha_i}) \leq \prod_{i \in [s]} \lambda_{\alpha_i}$.

Remark 11. Note that inequality (14) implies $\varphi(\varepsilon) \geq \varepsilon^2/[2(1 + \varepsilon/3)] \geq \min\{\varepsilon^2, \varepsilon\}/3$ for $\varepsilon \geq 0$.

Remarks 9–10 suggest that the proof of Theorem 7 is fairly robust (it exploits independence only in a limited way; see also the discussion in [35] and the proof of Lemma 4.5 in [33]).

2.2. The BK-inequality

In this subsection we state a convenient consequence of the BK-inequality of van den Berg and Kesten [3] and Reimer [20]. As usual in this context, we consider a sample space $\Omega = \Omega_1 \times \cdots \times \Omega_M$ with finite Ω_i, and write $\omega = (\omega_1, \ldots, \omega_M) \in \Omega$. Given an event $\mathcal{E} \subseteq \Omega$ and an index set $I \subseteq [M] = \{1, \ldots, M\}$, we define

$$\mathcal{E}|_I = \big\{\omega \in \mathcal{E} \,:\, \text{for all } \pi \in \Omega \text{ we have } \pi \in \mathcal{E} \text{ whenever } \pi_j = \omega_j \text{ for all } j \in I\big\}.$$

In intuitive words, the event $\mathcal{E}|_I$ occurs if knowledge of the variables indexed by I already 'guarantees' the occurrence of $\mathcal{E}$ (note that all other variables are irrelevant for $\mathcal{E}|_I$). Given a collection $(\mathcal{E}_i)_{i \in \mathcal{C}}$ of events, for the purposes of this paper it seems easiest to introduce the convenient definition

$$\Box_{i \in \mathcal{C}} \mathcal{E}_i = \Big\{\text{there are pairwise disjoint } I_i \subseteq [M] \text{ such that } \bigcap_{i \in \mathcal{C}} \mathcal{E}_i|_{I_i} \text{ occurs}\Big\}. \tag{15}$$

The event $\Box_{i \in \mathcal{C}} \mathcal{E}_i$ intuitively states that all $\mathcal{E}_i$ 'occur disjointly', i.e., that there are *disjoint* subsets of variables which guarantee the occurrence of each event $\mathcal{E}_i$ (the definition of $\Box$ sidesteps that the usual box product $\Box$ is, in general, not associative). The general BK-inequality of Reimer [20] implies the following estimate.

Theorem 12. *Let $\mathbb{P}$ be a product measure on $\Omega = \Omega_1 \times \cdots \times \Omega_M$ with finite Ω_i. Then for any collection $(\mathcal{E}_i)_{i \in \mathcal{C}}$ of events we have*

$$\mathbb{P}\big(\Box_{i \in \mathcal{C}} \mathcal{E}_i\big) \leq \prod_{i \in \mathcal{C}} \mathbb{P}(\mathcal{E}_i). \tag{16}$$

Remark 13. For increasing events $\mathcal{E}_i$, [4] implies that inequality (16) also holds for $\mathbb{P}$ assigning equal probability to all outcomes $\omega \in \{0, 1\}^M$ with exactly m ones (as usual, an event $\mathcal{E}$ is called increasing if for all $\omega \in \mathcal{E}$ and $\pi \in \Omega$ we have $\pi \in \mathcal{E}$ whenever $\omega_j \leq \pi_j$ for all $j \in [M]$).

3. Core ideas and arguments

In this section we present our core combinatorial and probabilistic arguments in a slightly simplified setup. Our main focus is on the new proof ideas and methods (which

we believe are more useful to the reader than the theorems), so we defer applications and concrete upper tail inequalities to Sections 5–6. This organization of the paper also makes the extension to the more general setup of Section 4 more economical. Indeed, similar to the high-level proof strategy discussed in Section 1.2, the main results of this section are Theorem 15 of form $\mathbb{P}(X \geq (1+\varepsilon)\mathbb{E}X) \leq \sum_i \mathbb{P}(\neg \mathcal{E}_i)$ and Theorem 18 of form $\mathbb{P}(\neg \mathcal{E}_i) \leq \exp(-\Psi_i)$. Together they yield upper tail inequalities, and in Section 4.2 we adapt both to our more general setup.

In Section 3.1 we give a detailed proof overview, and introduce the simpler random induced subhypergraphs setup (where our main arguments and ideas are more natural). As a warm-up, in Section 3.2 we revisit existing inductive concentration methods, and reinterpret some of the underlying ideas. Section 3.3 contains our key combinatorial arguments, which hinge on 'sparsification' ideas and the BK-inequality. In Section 3.4 these arguments are complemented by probabilistic estimates, which rely on the Chernoff-type tail inequality Theorem 7. Finally, in Section 3.5 we demonstrate that our proofs are somewhat 'robust'.

3.1. Overview

3.1.1. Simplified setup: random induced subhypergraph $\mathcal{H}_p$

Our basic setup concerns random induced subhypergraphs. For a hypergraph $\mathcal{H}$ with vertex set $V(\mathcal{H})$, let $V_p(\mathcal{H})$ denote the binomial random vertex subset where each $v \in V(\mathcal{H})$ is included independently with probability p. We define the subhypergraph of $\mathcal{H}$ induced by $V_p(\mathcal{H})$ as

$$\mathcal{H}_p = \mathcal{H}[V_p(\mathcal{H})]. \tag{17}$$

Given non-negative weights $(w_f)_{f \in \mathcal{H}}$, for every $\mathcal{G} \subseteq \mathcal{H}$ we set

$$w(\mathcal{G}) = \sum_{f \in \mathcal{G}} w_f \mathbb{1}_{\{f \in \mathcal{H}_p\}}, \tag{18}$$

where our main focus is on the weighted number of induced edges $w(\mathcal{H}) = w(\mathcal{H}_p)$. The 'unweighted' case with $w_f = 1$ occurs frequently in the literature (see, e.g., [14, 22,16,35,37]), where the random variable $w(\mathcal{H}) = e(\mathcal{H}_p)$ simply counts the number of edges of $\mathcal{H}$ induced by $V_p(\mathcal{H})$. Our arguments will also carry over to the uniform variant $\mathcal{H}_m = \mathcal{H}[V_m(\mathcal{H})]$ defined in Section 3.5 (see Remark 19).

To formulate our results, we need some more notation and definitions. As usual, we write

$$\Gamma_U(\mathcal{H}) = \{f \in \mathcal{H} : U \subseteq f\}, \tag{19}$$

$$\Delta_j(\mathcal{H}) = \max_{U \subseteq V(\mathcal{H}):|U|=j} |\Gamma_U(\mathcal{H})|. \tag{20}$$

In concrete words, $\Gamma_U(\mathcal{H})$ corresponds to the set of all edges $f \in \mathcal{H}$ that contain the vertex subset $U \subseteq V(\mathcal{H})$, and $\Delta_j(\mathcal{H})$ denotes the maximum number of edges that contain j given vertices (which we think of as a 'maximum degree' parameter). Inspired by [15,19,30, 32], we now define the following two crucial *assumptions* (P') and (Pq), where $q \in \mathbb{N}$ is a parameter:

(P') Assume that $\max_{f \in \mathcal{H}} |f| \leq k$, $\max_{f \in \mathcal{H}} w_f \leq L$ and $v(\mathcal{H}) \leq N$. Define $\mu = \mathbb{E}w(\mathcal{H})$ and

$$\mu_j = \max_{U \subseteq V(\mathcal{H}):|U|=j} \sum_{f \in \Gamma_U(\mathcal{H})} p^{|f|-|U|}. \tag{21}$$

(Pq) Assume that $\Delta_q(\mathcal{H}) \leq D$.

Property (P') ensures that every edge $f \in \mathcal{H}$ has at most k vertices, that the associated edge weights satisfy $0 \leq w_f \leq L$, and that $\mathcal{H}$ contains at most $v(\mathcal{H}) \leq N$ vertices. Although we shall not assume this, our main focus is on the common case where $k + L = O(1)$ and $N = \omega(1)$ holds. Property (Pq) will be useful when $D = O(1)$ holds for $q < k$ (this is trivial for $q = k$). The key parameters μ_j intuitively quantify the 'dependencies' between the edges, and we think of them as average variants of the 'maximum degree' parameter $\Delta_j(\mathcal{H}_p)$ from (20). To see this, note that $\mathbb{P}(f \in \mathcal{H}_p \mid U \subseteq V_p(\mathcal{H})) = p^{|f|-|U|}$, so (21) equals

$$\mu_j = \max_{U \subseteq V(\mathcal{H}):|U|=j} \mathbb{E}\big(|\Gamma_U(\mathcal{H}_p)| \mid U \subseteq V_p(\mathcal{H})\big). \tag{22}$$

In concrete words, after conditioning on the presence of any vertex subset $U \subseteq V_p(\mathcal{H})$ of size $|U| = j$, the expected number of edges in $\mathcal{H}_p$ that contain U is at most μ_j (for this reason, μ_j can be interpreted as the 'maximum average effect' of any j vertices or variables, see also [19,32]). For example, if the edges of the k-uniform hypergraph $\mathcal{H} = \mathcal{H}_n$ correspond to k-term arithmetic progressions, then we can take $V(\mathcal{H}) = [n]$, $N = n$, $L = 1$, $\mu = \Theta(n^2 p^k)$ and $\mu_j = \Theta(n^{2-j}p^{k-j})$ for $1 \leq j \leq q = 2$ (note that $\Delta_2(\mathcal{H}) = O(1)$ holds).

3.1.2. The basic form of our tail estimates

In this subsection we discuss the approximate form of our upper tail estimates. As we shall see in Section 3.2, for hypergraphs $\mathcal{H}$ with $\Delta_q(\mathcal{H}) \leq D$ the usual inductive concentration of measure methods [19,15,32] yield basic inequalities of the following form (omitting several technicalities). Given positive parameters $(R_j)_{1 \leq j \leq q}$ with $R_q \geq D$, for every $\varepsilon > 0$ there are positive constants $a = a(\varepsilon, k)$ and $b = b(k)$ such that roughly

$$\mathbb{P}(e(\mathcal{H}_p) \geq (1+\varepsilon)\mu) \leq \exp\big(-a\mu/R_1\big) + \sum_{1 \leq j < q} \left(\frac{\mu_j}{R_j}\right)^{bR_j/R_{j+1}}, \tag{23}$$

say (see (76) of Claim 33; the freedom of choosing the parameters $(R_j)_{1 \le j \le q}$ is part of the method, though one naturally aims at roughly $\mu/R_1 \approx R_j/R_{j+1}$). The 'prepackaged versions' of these inequalities usually assume that the parameters satisfy roughly $\mu/R_1 \ge \lambda$ and $R_j \ge \max\{2\mu_j, \lambda R_{j+1}\}$ (see, e.g., Theorem 4.2 in [32] or Theorem 3.10 in [15]). In this case there are positive constants $c = c(a,b)$ and $C = C(q)$ such that

$$\mathbb{P}(e(\mathcal{H}_p) \ge (1+\varepsilon)\mu) \le C \exp(-c\lambda). \tag{24}$$

The punchline of this paper is that we can often improve the exponential decay of (24) if stronger bounds than $R_j \ge 2\mu_j$ hold. For example, setting $\lambda \approx \mu^{1/q}$ and $R_j \approx \lambda^{q-j}$ (similar to, e.g., the proof of Corollary 6.3 in [32] or Theorem 2.1 in [31]), in the applications of Section 6.1 we naturally arrive at bounds of form

$$\max_{1 \le j < q} \frac{\mu_j}{R_j} \approx \max_{1 \le j < q} \frac{\mu_j}{\mu^{(q-j)/j}} = O(p^\alpha). \tag{25}$$

It might be instructive to check that (25) holds with $\alpha = 1/2$ for k-term arithmetic progressions with $k \ge 3$. Intuitively, replacing $R_j \ge 2\mu_j$ by the stronger assumption (25) improves the exponential decay of the sum-terms in (23) by a factor of roughly $\log(1/p)$ for small p. Hence the $\exp(-a\mu/R_1)$ term in (23) is the main obstacle for improving inequality (24). Here our new 'sparsification' based approach is key: after some technical work it essentially allows us to replace R_1 by

$$Q_1 = \max\{R_1/\log(1/p),\, B\},$$

where $B \ge 1$ is some constant (of course, we later need to be a bit careful when $p \approx 1$ holds, e.g., replacing $\log(1/p)$ with $\log(e/p)$, say). More concretely, assuming (25), for $\mu/R_1 \ge \lambda$, $R_j \ge \lambda R_{j+1}$ and $p = o(1)$ we eventually arrive (ignoring some technicalities) at a bound that is roughly of the form

$$\mathbb{P}(e(\mathcal{H}_p) \ge (1+\varepsilon)\mu) \le \exp(-a\mu/Q_1) + \sum_{1 \le j < q} \left[\left(\frac{\mu_j}{R_j} \right)^{bR_j/R_{j+1}} + \left(\frac{\mu_j}{R_j} \right)^{a\mu/R_1} \right]$$

$$\le C \exp\left(-c \min\{\mu,\, \lambda \log(1/p)\}\right), \tag{26}$$

with $c = c(a,b,\alpha,B) > 0$ and $C = q$ (see (80) of Theorem 34). In words, (26) essentially adds a logarithmic factor to the exponent of the classical bound (24). This improvement of (23)–(24) is conceptually important, since in several interesting examples the resulting estimate (26) is qualitatively best possible (see Section 6.1).

3.1.3. Sketch of the argument

In this subsection we expand on the high-level proof strategy from Section 1.2, and give a rough sketch of our main *combinatorial* line of reasoning (the full details are

deferred to Sections 3.2–3.4 and 4.2). As we shall argue in Section 3.2, at the conceptual heart of the usual inductive concentration approaches lies the following combinatorial 'degree' event $\mathcal{D}_j$: $\Delta_{j+1}(\mathcal{H}_p) \leq R_{j+1}$ implies $\Delta_j(\mathcal{H}_p) \leq R_j$. Given a hypergraph $\mathcal{H}$ with $\Delta_q(\mathcal{H}) \leq R_q$, for the induced number of edges $e(\mathcal{H}_p)$ the basic idea is that an iterative application of the events $\mathcal{D}_{q-1} \cap \cdots \cap \mathcal{D}_1$ reduces the upper tail problem to

$$
\begin{aligned}
\mathbb{P}(e(\mathcal{H}_p) \geq (1+\varepsilon)\mu) &\leq \mathbb{P}(e(\mathcal{H}_p) \geq (1+\varepsilon)\mu \text{ and } \Delta_q(\mathcal{H}_p) \leq R_q) \\
&\leq \mathbb{P}(e(\mathcal{H}_p) \geq (1+\varepsilon)\mu \text{ and } \Delta_1(\mathcal{H}_p) \leq R_1) + \sum_{1 \leq j < q} \mathbb{P}(\neg \mathcal{D}_j).
\end{aligned}
\tag{27}
$$

It turns out that all the probabilities on the right hand side of (27) can easily be estimated by the concentration inequality Theorem 7 (see Claim 14 and Theorem 18), which eventually yields a variant of the upper tail estimate (23). As before, the crux is that smaller values of the 'maximum degree' R_1 translate into better tail estimates. To surpass the usual inductive approaches, similar to (26) our plan is thus to reduce the 'degree bound' R_1 down to Q_1, and here our new 'sparsification idea' will be key, achieving this 'degree reduction' by deleting up to $\varepsilon\mu/2$ edges.

Our starting point is the observation that, via Theorem 7, we can strengthen the degree event $\mathcal{D}_j$ to all subhypergraphs $\mathcal{G} \subseteq \mathcal{H}_p$ (see Claim 14 and Theorem 18). Namely, let $\mathcal{D}_j^+$ denote the event that $\Delta_{j+1}(\mathcal{G}) \leq Q_{j+1}$ implies $\Delta_j(\mathcal{G}) \leq Q_j$ for all $\mathcal{G} \subseteq \mathcal{H}_p$. A crucial aspect of our argument is that the events $\mathcal{D}_j, \mathcal{D}_j^+$ work hand in hand with the following combinatorial 'sparsification' event $\mathcal{E}_q$: $\Delta_1(\mathcal{H}_p) \leq R_1$ implies existence of a subhypergraph $\mathcal{G} \subseteq \mathcal{H}_p$ with $e(\mathcal{H}_p \setminus \mathcal{G}) \leq \varepsilon\mu/2$ and $\Delta_{q-1}(\mathcal{G}) \leq Q_{q-1}$ (tacitly assuming $q \geq 2$). Intuitively, $\mathcal{E}_q$ states that the deletion of 'few' edges reduces the degree $\Delta_{q-1}(\mathcal{H}_p)$ down to $\Delta_{q-1}(\mathcal{G}) \leq Q_{q-1}$.

The basic combinatorial idea of our approach is roughly as follows (see Section 3.3 for the more involved details). We first (i) obtain the coarse degree bound $\Delta_1(\mathcal{H}_p) \leq R_1$ via an iterative application of the degree events $\mathcal{D}_{q-1} \cap \cdots \cap \mathcal{D}_1$, then (ii) exploit the sparsification event $\mathcal{E}_q$ to find a subhypergraph $\mathcal{G} \subseteq \mathcal{H}_p$ with $e(\mathcal{H}_p \setminus \mathcal{G}) \leq \varepsilon\mu/2$ and $\Delta_{q-1}(\mathcal{G}) \leq Q_{q-1}$, and finally (iii) deduce the improved degree bound $\Delta_1(\mathcal{G}) \leq Q_1$ via an iterative application of the degree events $\mathcal{D}_{q-2}^+ \cap \cdots \cap \mathcal{D}_1^+$. Taking into account that we obtain $\mathcal{G} \subseteq \mathcal{H}_p$ by deleting up to $\varepsilon\mu/2$ edges, for hypergraphs $\mathcal{H}$ with $\Delta_q(\mathcal{H}) \leq R_q$ we eventually arrive at

$$
\begin{aligned}
\mathbb{P}(e(\mathcal{H}_p) \geq (1+\varepsilon)\mu) &\leq \mathbb{P}(e(\mathcal{G}) \geq (1+\varepsilon/2)\mu \text{ and } \Delta_1(\mathcal{G}) \leq Q_1 \text{ for some } \mathcal{G} \subseteq \mathcal{H}_p) \\
&\quad + \sum_{1 \leq j < q} \mathbb{P}(\neg \mathcal{D}_j) + \mathbb{P}(\neg \mathcal{E}_q) + \sum_{1 \leq j < q-1} \mathbb{P}(\neg \mathcal{D}_j^+).
\end{aligned}
\tag{28}
$$

The crux is that we can again obtain good tail estimates for $\mathbb{P}(e(\mathcal{G}) \geq (1+\varepsilon/2)\mu \cdots)$ and $\mathbb{P}(\neg \mathcal{D}_j) + \mathbb{P}(\neg \mathcal{D}_j^+)$ via Theorem 7 (see Claim 14 and Theorem 18), so in (28) it remains to bound $\mathbb{P}(\neg \mathcal{E}_q)$.

To estimate the probability that the sparsification event $\mathcal{E}_q$ fails, we shall rely on combinatorial arguments and the BK-inequality, developing a 'maximal matching' based idea from [35]. Simplifying slightly (see Section 3.3.1 for the full details), for any vertex set $U \subseteq V(\mathcal{H})$ with $|U| = q - 1$ we tentatively call $\mathcal{K}_U \subseteq \Gamma_U(\mathcal{H}) = \{f \in \mathcal{H} : U \subseteq f\}$ with $|\mathcal{K}_U| = r$ an *r-star*, where we set $r = Q_{q-1}$ for brevity. The basic idea is to take a maximal vertex disjoint collection of r-stars in $\mathcal{H}_p$, which we denote by $\mathcal{M}$ (to clarify: the edges from any two distinct r-stars $\mathcal{K}_U, \mathcal{K}_W \in \mathcal{M}$ are vertex disjoint), and remove all edges $f \in \mathcal{H}_p$ that are incident to $\mathcal{M}$, i.e., which share at least one vertex with some r-star from $\mathcal{M}$. Denoting the resulting subhypergraph by $\mathcal{G} \subseteq \mathcal{H}_p$, using maximality of $\mathcal{M}$ it is not difficult to argue that $\Delta_{q-1}(\mathcal{G}) < r = Q_{q-1}$ holds (otherwise we could add another r-star to $\mathcal{M}$). Furthermore, by construction the deleted number of edges is at most

$$e(\mathcal{H}_p \setminus \mathcal{G}) \leq \sum_{\mathcal{K}_U \in \mathcal{M}} \sum_{f \in \mathcal{K}_U} \sum_{v \in f} |\Gamma_{\{v\}}(\mathcal{H}_p)| \leq |\mathcal{M}| \cdot r \cdot k \cdot \Delta_1(\mathcal{H}_p). \tag{29}$$

Since the event $\mathcal{E}_q$ presupposes $\Delta_1(\mathcal{H}_p) \leq R_1$, we thus see that $|\mathcal{M}| \leq \varepsilon\mu/(2rkR_1)$ implies $|\mathcal{H}_p \setminus \mathcal{G}| \leq \varepsilon\mu/2$. It remains to estimate the probability that $|\mathcal{M}|$ is big, and here we shall exploit the fact that the r-stars $\mathcal{K}_U \in \mathcal{M}$ satisfy two properties: they (i) are pairwise vertex disjoint, and (ii) each 'guarantee' that $|\Gamma_U(\mathcal{H}_p)| \geq r$ holds. Intuitively, the point of (i) and (ii) is that $|\mathcal{M}|$ events of from $|\Gamma_U(\mathcal{H}_p)| \geq r$ 'occur disjointly' in the sense of Section 2.2, which allows us to bring the BK-inequality (16) into play. Indeed, by analyzing a $\square$-based moment of $\sum_{U:|U|=q-1} \mathbb{1}_{\{|\Gamma_U(\mathcal{H}_p)| \geq r\}}$, we then eventually obtain sufficiently good estimates for $\mathbb{P}(\neg\mathcal{E}_q)$, as desired (see the proofs of Lemma 16 and inequality (48) of Theorem 18).

As the reader can guess, the actual details are more involved. For example, instead of just $\mathcal{E}_q$ for $\Delta_{q-1}(\cdot)$, we also need to consider similar sparsification events for the others degrees $\Delta_j(\cdot)$ with $1 \leq j < q$. In fact, analogous to $\mathcal{D}_j^+$, these events must moreover apply to all subhypergraphs $\mathcal{G} \subseteq \mathcal{H}_p$ simultaneously (see $\mathcal{E}_{j,\ell}(x, r, y, z)$ defined in Section 3.3). Furthermore, due to technical reasons, the decomposition (28) requires some extra bells and whistles (see (33) of Theorem 15). Finally, we have also ignored how Theorem 7 and the BK-inequality (16) eventually allow us to convert the decompositions (27)–(28) into concrete upper tail inequalities of form (23) and (26); see Sections 3.3.1, 3.4, 4.2 and 5.3 for these technical calculations.

3.2. Inductive concentration proofs revisited

The goal of this warm-up section is to reinterpret the classical inductive concentration proofs from [15,19,32] using the following 'degree intuition': an (improved) upper bound for $\Delta_{j+1}(\mathcal{H}_p)$ and $\Delta_1(\mathcal{H}_p)$ translates into an improved upper tail estimate for $\Delta_j(\mathcal{H}_p)$ and $w(\mathcal{H}_p)$, respectively. We exemplify this with the following claim, which is usually stated for $\mathcal{G} = \mathcal{H}_p$ only (the proof of is based on routine applications of Theorem 7, and

thus deferred to Section 3.4). We find inequalities (30)–(31) below remarkable, since they intuitively yield bounds for all subhypergraphs $\mathcal{G} \subseteq \mathcal{H}_p$ *without* taking a union bound.

Claim 14. *Given $\mathcal{H}$, assume that (P') holds. Then for all $t, x, y > 0$ and $1 \leq j < k$ we have*

$$\mathbb{P}\big(w(\mathcal{G}) \geq \mu + t \text{ and } \Delta_1(\mathcal{G}) \leq y \text{ for some } \mathcal{G} \subseteq \mathcal{H}_p\big) \leq \left(1 + \frac{t}{\mu}\right)^{-t/(4Lky)}, \qquad (30)$$

$$\mathbb{P}\big(\Delta_j(\mathcal{G}) \geq \mu_j + x \text{ and } \Delta_{j+1}(\mathcal{G}) \leq y \text{ for some } \mathcal{G} \subseteq \mathcal{H}_p\big) \leq N^j \left(1 + \frac{x}{\mu_j}\right)^{-x/(4ky)}. \tag{31}$$

Now, by a straightforward iterative degree argument similar to (27), we obtain the simple estimate

$$\begin{aligned} \mathbb{P}\big(&w(\mathcal{G}) \geq \mu + t \text{ and } \Delta_q(\mathcal{G}) \leq R_q \text{ for some } \mathcal{G} \subseteq \mathcal{H}_p\big) \\ &\leq \mathbb{P}\big(w(\mathcal{G}) \geq \mu + t \text{ and } \Delta_1(\mathcal{G}) \leq R_1 \text{ for some } \mathcal{G} \subseteq \mathcal{H}_p\big) \\ &\quad + \sum_{1 \leq j < q} \mathbb{P}\big(\Delta_j(\mathcal{G}) > R_j \text{ and } \Delta_{j+1}(\mathcal{G}) \leq R_{j+1} \text{ for some } \mathcal{G} \subseteq \mathcal{H}_p\big). \end{aligned} \tag{32}$$

Restricting to the special case $w(\mathcal{H}_p)$, using Claim 14 it turns out that inequality (32) is essentially equivalent to the basic induction of Janson and Ruciński [15] (see the proof of Theorem 3.10 in [15]), which in turn qualitatively recovers the upper tail part of Kim and Vu [19] (see Section 5 of [15,13]). The iterative point of view (32) is somewhat more flexible than induction, making the arguments subjectively easier to modify (as there is no need to formulate a suitable induction hypothesis). Estimates for all subhypergraphs $\mathcal{G} \subseteq \mathcal{H}_p$ also make room for additional combinatorial arguments, which is crucial for the purposes of this paper.

3.3. Combinatorial sparsification: degree reduction by deletion

In this section we introduce our key combinatorial arguments, which eventually allow us to obtain improved upper tail estimates by 'sparsifying' $\mathcal{H}_p$, i.e., deleting edges from $\mathcal{H}_p$. Loosely speaking, via this *sparsification idea* we can effectively ignore certain 'exceptional' edges from $\mathcal{H}_p$ (which contain vertices with extremely high degree, say). For the purpose of this paper, we encapsulate this heuristic idea with the definition below. In intuitive words, for $\ell = 1$ the 'sparsification' event $\mathcal{E}_{j,1}(x, r, y, z)$ essentially ensures that every $\mathcal{G} \subseteq \mathcal{H}_p$ with bounded $\Delta_{j+1}(\mathcal{G})$ and $\Delta_1(\mathcal{G})$ contains a large subhypergraph $\mathcal{J} \subseteq \mathcal{G}$ with small $\Delta_j(\mathcal{J})$.

Definition *(Sparsification event).* Let $\mathcal{E}_{j,\ell}(x, r, y, z)$ denote the event that for every $\mathcal{G} \subseteq \mathcal{H}_p$ with $\Delta_{j+1}(\mathcal{G}) \leq y$ and $\Delta_\ell(\mathcal{G}) \leq z$ there is $\mathcal{J} \subseteq \mathcal{G}$ with $\Delta_j(\mathcal{J}) \leq x$ and $e(\mathcal{G} \setminus \mathcal{J}) \leq r$.

Here one conceptual difference to the 'deletion lemma' of Rödl and Ruciński [22,14] is that our focus is on 'local properties' such as degrees (somewhat in the spirit of [29]), and not on 'global properties' such as subgraph counts. Furthermore, we are deleting edges from $\mathcal{H}_p = \mathcal{H}[V_p(\mathcal{H})]$, whereas the classical approach corresponds to deleting vertices from $V_p(\mathcal{H}) = E(G_{n,p})$, say.

With $\mathcal{E}_{j,1}(x, r, y, z)$ in hand, we now refine[4] the basic estimate (32) via the strategy outlined in Section 3.1.3 (see also (28) therein). We believe that the ideas used in the proof of Theorem 15 below are more important than its concrete statement (which is optimized for the purposes of this paper). Here one new ingredient is the edge deletion of the sparsification events in $\mathbb{P}_{j,3,\ell}$ of (36), which allows us to decrease certain maximum degrees. The total weight of the deleted edges can be as large as $t/2$, which is the reason why in (33) we need to relax $w(\mathcal{G}) \geq \mu + t$ to $w(\mathcal{G}) \geq \mu + t/2$. In later applications we shall use $S_j \approx R_j/s$ with $s = \omega(1)$, and then the parametrization $Q_j = \max\{S_j, D_j\}$ allows us to easily deal with $S_j = o(1)$ border cases. The indicators in (35)–(36) can safely be ignored on first reading (they mainly facilitate certain technical estimates). A key aspect of (33) is that we intuitively replace $\Delta_1(\mathcal{G}) \leq R_1$ of (32) with $\Delta_1(\mathcal{G}) \leq \min\{Q_1, R_1\}$, which by the discussion of Section 3.2 is crucial for obtaining improved tail estimates (see also Theorem 18).

Theorem 15 *(Combinatorial decomposition of the upper tail). Given $\mathcal{H}$ with $1 \leq q \leq k$, assume that (P') holds. Suppose that $t > 0$. Given positive $(D_j)_{1 \leq j \leq q}$, $(R_j)_{1 \leq j < q}$ and $(S_j)_{1 \leq j < q}$, define $R_q = Q_q = D_q$ and $Q_j = \max\{S_j, D_j\}$ for $1 \leq j < q$. Then we have*

$$\mathbb{P}\left(w(\mathcal{G}) \geq \mu + t \text{ and } \Delta_q(\mathcal{G}) \leq D_q \text{ for some } \mathcal{G} \subseteq \mathcal{H}_p\right)$$
$$\leq \mathbb{P}\left(w(\mathcal{G}) \geq \mu + t/2 \text{ and } \Delta_1(\mathcal{G}) \leq \min\{Q_1, R_1\} \text{ for some } \mathcal{G} \subseteq \mathcal{H}_p\right) \qquad (33)$$
$$+ \sum_{1 \leq j < q} \left[\mathbb{P}_{j,1} + \mathbb{P}_{j,2} + \mathbb{P}_{j,3,1}\right],$$

where

$$\mathbb{P}_{j,1} = \mathbb{P}\left(\Delta_j(\mathcal{G}) > R_j \text{ and } \Delta_{j+1}(\mathcal{G}) \leq R_{j+1} \text{ for some } \mathcal{G} \subseteq \mathcal{H}_p\right), \qquad (34)$$

$$\mathbb{P}_{j,2} = \mathbb{1}_{\{Q_j < R_j \text{ and } Q_{j+1} > D_{j+1}\}} \mathbb{P}\left(\Delta_j(\mathcal{G}) > Q_j \text{ and } \Delta_{j+1}(\mathcal{G}) \leq S_{j+1} \text{ for some } \mathcal{G} \subseteq \mathcal{H}_p\right), \qquad (35)$$

$$\mathbb{P}_{j,3,\ell} = \mathbb{1}_{\{Q_j < R_j \text{ and } Q_{j+1} = D_{j+1}\}} \mathbb{P}\left(\neg \mathcal{E}_{j,\ell}(Q_j, t/(2Lq), D_{j+1}, R_\ell)\right). \qquad (36)$$

The combinatorial proof proceeds in two sparsification rounds. In the first round we use our usual iterative degree argument to deduce that $\Delta_q(\mathcal{G}) \leq R_q$ implies $\Delta_j(\mathcal{G}) \leq R_j$ for all $1 \leq j \leq q$. We start the second round with the sparsification event, by deleting edges

---

[4] Note that by setting $D_j = R_j = S_j$ the indicators in (35)–(36) are zero, so (33) qualitatively reduces to (32).

such that $\mathcal{J} \subseteq \mathcal{G}$ satisfies $\Delta_{q-1}(\mathcal{J}) \leq Q_{q-1}$ (tacitly assuming $Q_{q-1} < R_{q-1}$, say). The idea is that our usual iterative degree argument should then allow us to deduce that $\Delta_{j+1}(\mathcal{J}) \leq Q_{j+1}$ implies $\Delta_j(\mathcal{J}) \leq Q_j$ for all $1 \leq j < q - 1$. Unfortunately, our later probabilistic estimates break down if the parameter Q_{j+1} is 'too small'. With foresight we thus use our alternative 'degree reduction' argument whenever $Q_{j+1} = D_{j+1}$ holds, i.e., we again delete edges.

Proof of Theorem 15. Inequality (33) is trivial for $q = 1$ (since $R_1 = Q_1 = D_1$). For $q \geq 2$ the plan is to show that properties (a)–(d) below *deterministically* imply that $w(\mathcal{G}) < \mu + t$ for every $\mathcal{G} \subseteq \mathcal{H}_p$ with $\Delta_q(\mathcal{G}) \leq D_q$. Using a union bound argument this then completes the proof (it is routine to check that (a)–(d) correspond to the complements of the events on the right hand side of (33), since $Q_{j+1} > D_{j+1}$ implies $S_{j+1} = Q_{j+1}$). Turning to the details, we henceforth assume that the following properties hold for all $\mathcal{G} \subseteq \mathcal{H}_p$ and $1 \leq j < q$:

(a) $\Delta_1(\mathcal{G}) \leq \min\{Q_1, R_1\}$ implies $w(\mathcal{G}) < \mu + t/2$,
(b) $\Delta_{j+1}(\mathcal{G}) \leq R_{j+1}$ implies $\Delta_j(\mathcal{G}) \leq R_j$,
(c) if $Q_j < R_j$ and $Q_{j+1} > D_{j+1}$, then $\Delta_{j+1}(\mathcal{G}) \leq Q_{j+1}$ implies $\Delta_j(\mathcal{G}) \leq Q_j$, and
(d) if $Q_j < R_j$ and $Q_{j+1} = D_{j+1}$, then $\Delta_{j+1}(\mathcal{G}) \leq Q_{j+1}$ and $\Delta_1(\mathcal{G}) \leq R_1$ implies existence of $\mathcal{J} \subseteq \mathcal{G}$ with $\Delta_j(\mathcal{J}) \leq Q_j$ and $e(\mathcal{G} \setminus \mathcal{J}) \leq t/(2Lq)$.

For the remaining deterministic argument we fix $\mathcal{G} \subseteq \mathcal{H}_p$ with $\Delta_q(\mathcal{G}) \leq D_q$, and claim that we can construct a hypergraph sequence $\mathcal{G} = \mathcal{J}_q \supseteq \cdots \supseteq \mathcal{J}_1$ such that

$$\Delta_i(\mathcal{J}_j) \leq \begin{cases} R_i, & \text{if } 1 \leq i < j, \\ \min\{Q_i, R_i\}, & \text{if } j \leq i \leq q, \end{cases} \tag{37}$$

$$e(\mathcal{J}_{j+1} \setminus \mathcal{J}_j) \leq t/(2Lq). \tag{38}$$

With this sequence in hand, using (38) we have

$$w(\mathcal{J}_{j+1} \setminus \mathcal{J}_j) = \sum_{f \in \mathcal{J}_{j+1} \setminus \mathcal{J}_j} w_f \leq \big(\max_{f \in \mathcal{J}_{j+1} \setminus \mathcal{J}_j} w_f\big) \cdot e(\mathcal{J}_{j+1} \setminus \mathcal{J}_j) \leq L \cdot t/(2Lq) = t/(2q),$$

which together with $\Delta_1(\mathcal{J}_1) \leq \min\{Q_1, R_1\}$ of (37) and (a) then yields

$$w(\mathcal{G}) = w(\mathcal{J}_1) + \sum_{1 \leq j < q} w(\mathcal{J}_{j+1} \setminus \mathcal{J}_j) < (\mu + t/2) + (q - 1) \cdot t/(2q) \leq \mu + t. \tag{39}$$

It thus remains to construct $\mathcal{G} = \mathcal{J}_q \supseteq \cdots \supseteq \mathcal{J}_1$ with the claimed properties. For the base case $\mathcal{G} = \mathcal{J}_q$, using $\Delta_q(\mathcal{J}_q) = \Delta_q(\mathcal{G}) \leq D_q = R_q$ repeated applications of (b) yield that $\Delta_i(\mathcal{J}_q) \leq R_i$ for all $1 \leq i \leq q$, so (37) holds since $\Delta_q(\mathcal{J}_q) \leq R_q = \min\{R_q, Q_q\}$. Given $\mathcal{J}_{j+1}$ with $1 \leq j < q$, our construction of $\mathcal{J}_j \subseteq \mathcal{J}_{j+1}$ distinguishes several cases; in view of $\Delta_i(\mathcal{J}_j) \leq \Delta_i(\mathcal{J}_{j+1})$ it clearly suffices to check (37) for $\Delta_j(\mathcal{J}_j)$ only.

If $Q_j \geq R_j$, then we set $\mathcal{J}_j = \mathcal{J}_{j+1}$, which satisfies $\Delta_j(\mathcal{J}_j) = \Delta_j(\mathcal{J}_{j+1}) \leq R_j = \min\{Q_j, R_j\}$ by (37).

If $Q_j < R_j$ and $Q_{j+1} > D_{j+1}$, then we set $\mathcal{J}_j = \mathcal{J}_{j+1}$, which by (37) satisfies $\Delta_{j+1}(\mathcal{J}_j) = \Delta_{j+1}(\mathcal{J}_{j+1}) \leq Q_{j+1}$. Hence (c) implies $\Delta_j(\mathcal{J}_j) \leq Q_j = \min\{Q_j, R_j\}$.

Finally, if $Q_j < R_j$ and $Q_{j+1} = D_{j+1}$, then by (37) we have $\Delta_{j+1}(\mathcal{J}_{j+1}) \leq Q_{j+1}$ and $\Delta_1(\mathcal{J}_{j+1}) \leq R_1$. Hence (d) implies existence of $\mathcal{J}_j \subseteq \mathcal{J}_{j+1}$ satisfying $\Delta_j(\mathcal{J}_j) \leq Q_j = \min\{Q_j, R_j\}$ and $e(\mathcal{J}_{j+1} \setminus \mathcal{J}_j) \leq t/(2Lq)$, completing the proof.   □

The above proof demonstrates that estimates for all subhypergraphs $\mathcal{G} \subseteq \mathcal{H}_p$ are extremely powerful along with combinatorial arguments. It seems likely that the above sparsification approach can be sharpened in specific applications, i.e., that there is room for alternative (ad-hoc) arguments which apply the 'degree reduction' idea differently. For example, in [35] the degrees are iteratively reduced by a factor of two, say (replacing the finite sum in (39) by a convergent geometric series). In [27] the iterative argument also takes 'trivial' upper bounds for the $\Delta_j(\mathcal{H})$ into account (which can be smaller than R_j or Q_j).

3.3.1. A combinatorial local deletion argument

The goal of this subsection is to estimate $\mathbb{P}(\neg\mathcal{E}_{j,1}(x, r, y, z))$, i.e., the probability that our 'sparsification' event fails. As indicated in Section 3.1.3, our proof uses a *maximal matching based idea* which relies on combinatorial arguments and the BK-inequality. The following auxiliary event $\mathcal{D}_{U,x,y}$ intuitively states that, in $\mathcal{H}_p$, the vertex set U is the centre of a 'star' with at least x spikes (satisfying some degree constraint).

Definition *(Auxiliary degree event).* Let $\mathcal{D}_{U,x,y}$ denote the event that there is $\mathcal{K} \subseteq \Gamma_U(\mathcal{H}_p)$ with $|\mathcal{K}| \geq x$ and $\Delta_{|U|+1}(\mathcal{K}) \leq y$.

To put this definition into our 'all subhypergraphs' context, note that $\neg\mathcal{D}_{U,x,y}$ implies $|\Gamma_U(\mathcal{G})| < x$ for all $\mathcal{G} \subseteq \mathcal{H}_p$ with $\Delta_{|U|+1}(\mathcal{G}) \leq y$. It might also be instructive to note that a union bound argument yields

$$\mathbb{P}\big(\Delta_j(\mathcal{G}) \geq x \text{ and } \Delta_{j+1}(\mathcal{G}) \leq y \text{ for some } \mathcal{G} \subseteq \mathcal{H}_p\big) \leq \sum_{U \subseteq V(\mathcal{H}):|U|=j} \mathbb{P}(\mathcal{D}_{U,x,y}). \qquad (40)$$

The next result relates the auxiliary event $\mathcal{D}_{U,x,y}$ with the sparsification event $\mathcal{E}_{j,1}(x, r, y, z)$. For example, $\sum_U \mathbb{P}(\mathcal{D}_{U,x,y}) \leq B^{-x/y}$ translates into $\mathbb{P}(\neg\mathcal{E}_{j,1}(x, r, y, z)) \leq B^{-r/(kyz)}$ by inequality (41).

Lemma 16 *(Auxiliary result for the sparsification event). Given $\mathcal{H}$, assume that $\max_{f \in \mathcal{H}} |f| \leq k$ holds. Then for all $x, r, y, z > 0$ and $1 \leq j < k$ we have*

$$\mathbb{P}\big(\neg\mathcal{E}_{j,1}(x, r, y, z)\big) \leq \left(\sum_{U \subseteq V(\mathcal{H}):|U|=j} \mathbb{P}(\mathcal{D}_{U,x,y}) \right)^{\lceil r/(k\lceil x\rceil z)\rceil}. \qquad (41)$$

Remark 17. Inequality (41) remains valid after dividing the right hand side by $\lceil r/(k\lceil x \rceil z) \rceil!$.

The proof of Lemma 16 develops a combinatorial idea from [35], which in turn was partially inspired by [28,14]. We call $(U, \mathcal{K}_U)$ an (j, x, y)-*star in* $\mathcal{G}$ if $U \subseteq V(\mathcal{G})$ and $\mathcal{K}_U \subseteq \Gamma_U(\mathcal{G}) = \{ f \in \mathcal{G} : U \subseteq f \}$ satisfy $|U| = j$, $|\mathcal{K}_U| = \lceil x \rceil$ and $\Delta_{j+1}(\mathcal{K}_U) \le y$. Note that we allow for overlaps of the edges $f, g \in \mathcal{K}_U$ outside of the 'centre' U. Writing $\mathcal{S}_{j,x,y}(\mathcal{G})$ for the collection of all (j, x, y)-stars in $\mathcal{G}$, we define $M_{j,x,y}(\mathcal{G})$ as the size of the largest $\mathcal{M} \subseteq \mathcal{S}_{j,x,y}(\mathcal{G})$ satisfying $V(\mathcal{K}_U) \cap V(\mathcal{K}_W) = \emptyset$ for all distinct $(U, \mathcal{K}_U), (W, \mathcal{K}_W) \in \mathcal{M}$. In intuitive words, $M_{j,x,y}(\mathcal{G})$ denotes the size of the 'largest (j, x, y)-star matching' in $\mathcal{G}$, i.e., vertex-disjoint collection of stars. We are now ready to follow the strategy sketched in Section 3.1.3 (see also (29) therein).

Proof of Lemma 16. Let $\tilde{r} = r/(k\lceil x \rceil z)$ and $R = \lceil \tilde{r} \rceil$. We first assume that $M_{j,x,y}(\mathcal{H}_p) \le \tilde{r}$ holds, and claim that this implies the occurrence of $\mathcal{E}_{j,1}(x, r, y, z)$. For any $\mathcal{G} \subseteq \mathcal{H}_p$ with $\Delta_{j+1}(\mathcal{G}) \le y$ and $\Delta_1(\mathcal{G}) \le z$, it clearly suffices to show that there is $\mathcal{J} \subseteq \mathcal{G}$ with $\Delta_j(\mathcal{J}) \le x$ and $e(\mathcal{G} \setminus \mathcal{J}) \le r$. Let $\mathcal{M} \subseteq \mathcal{S}_{j,x,y}(\mathcal{G})$ attain the maximum in the definition of $M_{j,x,y}(\mathcal{G})$. We then remove all edges $f \in \mathcal{G}$ which overlap some star $(U, \mathcal{K}_U) \in \mathcal{M}$, where overlap means that $f \cap g \ne \emptyset$ for some edge $g \in \mathcal{K}_U$. We denote the resulting subhypergraph by $\mathcal{J} \subseteq \mathcal{G}$. Using $\Delta_{j+1}(\mathcal{J}) \le \Delta_{j+1}(\mathcal{G}) \le y$ and maximality of $\mathcal{M}$, we then infer $\Delta_j(\mathcal{J}) \le \lceil x \rceil - 1 < x$ (because otherwise we could add another (j, x, y)-star to $\mathcal{M}$). Furthermore, since $|\mathcal{M}| = M_{j,x,y}(\mathcal{G}) \le M_{j,x,y}(\mathcal{H}_p) \le \tilde{r}$ and $\Delta_1(\mathcal{G}) \le z$, by construction the number of deleted edges is at most

$$e(\mathcal{G} \setminus \mathcal{J}) \le \sum_{K_U \in \mathcal{M}} \sum_{f \in \mathcal{K}_U} \sum_{v \in f} |\Gamma_{\{v\}}(\mathcal{G})| \le |\mathcal{M}| \cdot \lceil x \rceil \cdot \left(\max_{f \in \mathcal{G}} |f| \right) \cdot \Delta_1(\mathcal{G}) \le \tilde{r} \cdot \lceil x \rceil k z = r. \quad (42)$$

It follows that $M_{j,x,y}(\mathcal{H}_p) \le \tilde{r}$ implies $\mathcal{E}_{j,1}(x, r, y, z)$, as claimed.

For (41) it remains to estimate $\mathbb{P}(M_{j,x,y}(\mathcal{H}_p) > \tilde{r})$. Similar to the proof of Theorem 11 in [35], we set

$$Z_R = \sum_{\substack{(U_1, \ldots, U_R): \\ U_i \subseteq V(\mathcal{H}) \text{ and } |U_i| = j}} \mathbb{1}_{\{ \square_{i \in [R]} \mathcal{D}_{U_i, x, y} \}}, \quad (43)$$

where $\square$ is defined as in (15). If $M_{j,x,y}(\mathcal{H}_p) > \tilde{r}$, then there is $\mathcal{M} \subseteq \mathcal{S}_{j,x,y}(\mathcal{H}_p)$ of size $|\mathcal{M}| = \lceil \tilde{r} \rceil = R$ which satisfies $V(\mathcal{K}_U) \cap V(\mathcal{K}_W) = \emptyset$ for all distinct $(U, \mathcal{K}_U), (W, \mathcal{K}_W) \in \mathcal{M}$. So, since the disjoint vertex sets $V(\mathcal{K}_U) \subseteq V_p(\mathcal{H})$ guarantee the occurrence of each event $\mathcal{D}_{U,x,y}$, it follows that $\square_{(U, \mathcal{K}_U) \in \mathcal{M}} \mathcal{D}_{U,x,y}$ occurs. As $U \subseteq V(\mathcal{K}_U)$ holds, by vertex disjointness of the $V(\mathcal{K}_U)$ we deduce that the corresponding 'star-centres' U are distinct. Since Z_R counts ordered R-tuples, we thus infer $Z_R \ge R!$. Hence, Markov's inequality yields

$$\mathbb{P}(M_{j,x,y}(\mathcal{H}_p) > \tilde{r}) \le \mathbb{P}(Z_R \ge R!) \le (\mathbb{E} Z_R)/R!. \quad (44)$$

Turning to $\mathbb{E}Z_R$, using the BK-inequality (16) we readily obtain

$$
\begin{aligned}
\mathbb{E}Z_R &= \sum_{\substack{(U_1,\dots,U_R): \\ U_i \subseteq V(\mathcal{H})s \text{ and } |U_i|=j}} \mathbb{P}\Big(\square_{i \in [R]} \mathcal{D}_{U_i,x,y}\Big) \\
&\leq \sum_{\substack{(U_1,\dots,U_R): \\ U_i \subseteq V(\mathcal{H}) \text{ and } |U_i|=j}} \prod_{i \in [R]} \mathbb{P}\big(\mathcal{D}_{U_i,x,y}\big) \leq \Big(\sum_{U \subseteq V(\mathcal{H}): |U|=j} \mathbb{P}\big(\mathcal{D}_{U,x,y}\big)\Big)^R,
\end{aligned}
\tag{45}
$$

which together with (44) and $R \geq 1$ completes the proof. $\quad\square$

The 'star-matching' based deletion argument used in the above proof seems of independent interest. In applications it might be easier to avoid $\mathcal{E}_{j,1}(x,r,y,z)$, and directly work with the random variable $M_{j,x,y}(\mathcal{H}_p)$, see also [35,27]. The above estimates (44)–(45) exploit the BK-inequality to relate $M_{j,x,y}(\mathcal{H}_p)$ with the simpler events $\mathcal{D}_{U,x,y}$. In $\mathcal{H}_p$ and other probability spaces one can sometimes also estimate $\mathbb{P}(M_{j,x,y}(\mathcal{H}_p) \geq z)$ more directly (see, e.g., the remark after the proof of Lemma 17 in [35], or the proof of Lemma 9 in [27]).

3.4. Probabilistic estimates

In this section we introduce our key probabilistic estimates, which complement the combinatorial decomposition of Theorem 15, i.e., allow us to bound the right hand side of (33). A key aspect of inequalities (46)–(47) is that improved degree constraints $\Delta_i(\mathcal{G}) \leq y$ translate into improved tail estimates. In our applications (48) below often reduces to $\mathbb{P}\big(\neg\mathcal{E}_{j,1}(x,r,y,z)\big) \leq (e\mu_j/x)^{-\Theta(r/(yz))}$, say (see, e.g., the proof of Theorem 34).

Theorem 18 (Probabilistic upper tail estimates). *Given $\mathcal{H}$, assume that (P') holds. Set $\varphi(x) = (1+x)\log(1+x) - x$. Then for all $x,r,y,z,t > 0$ and $1 \leq j < k$ we have*

$$
\mathbb{P}\big(w(\mathcal{G}) \geq \mu + t/2 \text{ and } \Delta_1(\mathcal{G}) \leq y \text{ for some } \mathcal{G} \subseteq \mathcal{H}_p\big) \leq \exp\Big(-\frac{\varphi(t/\mu)\mu}{4Lky}\Big),
\tag{46}
$$

$$
\mathbb{P}\big(\Delta_j(\mathcal{G}) \geq x \text{ and } \Delta_{j+1}(\mathcal{G}) \leq y \text{ for some } \mathcal{G} \subseteq \mathcal{H}_p\big) \leq N^j \Big(\frac{e\mu_j}{x}\Big)^{x/(ky)},
\tag{47}
$$

$$
\mathbb{P}\big(\neg\mathcal{E}_{j,1}(x,r,y,z)\big) \leq \Big(N^j \Big(\frac{e\mu_j}{\lceil x \rceil}\Big)^{\lceil x \rceil/(ky)}\Big)^{\lceil r/(k\lceil x \rceil z)\rceil}.
\tag{48}
$$

The proofs of (46)–(47) are based on fairly *routine* applications of Theorem 7. The crux is that the restrictions $\Delta_1(\mathcal{G}) \leq y$ and $\Delta_{j+1}(\mathcal{G}) \leq y$ translate into bounds for the parameter C in (14), which intuitively controls the 'largest dependencies' ($\Delta_1(\mathcal{G}) \leq y$ ensures that every edge $f \in \mathcal{G}$ overlaps at most $|f|\cdot\Delta_1(\mathcal{G}) \leq ky$ edges $e \in \mathcal{G}$). For verifying

the independence assumption of Theorem 7, we use the following simple observation: $e \cap f = \emptyset$ implies that $\mathbb{1}_{\{e \in \mathcal{H}_p\}} = \mathbb{1}_{\{e \subseteq V_p(\mathcal{H})\}}$ and $\mathbb{1}_{\{f \in \mathcal{H}_p\}} = \mathbb{1}_{\{f \subseteq V_p(\mathcal{H})\}}$ are independent, since both depend on *disjoint* sets of independent variables $\xi_\sigma = \mathbb{1}_{\{\sigma \in V_p(\mathcal{H})\}}$. Assuming $(e \cap f) \setminus U = \emptyset$, we below exploit that an analogous (conditional independence) reasoning works after conditioning on $U \subseteq V_p(H)$.

Proof of Theorem 18. With an eye on Theorem 7, inspired by Remark 8 we set $\xi_\sigma = \mathbb{1}_{\{\sigma \in V_p(\mathcal{H})\}}$.

We first prove (46). Let $Y_f = w_f \mathbb{1}_{\{f \in \mathcal{H}_p\}}$, which satisfies $Y_f = w_f \prod_{\sigma \in f} \xi_\sigma$ and $\sum_{f \in \mathcal{H}} \mathbb{E} Y_f = \mathbb{E} w(\mathcal{H}) = \mu$. Furthermore, $w(\mathcal{G}) = \sum_{w \in \mathcal{G}} Y_f$ for any $\mathcal{G} \subseteq \mathcal{H}_p$. Defining $\alpha \sim \beta$ if $\alpha \cap \beta \neq \emptyset$, the independence assumption of Theorem 7 holds by Remark 8. Observe that for any $f \in \mathcal{G} \subseteq \mathcal{H}$ with $\Delta_1(\mathcal{G}) \leq y$ we have

$$\sum_{e \in \mathcal{G}: e \sim f} Y_e \leq \left(\max_{e \in \mathcal{G}} w_e \right) \cdot \sum_{e \in \mathcal{G}: e \cap f \neq \emptyset} \mathbb{1}_{\{e \in \mathcal{H}_p\}} \leq L \cdot \sum_{v \in f} |\Gamma_{\{v\}}(\mathcal{G})| \leq L \cdot |f| \cdot \Delta_1(\mathcal{G}) \leq Lky.$$

To sum up, if $w(\mathcal{G}) \geq \mu + t/2$ and $\Delta_1(\mathcal{G}) \leq y$ for some $\mathcal{G} \subseteq \mathcal{H}_p$, then $Z_C \geq \mu + t/2$ holds with $C = Lky$, where Z_C is defined as in Theorem 7 with $\mathcal{I} = \mathcal{H}$. So, applying (14), we deduce

$$\mathbb{P}\left(w(\mathcal{G}) \geq \mu + t/2 \text{ and } \Delta_1(\mathcal{G}) \leq y \text{ for some} \right.$$

$$\left. \mathcal{G} \subseteq \mathcal{H}_p \right) \leq \mathbb{P}(Z_C \geq \mu + t/2) \leq \exp\left(-\frac{\varphi(t/(2\mu))\mu}{Lky} \right). \tag{49}$$

Using calculus (see, e.g., the proof of Lemma 13 in [35]) it is easy to check that $\varphi(t/(2\mu) \geq \varphi(t/\mu)/4$. In view of (49) and (14), inequality (46) now follows.

Next we turn to (47), which hinges on the union bound estimate (40). Note that $v(\mathcal{H}) < 1$ implies $\mathcal{H} = \emptyset$, so (47) is trivial for $N < 1$ (the left hand side is zero). Similarly, (47) is also trivial for $x \leq e\mu_j$ and $N \geq 1$ (the expression on the right hand side is at least one). To sum up, we henceforth may assume $x > e\mu_j$ and $N \geq 1$. Given $U \subseteq V(\mathcal{H})$ with $|U| = j$, set $\mathcal{I} := \Gamma_U(\mathcal{H}) = \{f \in \mathcal{H} : U \subseteq f\}$. Let $Y_f = \mathbb{1}_{\{f \in \mathcal{H}_p\}}$, and define $\alpha \sim \beta$ if $(\alpha \cap \beta) \setminus U \neq \emptyset$. Note that for any $f \in \mathcal{K} \subseteq \mathcal{I}$ with $\Delta_{|U|+1}(\mathcal{K}) \leq y$ we have

$$\sum_{e \in \mathcal{K}: e \sim f} Y_e = \sum_{e \in \mathcal{K}: (e \cap f) \setminus U \neq \emptyset} \mathbb{1}_{\{e \in \mathcal{H}_p\}} \leq \sum_{v \in f \setminus U} |\Gamma_{U \cup \{v\}}(\mathcal{K})| \leq |f \setminus U| \cdot \Delta_{|U|+1}(\mathcal{K}) \leq ky.$$

$$\tag{50}$$

So, if $\mathcal{D}_{U,x,y}$ occurs, then $Z_C \geq x$ holds with $C = ky$, where Z_C is defined as in Theorem 7 with $\mathcal{I} = \Gamma_U(\mathcal{H})$. For $f \in \mathcal{I}$, note that $U \nsubseteq V_p(\mathcal{H})$ implies $f \notin \mathcal{H}_p = \mathcal{H}[V_p(\mathcal{H})]$. Recalling $Y_f = \mathbb{1}_{\{f \in \mathcal{H}_p\}}$ and $\xi_\sigma = \mathbb{1}_{\{\sigma \in V_p(\mathcal{H})\}}$, using the definition of μ_j (see (21)) it follows that

$$\sum_{f \in \mathcal{I}} \mathbb{E}(Y_f \mid (\xi_\sigma)_{\sigma \in U}) = \sum_{f \in \Gamma_U(\mathcal{H})} \mathbb{P}(f \in \mathcal{H}_p \mid (\xi_\sigma)_{\sigma \in U}) \mathbb{1}_{\{U \subseteq V_p(\mathcal{H})\}}$$

$$\leq \sum_{f \in \Gamma_U(\mathcal{H})} \mathbb{P}(f \in \mathcal{H}_p \mid U \subseteq V_p(\mathcal{H})) = \sum_{f \in \Gamma_U(\mathcal{H})} p^{|f|-|U|} \leq \mu_{|U|} = \mu_j.$$

$$(51)$$

Furthermore, conditional on $(\xi_\sigma)_{\sigma \in U}$, the independence assumption of Theorem 7 holds by the same reasoning as in Remark 8 (in the conditional space, each Y_f is a function of the independent random variables $(\xi_\sigma)_{\sigma \in f \setminus U}$). So, applying (14) with $\mu = \mu_j$ and $\mu + t = x > e\mu_j$, we deduce the conditional inequality

$$\mathbb{P}(\mathcal{D}_{U,x,y} \mid (\xi_\sigma)_{\sigma \in U}) \leq \mathbb{P}(Z_C \geq x \mid (\xi_\sigma)_{\sigma \in U}) \leq \left(\frac{e\mu_j}{x}\right)^{x/(ky)}. \qquad (52)$$

Taking expectations, by summing over all relevant $U \subseteq V(\mathcal{H})$ we thus infer

$$\sum_{U \subseteq V(\mathcal{H}):|U|=j} \mathbb{P}(\mathcal{D}_{U,x,y}) = \sum_{U \subseteq V(\mathcal{H}):|U|=j} \mathbb{E}\mathbb{P}(\mathcal{D}_{U,x,y} \mid (\xi_\sigma)_{\sigma \in U}) \leq N^j \left(\frac{e\mu_j}{x}\right)^{x/(ky)}, \quad (53)$$

and (47) follows in view of (40).

It remains to establish (48). Exploiting integrality of the underlying variables, note in (52) we can strengthen $Z_C \geq x$ to $Z_C \geq \lceil x \rceil$. In (52)–(53) we thus may replace $(e\mu_j/x)^{x/(ky)}$ by $(e\mu_j/\lceil x \rceil)^{\lceil x \rceil/(ky)}$, and so (48) follows from (41) of Lemma 16, with room to spare. $\quad \square$

The proof of Claim 14 (only used in our informal discussion) is very similar, and thus left to the reader.

3.5. Extension: uniform random induced subhypergraph $\mathcal{H}_m$

The proofs in Sections 3.3–3.4 exploited the independence of $\mathcal{H}_p = \mathcal{H}[V_p(\mathcal{H})]$ in a limited way. In this section we record that they extend to the uniform model $\mathcal{H}_m = \mathcal{H}[V_m(\mathcal{H})]$, where the vertex subset $V_m(\mathcal{H}) \subseteq V(\mathcal{H})$ of size $|V_m(\mathcal{H})| = m$ is chosen uniformly at random (this is a natural variant of $\mathcal{H}_p$ with mild dependencies).

Remark 19. Theorems 15 and 18 carry over to $\mathcal{H}_m$ after setting $p = m/v(\mathcal{H})$ in (21).

Proof. The proof of Theorem 15 is based on (deterministic) combinatorial arguments, and after replacing $\mathcal{H}_p$ with $\mathcal{H}_m$ thus carries over word-for-word to $\mathcal{H}_m$.

Turning to Theorem 18, using Remark 9 it is easy to see that the proof of (46) carries over to $\mathcal{H}_m$ (with minor notational changes).

For (47) more care is needed. To avoid conditional probabilities and expectations, set $Y_f = \mathbb{1}_{\{f \setminus U \subseteq V_m(\mathcal{H})\}}$ for all $f \in \mathcal{I} := \Gamma_U(\mathcal{H})$. Writing $\alpha \sim \beta$ if $(\alpha \cap \beta) \setminus U \neq \emptyset$, note

that inequality (50) readily carries over. It is folklore (analogous to, e.g., the proof of Theorem 15 in [17]) that $\mathbb{E}Y_f = \mathbb{P}(f \setminus U \subseteq V_m(\mathcal{H})) \leq p^{|f|-|f \cap U|}$ for $p = m/v(\mathcal{H})$, so that $\sum_{f \in \mathcal{I}} \mathbb{E}Y_f \leq \sum_{f \in \Gamma_U(\mathcal{H})} p^{|f|-|U|} \leq \mu_j$ by (21). Recalling the definition of $\sim$, it is similarly folklore that the random variables $Y_f = \mathbb{1}_{\{f \setminus U \subseteq V_m(\mathcal{H})\}}$ satisfy the negative correlation condition of Remark 9. Mimicking the argument leading to (52), using Theorem 7 we obtain $\mathbb{P}(\mathcal{D}_{U,x,y}) \leq \mathbb{P}(Z_C \geq x) \leq (e\mu_j/x)^{x/(ky)}$ for $\mathcal{H}_m$, which by a simpler variant of (53) then establishes (47).

As the proof of (47) carries over, for (48) it remains to check that (41) holds for $\mathcal{H}_m$. A close inspection of the proof of Lemma 16 reveals that only the usage of the BK-inequality in (45) needs to be justified. But, since $\mathcal{D}_{U,x,y}$ is an increasing event, this application of (16) is valid by Remark 13, completing the proof. $\quad\square$

4. More general setup

In this section we introduce our general Kim–Vu/Janson–Ruciński type setup, and show that the combinatorial and probabilistic arguments of Section 3 carry over with somewhat minor changes. Readers only interested in random induced subhypergraphs $\mathcal{H}_p$ may wish to skip to Section 5 (see Remark 29).

4.1. Setup

Our general setup is based on certain independence assumptions, i.e., we do not restrict ourselves to polynomials of independent random variables (and we also do not make any monotonicity assumptions). Given a hypergraph $\mathcal{H}$ and non-negative random variables $(Y_f)_{f \in \mathcal{H}}$, for every $\mathcal{G} \subseteq \mathcal{H}$ we set

$$X(\mathcal{G}) = \sum_{f \in \mathcal{G}} Y_f, \tag{54}$$

where our main focus[5] is on the sum $X(\mathcal{H})$ of all the variables Y_f (sometimes $\mathcal{H}$ is also called the 'supporting' or 'underlying' hypergraph, see [19,32]). Loosely speaking, the plan is to adapt the combinatorial arguments of Sections 3.3–3.4 to the associated random subhypergraph

$$\mathcal{H}_p = \{f \in \mathcal{H} : Y_f > 0\}, \tag{55}$$

which due to $X(\mathcal{H}) = X(\mathcal{H}_p)$ loosely encodes all 'relevant' variables (recall that $Y_f \geq 0$). Similar to [15], we shall use the following *independence assumption* (Hℓ), where $\ell \in \mathbb{N}$ is a parameter:

---

[5] Usually we have $X = \sum_{f \in \mathcal{H}} w_f I_f$ in mind, for random variables $I_f \in \{0,1\}$ and constants $w_f \in (0,\infty)$. All examples and applications in [19,30,32,15,14,16] are of this form, with $w_f = 1$ (possibly after rescaling X by a constant factor).

(Hℓ) Let $(\xi_\sigma)_{\sigma \in \mathcal{A}}$ be a family of independent finite random variables. Suppose that there are families of subsets $\mathcal{A}_U \subseteq \mathcal{A}$ such that (i) each non-negative random variable Y_f with $f \in \mathcal{H}$ is a function of the variables $(\xi_\sigma)_{\sigma \in \mathcal{A}_f}$, (ii) we have $\mathcal{A}_e \cap \mathcal{A}_f \subseteq \mathcal{A}_{e \cap f}$ for all $e, f \in \mathcal{H}$, and (iii) we have $\mathcal{A}_e \cap \mathcal{A}_f = \emptyset$ for all $e, f \in \mathcal{H}$ with $|e \cap f| < \ell$.

The setup of Section 3.1.1 corresponds to the special case $\xi_\sigma = \mathbb{1}_{\{\sigma \in V_p(\mathcal{H})\}}$, $\mathcal{A}_f = f$ and $Y_f = w_f \prod_{\sigma \in \mathcal{A}_f} \xi_\sigma$. A key consequence of (Hℓ) is that Y_e and Y_f are independent whenever $|e \cap f| < \ell$, since by (i) and (iii) then both depend on *disjoint* sets of variables ξ_σ. The 'structural' assumption (i) that each Y_f depends only on the variables ξ_σ with $\sigma \in \mathcal{A}_f$ is very common in applications; often $\mathcal{A}_U = U$ suffices. The 'consistency' assumption (ii) and 'independence' assumption (iii) of the index sets $\mathcal{A}_U$ are also very natural. For example, in the frequent case $\mathcal{A}_U = U$ we have $\mathcal{A}_e \cap \mathcal{A}_f = \mathcal{A}_{e \cap f}$, so $\mathcal{A}_e \cap \mathcal{A}_f = \emptyset$ if $|e \cap f| < 1$. Example 22 in Section 4.1.1 illustrates the case $\ell \neq 1$ with $\mathcal{A}_U = \{f \in E(K_n) : f \subseteq U\}$.

We now introduce the modified key parameters μ_j, which intuitively quantify the 'dependencies' among the variables Y_f (in the spirit of [15,19,30,32]). Recalling $\Gamma_U(\mathcal{H}) = \{f \in \mathcal{H} : U \subseteq f\}$, with Section 3.1.1 in mind we now define the following two crucial *assumptions* (P) and (Pq), where $q \in \mathbb{N}$ is a parameter:

(P) Assume that $\max_{f \in \mathcal{H}} |f| \leq k$, $\max_{f \in \mathcal{H}} \sup Y_f \leq L$ and $v(\mathcal{H}) \leq N$. Define $\mu = \mathbb{E} X(\mathcal{H})$ and

$$\mu_j = \max_{U \subseteq V(\mathcal{H}): |U| = j} \sup \mathbb{E}\big(|\Gamma_U(\mathcal{H}_p)| \,\big|\, (\xi_\sigma)_{\sigma \in \mathcal{A}_U}\big), \tag{56}$$

where the supremum is over all values of the variables ξ_σ with $\sigma \in \mathcal{A}_U$.

(Pq) Assume that $\Delta_q(\mathcal{H}) \leq D$.

In view of (22), property (P) is a natural extension of (P') from the basic setup of Section 3.1.1. Our general setup lacks monotonicity, and so the conditioning in (56) is with respect to all possible values of the ξ_σ.

For the interested reader, we now briefly discuss how our setup and assumptions differ in some (usually irrelevant) minor details from the literature [15,19,30,32]. Firstly, the 'normal' assumption of Vu implies $\max_{f \in \mathcal{H}} \sup Y_f \leq 1$ in (P) above (see, e.g., Theorem 1.2 in [30] and Theorem 4.2 in [32]). Secondly, classical variants of the 'maximum average effect' parameter μ_j (see, e.g., Sections 3 in [15] and Section 4 in [32]) are roughly defined as the maximum over all $\sup \mathbb{E}(\sum_{f \in \Gamma_U(\mathcal{H}_p)} Y_f \mid (\xi_\sigma)_{\sigma \in \mathcal{A}_U})$ with $|U| = j$, but in most applications $\sum_{f \in \Gamma_U(\mathcal{H}_p)} Y_f = \Theta(|\Gamma_U(\mathcal{H}_p)|)$ holds, so the difference is usually immaterial. Thirdly, in (Hℓ) our assumptions for the index sets $\mathcal{A}_U$ are slightly simpler than in Section 3 of [15]. Finally, in contrast to [15], we assume that the $(\xi_\sigma)_{\sigma \in \mathcal{A}}$ are *finite* random variables, which is very natural in combinatorial applications (this technicality can presumably be removed by approximation arguments, but we have not pursued this).

4.1.1. Examples

The above assumptions (Hℓ) and (P) might seem a bit technical at first sight, and for this reason we shall below spell out three pivotal examples (see Section 3 of [15] for more examples).

Example 20 *(Random induced subhypergraphs).* For a given k-uniform hypergraph $\mathcal{H}$, analogous to Section 3.1.1 we consider $X = e(\mathcal{H}_p) = \sum_{f \in \mathcal{H}} \mathbb{1}_{\{f \in \mathcal{H}_p\}}$. Note that $\mathcal{A} = \mathcal{H}$, $\xi_\sigma = \mathbb{1}_{\{\sigma \in V_p(\mathcal{H})\}}$, $\mathcal{A}_f = f$ and $Y_f = \prod_{\sigma \in \mathcal{A}_f} \xi_\sigma \in \{0, 1\}$ satisfy properties (H1) and (Pk). In fact, for (P) we can simplify the definition of μ_j. Namely, since $U \not\subseteq V_p(\mathcal{H})$ implies $f \notin \mathcal{H}_p = \mathcal{H}[V_p(\mathcal{H})]$ for all $f \in \Gamma_U(\mathcal{H})$, we have

$$\sup \mathbb{E}\big(|\Gamma_U(\mathcal{H}_p)| \,\big|\, (\xi_\sigma)_{\sigma \in \mathcal{A}_U}\big) = \mathbb{E}\big(|\Gamma_U(\mathcal{H}_p)| \,\big|\, U \subseteq V_p(\mathcal{H})\big)$$
$$= \sum_{f \in \Gamma_U(\mathcal{H})} \mathbb{P}\big(f \in \mathcal{H}_p \,\big|\, U \subseteq V_p(\mathcal{H})\big).$$

As $\mathcal{H}$ is k-uniform, for any $f \in \Gamma_U(\mathcal{H})$ it is easy to see that $\mathbb{P}\big(f \in \mathcal{H}_p \,\big|\, U \subseteq V_p(\mathcal{H})\big) = \mathbb{P}\big(f \setminus U \subseteq V_p(\mathcal{H})\big) = p^{k-|U|}$. Combining these observations, it follows that (56) simplifies for $1 \le j \le k$ to

$$\mu_j = \max_{U \subseteq V(\mathcal{H}): |U| = j} |\Gamma_U(\mathcal{H})| \cdot p^{k-j}. \tag{57}$$

Example 21 *(Subgraph counts in $G_{n,p}$: induced subhypergraphs approach).* Subgraph counts in $G_{n,p}$ can be viewed as a special case of Example 20, i.e., random induced subhypergraphs. Given a fixed subgraph H with $e = e_H$ edges, $v = v_H$ vertices and minimum degree $\delta = \delta_H \ge 1$, we consider the e-uniform hypergraph $\mathcal{H}$ with vertex set $V(\mathcal{H}) = E(K_n)$, where edges correspond to copies of H. Clearly, $k = e$ and $N = n^2$ suffice. Note that for the copy of H counted by Y_f, any subset of the edges $U \subseteq f \cap E(K_n) \subseteq V(\mathcal{H})$ is isomorphic to some subgraph $J \subseteq H$. So, taking all subgraphs of H with exactly $|U| = j$ edges into account, using (57) with $k = e$ and $V(\mathcal{H}) = E(K_n)$ there is universal constant $B = B(H) > 0$ such that for $1 \le j \le e$ we have

$$\mu_j \le \sum_{J \subseteq H: e_J = j} \max_{U \subseteq E(K_n): U \cong J} |\Gamma_U(\mathcal{H})| \cdot p^{e-j} \le B \sum_{J \subseteq H: e_J = j} n^{v - v_J} p^{e-j}. \tag{58}$$

Note that any $q = e - \delta + 1 \le e$ edges already determine the vertex set, so (Pq) holds with $D = O(1)$. Finally, a minor variant of the described approach also applies to *induced* subgraph counts (with $k = \binom{v_H}{2}$), by letting $E(\mathcal{H})$ correspond to copies of the complete graph K_{v_H}, and defining Y_f as the indicator for the event that the subgraph of $G_{n,p}$ defined by the edges in f is isomorphic to H).

Example 22 *(Subgraph counts in $G_{n,p}$: vertex exposure approach).* Subgraph counts in $G_{n,p}$ can also be treated via a 'vertex exposure' based approach. Given a fixed sub-

graph H with $e = e_H$ edges and $v = v_H$ edges, we consider the complete v-uniform hypergraph $\mathcal{H}$ with vertex set $V(\mathcal{H}) = [n]$, so $N = n$ and $k = v$. For $I \subseteq V(\mathcal{H})$ with $|I| = v$ the random variable Y_I counts the number of copies of H in $G_{n,p}$ that have vertex set I. Note that $0 \leq Y_I \leq L = O(1)$. Since $X = \sum_{I \in \mathcal{H}} Y_I$, we take $\mathcal{A} = E(K_n)$, $\xi_\sigma = \mathbb{1}_{\{\sigma \in V_p(\mathcal{H})\}}$, and $\mathcal{A}_I = \{f \in E(K_n) : f \subseteq I\}$. As $\mathcal{A}_I \cap \mathcal{A}_J = \mathcal{A}_{I \cap J}$ is empty whenever $|I \cap J| < 2$, for $\ell = 2$ properties (Hℓ) and (Pk) are satisfied. Conditioning on $(\xi_\sigma)_{\sigma \in \mathcal{A}_U}$ corresponds to conditioning on $G_{n,p}[U]$, so bounding μ_j is conceptually analogous (58). Indeed, by similar reasoning as in Example 21, we arrive for $1 \leq j \leq v$ at

$$\mu_j \leq B \sum_{\text{induced } J \subseteq H : v_J = j} n^{v-j} p^{e-e_J}, \tag{59}$$

where $B = B(H) > 0$. Finally, *induced* subgraph counts can clearly be treated analogously.

4.2. Adapting the arguments of Sections 3.3–3.4

In this section we adapt the key results Theorem 15 and 18 from Sections 3.3–3.4 to our more general setup. The crux is that the random variables $(Y_f)_{f \in \mathcal{H}}$ satisfy $Y_f = Y_f(\xi_\sigma : \sigma \in \mathcal{A}_f)$ by the independence assumption (Hℓ), so that the intersection properties of the index sets $\mathcal{A}_f$ give us a handle on the dependencies. This allows us to adapt our combinatorial arguments to the auxiliary subhypergraph $\mathcal{H}_p = \{f \in \mathcal{H} : Y_f > 0\}$.

We start with a natural analogue of Theorem 15, which is at the heart of our arguments.

Theorem 23 (*Combinatorial decomposition of the upper tail: general setup*). *Given $\mathcal{H}$ with $1 \leq \ell \leq q \leq k$, assume that (H$\ell$) and (P) hold. Suppose that $t > 0$. Given positive $(R_j)_{\ell \leq j < q}$ and $(D_j)_{\ell \leq j \leq q}$, define $R_q = Q_q = D_q$ and $Q_j = \max\{S_j, D_j\}$ for $\ell \leq j < q$. Then we have*

$$\begin{aligned}
\mathbb{P}\big(X(\mathcal{G}) &\geq \mu + t \text{ and } \Delta_q(\mathcal{G}) \leq D_q \text{ for some } \mathcal{G} \subseteq \mathcal{H}_p\big) \\
&\leq \mathbb{P}\big(X(\mathcal{G}) \geq \mu + t/2 \text{ and } \Delta_\ell(\mathcal{G}) \leq \min\{Q_\ell, R_\ell\} \text{ for some } \mathcal{G} \subseteq \mathcal{H}_p\big) \\
&\quad + \sum_{\ell \leq j < q} \big[\mathbb{P}_{j,1} + \mathbb{P}_{j,2} + \mathbb{P}_{j,3,\ell}\big],
\end{aligned} \tag{60}$$

where $\mathbb{P}_{j,1}$, $\mathbb{P}_{j,2}$ and $\mathbb{P}_{j,3,\ell}$ are defined as in (34)–(36).

Recalling $X(\mathcal{G}) = \sum_{f \in \mathcal{G}} Y_f$ and $\mathcal{H}_p = \{f \in \mathcal{H} : Y_f > 0\}$, the deterministic proof of Theorem 15 carries over to Theorem 23 with minor obvious changes (inequality (60) is trivial if $q = \ell$; for $q > \ell$ it suffices to construct $\mathcal{G} = \mathcal{J}_q \supseteq \cdots \supseteq \mathcal{J}_\ell$, with indices of form $\ell \leq i, j \leq q$ in (37)); we omit the routine details.

Next we state an analogue of Lemma 16 for the 'sparsification' event $\mathcal{E}_{j,\ell}(x, r, y, z)$ from Section 3.3.

Lemma 24 *(Auxiliary result for the sparsification event: general setup). Given $\mathcal{H}$ with $1 \le \ell \le k$, assume that (Hℓ) and $\max_{f \in \mathcal{H}} |f| \le k$ hold. Then for all $x, r, y, z > 0$ and $\ell \le j < k$ we have*

$$\mathbb{P}\left(\neg \mathcal{E}_{j,\ell}(x, r, y, z)\right) \le \left(\sum_{U \subseteq V(\mathcal{H}):|U|=j} \mathbb{P}(\mathcal{D}_{U,x,y}) \right)^{\left\lceil r/\left(\binom{k}{\ell} \lceil x \rceil z\right) \right\rceil}. \tag{61}$$

Remark 25. Inequality (61) remains valid after dividing the right hand side by $\left\lceil r/\left(\binom{k}{\ell} \lceil x \rceil z\right)\right\rceil!$.

For the proof of Lemma 24 we adapt the definition of $M_{j,x,y}(\mathcal{G})$ used for Lemma 16. Intuitively, the idea is to replace 'vertex disjoint' by 'depending on disjoint sets of variables'. Namely, here we define $M_{j,x,y}(\mathcal{G})$ as the size of the largest collection $\mathcal{M} \subseteq \mathcal{S}_{j,x,y}(\mathcal{G})$ of (j, x, y)-stars in $\mathcal{G}$ satisfying the following property for all distinct $(U, \mathcal{K}_U), (W, \mathcal{K}_W) \in \mathcal{M}$: we have $|e \cap f| < \ell$ for all $e \in \mathcal{K}_U$ and $f \in \mathcal{K}_W$. The point will be (i) that each Y_f is a function of the variables $(\xi_\sigma)_{\sigma \in \mathcal{A}_f}$, and (ii) that $|e \cap f| < \ell$ implies $\mathcal{A}_e \cap \mathcal{A}_f = \emptyset$ by (Hℓ).

Proof of Lemma 24. Using the above definition of $M_{j,x,y}(\mathcal{G})$, we shall adapt the proof of Lemma 16. Let $\tilde{r} = r/\left(\binom{k}{\ell} \lceil x \rceil z\right)$ and $R = \lceil \tilde{r} \rceil$. We first assume that $M_{j,x,y}(\mathcal{H}_p) \le \tilde{r}$ holds, and claim that this implies the occurrence of $\mathcal{E}_{j,\ell}(x, r, y, z)$. Fix $\mathcal{G} \subseteq \mathcal{H}_p$ with $\Delta_{j+1}(\mathcal{G}) \le y$ and $\Delta_\ell(\mathcal{G}) \le z$, and let $\mathcal{M} \subseteq \mathcal{S}_{j,x,y}(\mathcal{G})$ attain the maximum in the definition of $M_{j,x,y}(\mathcal{G})$. We remove all edges $f \in \mathcal{G}$ which 'overlap' some star $(U, \mathcal{K}_U) \in \mathcal{M}$, where overlap means that $|f \cap g| \ge \ell$ for some edge $g \in \mathcal{K}_U$. We denote the resulting subhypergraph by $\mathcal{J} \subseteq \mathcal{G}$. Recalling $\Delta_{j+1}(\mathcal{J}) \le \Delta_{j+1}(\mathcal{G}) \le y$, by maximality of $\mathcal{M}$ we infer $\Delta_j(\mathcal{J}) \le \lceil x \rceil - 1 < x$. Similar to (42), using $|\mathcal{M}| = M_{j,x,y}(\mathcal{G}) \le M_{j,x,y}(\mathcal{H}_p) \le \tilde{r}$ and $\Delta_\ell(\mathcal{G}) \le z$ it is easy to see that we removed at most

$$e(\mathcal{G} \setminus \mathcal{J}) \le |\mathcal{M}| \cdot \lceil x \rceil \cdot \left[\max_{f \in \mathcal{G}} \binom{|f|}{\ell} \right] \cdot \Delta_\ell(\mathcal{G}) \le \tilde{r} \cdot \lceil x \rceil \binom{k}{\ell} z = r \tag{62}$$

edges. It follows that $M_{j,x,y}(\mathcal{H}_p) \le \tilde{r}$ implies $\mathcal{E}_{j,\ell}(x, r, y, z)$, as claimed.

For (61) it remains to estimate $\mathbb{P}(M_{j,x,y}(\mathcal{H}_p) > \tilde{r})$. Suppose that $M_{j,x,y}(\mathcal{H}_p) > \tilde{r}$ occurs. If $\mathcal{M} \subseteq \mathcal{S}_{j,x,y}(\mathcal{H}_p)$ attains the maximum in the definition of $M_{j,x,y}(\mathcal{H}_p)$, then we know (i) that $|\mathcal{M}| \ge \lceil \tilde{r} \rceil = R$ holds, and (ii) that $\bigcap_{(U,\mathcal{K}_U) \in \mathcal{M}} \mathcal{D}_{U,x,y}$ occurs. In the following we argue that these events $\mathcal{D}_{U,x,y}$ 'occur disjointly' in the sense of Section 2.2. For each $(U, K_U) \in \mathcal{M}$, note that the variables indexed by

$$V(\mathcal{K}_U) = \bigcup_{f \in \mathcal{K}_U} \mathcal{A}_f$$

guarantee the occurrence of $\mathcal{D}_{U,x,y}$. The crux is now that for all distinct $(U, \mathcal{K}_U)$, $(W, \mathcal{K}_W) \in \mathcal{M}$, by (iii) of (H$\ell$) we have $\mathcal{A}_e \cap \mathcal{A}_f = \emptyset$ for all $e \in \mathcal{K}_u$ and $f \in \mathcal{K}_W$ (since $|e \cap f| < \ell$), so

$$V(\mathcal{K}_U) \cap V(\mathcal{K}_W) = \bigcup_{e \in \mathcal{K}_U} \bigcup_{f \in \mathcal{K}_W} (\mathcal{A}_e \cap \mathcal{A}_f) = \emptyset. \tag{63}$$

It follows that $\square_{(U,\mathcal{K}_U) \in \mathcal{M}} \mathcal{D}_{U,x,y}$ occurs (since the disjoint sets of variables indexed by $V(\mathcal{K}_U)$ guarantee the occurrence of each $\mathcal{D}_{U,x,y}$). Next we claim that all the corresponding sets U are distinct. To see this, note that for distinct $(U,\mathcal{K}_U),(W,\mathcal{K}_W) \in \mathcal{M}$ we have $\ell > |e \cap f| \geq |U \cap W|$ by definition of $\mathcal{M}$, which due to $|U| = |W| = j \geq \ell$ implies $U \neq W$. To sum up, $M_{j,x,y}(\mathcal{H}_p) > \tilde{r}$ implies $Z_R \geq R!$, where Z_R is defined as in (43). The arguments of (44) and (45) now carry over unchanged, completing the proof of (61). $\quad\square$

Finally, we state a natural analogue of Theorem 18, which contains our core probabilistic estimates (inequalities (64)–(66) allow us to bound the right hand side of (60) from Theorem 23).

Theorem 26 (*Probabilistic upper tail estimates: general setup*). *Given $\mathcal{H}$ with $1 \leq \ell \leq k$, assume that (Hℓ) and (P) hold. Set $\varphi(x) = (1+x)\log(1+x) - x$. Then for all $x, r, y, z, t > 0$ and $\ell \leq j < k$ we have*

$$\mathbb{P}\big(X(\mathcal{G}) \geq \mu + t/2 \text{ and } \Delta_\ell(\mathcal{G}) \leq y \text{ for some } \mathcal{G} \subseteq \mathcal{H}_p\big) \leq \exp\left(-\frac{\varphi(t/\mu)\mu}{4L\binom{k}{\ell}y}\right), \tag{64}$$

$$\mathbb{P}\big(\Delta_j(\mathcal{G}) \geq x \text{ and } \Delta_{j+1}(\mathcal{G}) \leq y \text{ for some } \mathcal{G} \subseteq \mathcal{H}_p\big) \leq N^j \left(\frac{e\mu_j}{x}\right)^{x/(ky)}, \tag{65}$$

$$\mathbb{P}\big(\neg\mathcal{E}_{j,\ell}(x,r,y,z)\big) \leq \left(N^j \left(\frac{e\mu_j}{\lceil x \rceil}\right)^{\lceil x \rceil/(4ky)}\right)^{\left\lceil r/\left(\binom{k}{\ell}\lceil x \rceil z\right)\right\rceil}. \tag{66}$$

The proof is based on a minor modification of the proof of Theorem 18. As we shall see, our main task is to adapt the definitions of the dependency relations $\sim$. To this end recall (i) that each Y_f is a function of the independent variables $(\xi_\sigma)_{\sigma \in \mathcal{A}_f}$, and (ii) that (H$\ell$) implies $\mathcal{A}_e \cap \mathcal{A}_f = \emptyset$ whenever $|e \cap f| < \ell$.

Proof of Theorem 26. For (64), note that $\sum_{f \in \mathcal{H}} \mathbb{E} Y_f = \mathbb{E} X(\mathcal{H}) = \mu$. We define $\alpha \sim \beta$ if $|\alpha \cap \beta| \geq \ell$. In view of properties (i) and (ii) discussed above, the independence assumption of Theorem 7 holds by analogous reasoning as in Remark 8. Furthermore, for any $f \in \mathcal{G} \subseteq \mathcal{H}$ with $\Delta_\ell(\mathcal{G}) \leq y$ we have

$$\sum_{e \in \mathcal{G}: e \sim f} Y_e \leq \big(\max_{e \in \mathcal{G}} \sup Y_e\big) \cdot \sum_{e \in \mathcal{G}: |e \cap f| \geq \ell} \mathbb{1}_{\{f \in \mathcal{G}\}} \leq L \cdot \sum_{U \subseteq f: |U| = \ell} |\Gamma_U(\mathcal{G})|$$

$$\leq L \cdot \binom{|f|}{\ell} \cdot \Delta_\ell(\mathcal{G}) \leq L\binom{k}{\ell}y.$$

Setting $C = L\binom{k}{\ell}y$, the remaining proof of (46) readily carries over to (64) with obvious notational changes.

Next we turn to (65), which is again based on (40). As before, we may assume that $x > e\mu_j$ and $N \geq 1$ (otherwise the claim is trivial). Furthermore, given $U \subseteq V(\mathcal{H})$ with $|U| = j$, we set $\mathcal{I} = \Gamma_U(\mathcal{H})$. With the random variables $\left(\mathbb{1}_{\{Y_f>0\}}\right)_{f \in \mathcal{I}}$ in mind, define $\alpha \sim \beta$ if $(\alpha \cap \beta) \setminus U \neq \emptyset$. Note that, for any $f \in \mathcal{K} \subseteq \mathcal{I}$ with $\Delta_{|U|+1}(\mathcal{K}) \leq y$, analogous to (50) we have $\sum_{e \in \mathcal{K}:e \sim f} \mathbb{1}_{\{Y_f>0\}} \leq |f \setminus U| \cdot \Delta_{|U|+1}(\mathcal{K}) \leq ky$. Furthermore, by definition of $\mathcal{I} = \Gamma_U(\mathcal{H})$, $\mathcal{H}_p = \{f \in \mathcal{H} : Y_f > 0\}$ and μ_j (see (56)) we obtain

$$\sum_{f \in \mathcal{I}} \mathbb{E}\left(\mathbb{1}_{\{Y_f>0\}} \mid (\xi_\sigma)_{\sigma \in \mathcal{A}_U}\right) = \mathbb{E}\left(|\Gamma_U(\mathcal{H}_p)| \mid (\xi_\sigma)_{\sigma \in \mathcal{A}_U}\right) \leq \mu_{|U|} = \mu_j.$$

Note that, conditional on $(\xi_\sigma)_{\sigma \in \mathcal{A}_U}$, each $\mathbb{1}_{\{Y_f>0\}}$ is now a function of the independent random variables $(\xi_\sigma)_{\sigma \in \mathcal{A}_f \setminus \mathcal{A}_U}$. Furthermore, for all $e, f \in \mathcal{I} = \{g \in \mathcal{H} : U \subseteq g\}$ we see that $(e \cap f) \setminus U = \emptyset$ implies $e \cap f = U$, so that (ii) of (Hℓ) yields $\mathcal{A}_e \cap \mathcal{A}_f \subseteq \mathcal{A}_{e \cap f} = \mathcal{A}_U$. For all $e, f \in \mathcal{I}$ we thus infer that $e \nsim f$ implies

$$(\mathcal{A}_e \setminus \mathcal{A}_U) \cap (\mathcal{A}_f \setminus \mathcal{A}_U) = (\mathcal{A}_e \cap \mathcal{A}_f) \setminus \mathcal{A}_U \subseteq \mathcal{A}_U \setminus \mathcal{A}_U = \emptyset.$$

Conditional on $(\xi_\sigma)_{\sigma \in \mathcal{A}_U}$, it follows (by the reasoning of Remark 8) that the independence assumption of Theorem 7 holds for the variables $\left(\mathbb{1}_{\{Y_f>0\}}\right)_{f \in \mathcal{I}}$. The remaining proof of (47) readily carries over to (65).

Finally, for (66) we recall that (48) is based on Lemma 16 and the argument leading to (47). In view of Lemma 24 and the above proof of (65), the same line of reasoning carries over, establishing (66). $\quad\square$

4.3. Adapting Section 3.5: vertex exposure approach for $\mathcal{H}_m$

In this section we partially adapt our arguments to the uniform random induced sub-hypergraph $\mathcal{F}_m = \mathcal{F}[V_m(\mathcal{F})]$. Generalizing the 'vertex exposure' approach of Example 22, we rely on the following assumption.

(HℓP) Suppose that $\mathcal{H}$, $\mathcal{E}$ and $\mathcal{F}$ are hypergraphs with $V(\mathcal{H}) = V(\mathcal{E})$, $V(\mathcal{F}) = \{h \in \mathcal{E}\}$ and $\min_{h \in \mathcal{E}} |h| \geq \ell$. Defining $\mathcal{A}_U = \{h \in \mathcal{E} : h \subseteq U\}$ for all $U \subseteq V(\mathcal{E})$, assume that $\mathcal{F} = \bigcup_{f \in \mathcal{H}} \mathcal{F}[\mathcal{A}_f]$ is a disjoint union of induced subhypergraphs. Suppose that $(w_g)_{g \in \mathcal{F}}$ are non-negative weights. For all $f \in \mathcal{H}$, let

$$Y_f = \sum_{g \in \mathcal{F}[\mathcal{A}_f]} w_g \mathbb{1}_{\{g \in \mathcal{F}_m\}}. \tag{67}$$

Assume that $\max_{f \in \mathcal{H}} |f| \leq k$, $\max_{f \in \mathcal{H}} Y_f \leq L$ and $v(\mathcal{H}) \leq N$. Define $\mu = \mathbb{E}X(\mathcal{H})$, $p = m/v(\mathcal{F})$, and

$$\mu_j = \max_{U \subseteq V(\mathcal{E}):|U|=j} \sum_{f \in \Gamma_U(\mathcal{H})} \sum_{g \in \mathcal{F}[\mathcal{A}_f]} p^{|g|-|g \cap \mathcal{A}_U|}. \tag{68}$$

Example 27. Using the 'vertex exposure' setup discussed in Example 22, subgraph counts in $G_{n,m}$ satisfy (HℓP) with $\ell = 2$ and $k = v_H$ (by setting $\mathcal{E} = K_n$, and defining $\mathcal{F}$ as the hypergraph $\mathcal{H}$ of Example 21). In (68) the modified parameter μ_j is again bounded from above by the right hand side of (59).

Remark 28. Theorems 23 and 26 remain valid after replacing the assumptions (Hℓ), (P) with (HℓP).

Proof. With the ideas of Remark 19 in mind, we only sketch the key modifications for (64)–(65) of Theorem 26.

For (64) it suffices to verify the negative correlation condition of Remark 9, writing $\alpha \sim \beta$ if $|\alpha \cap \beta| \geq \ell$. Using (67) and the negative correlation properties of $\mathcal{F}_m$ (see Remark 9), it is not hard to check that

$$\mathbb{E}\left(\prod_{i \in [s]} Y_{\alpha_i}\right) = \sum_{g_1 \in \mathcal{F}[\mathcal{A}_{\alpha_1}]} \cdots \sum_{g_s \in \mathcal{F}[\mathcal{A}_{\alpha_s}]} \mathbb{E}\left(\prod_{i \in [s]} w_{g_i} \mathbb{1}_{\{g_i \in \mathcal{F}_m\}}\right) \leq \prod_{i \in [s]} \mathbb{E} Y_{\alpha_i}, \qquad (69)$$

and so the proof of (64) carries over (above we used that $\alpha_i \nsim \alpha_j$ implies $\mathcal{F}[\mathcal{A}_{\alpha_i}] \cap \mathcal{F}[\mathcal{A}_{\alpha_j}] = \emptyset$).

For (65) we define $\alpha \sim \beta$ if $(\alpha \cap \beta) \setminus U \neq \emptyset$, and replace $\left(\mathbb{1}_{\{Y_f > 0\}}\right)_{f \in \mathcal{I}}$ by $\left(\mathbb{1}_{\{\mathfrak{Y}_f\}}\right)_{f \in \mathcal{I}}$, where $\mathfrak{Y}_f$ denotes the event that $g \setminus \mathcal{A}_U \subseteq V_m(\mathcal{F})$ for some $g \in \mathcal{F}[\mathcal{A}_f]$. Let $\lambda_f = \sum_{g \in \mathcal{F}[\mathcal{A}_f]} \mathbb{P}(g \setminus \mathcal{A}_U \subseteq V_m(\mathcal{F}))$. It is folklore that $\mathbb{P}(g \setminus \mathcal{A}_U \subseteq V_m(\mathcal{F})) \leq p^{|g| - |g \cap \mathcal{A}_U|}$ (see Remark 19), so $\mathcal{I} = \Gamma_U(\mathcal{H})$ and (68) yield $\sum_{f \in \mathcal{I}} \lambda_f \leq \mu_{|U|} = \mu_j$. Since $\mathbb{1}_{\{\mathfrak{Y}_f\}} \leq \sum_{g \in \mathcal{F}[\mathcal{A}_f]} \mathbb{1}_{\{g \setminus \mathcal{A}_U \subseteq V_m(\mathcal{F})\}}$, analogous to (69) we infer $\mathbb{E}\left(\prod_{i \in [s]} \mathbb{1}_{\{\mathfrak{Y}_{\alpha_i}\}}\right) \leq \prod_{i \in [s]} \lambda_{\alpha_i}$, establishing the correlation condition of Remark 10. Mimicking Remark 19, the proof of (47) then carries over to (65). $\quad\square$

5. Corollaries: upper tail inequalities

The main results of Sections 3–4 are Theorems 15, 23 of form $\mathbb{P}(X \geq (1 + \varepsilon)\mathbb{E}X) \leq \sum_i \mathbb{P}(\neg \mathcal{E}_i)$ and Theorems 18, 26 of form $\mathbb{P}(\neg \mathcal{E}_i) \leq \exp(-\Psi_i)$. In this section we derive upper tail inequalities that are convenient for the applications of Section 6, and briefly compare some of our more general estimates with the literature.

Remark 29 *(Random induced subhypergraph setup).* The results in Sections 5.1–5.2 are stated for the general setup of Section 4.1. But, with minor changes, they remain valid in the simpler random induced subhypergraph setup of Section 3.1.1. Indeed, setting $\ell = 1$ and replacing the assumptions (Hℓ), (P) with (P'), all results carry over to $\mathcal{H}_p$ by defining $X(\mathcal{J}) := w(\mathcal{J})$. After setting $p = m/v(\mathcal{H})$ in (21), these results for $\mathcal{H}_p$ then also carry over to the uniform variant $\mathcal{H}_m$ defined in Section 3.5. Finally, after replacing the assumptions (Hℓ), (P) with (HℓP), all results in Sections 5.1–5.2 also remain valid in the setup of Section 4.3.

Henceforth, we tacitly set $\varphi(x) = (1+x)\log(1+x) - x$ for brevity (as in Theorems 7, 18 and 26).

5.1. Easy-to-apply tail inequalities

In this section we state some simplified upper tail inequalities that suffice for all the applications in Section 6 (we have not optimized the usually irrelevant constants); the proofs are deferred to Section 5.3.

On first reading of the following upper tail inequality for $X(\mathcal{H}) = \sum_{f \in \mathcal{H}} Y_f$, the reader may wish to set $\ell = 1$ and $q = k$, so that (72) is of form $\mathbb{P}(X(\mathcal{H}) \geq 2\mu) \leq \exp(-d\min\{\mu, \mu^{1/k}\log(e/\pi)\})$. Here our main novelty is the $\log(e/\pi)$ term: it allows us to gain an extra logarithmic factor if $\pi \in \{N^{-1}, p\}$, which yields best possible tail estimates in the applications of Section 6.1. We think of (70) as a 'balancedness' condition, and mainly have parameters of form $\pi \in \{1, N^{-1}, p\}$ in mind. In fact, for $\pi \in \{N^{-1}, p\}$ the technical assumption (71) usually holds automatically for small τ (see Remark 31 and the proof of Theorem 36).

Theorem 30 (*Easy-to-apply upper tail inequality*). *Given $\mathcal{H}$ with $1 \leq \ell \leq q \leq k$, assume that (Hℓ), (P) and (Pq) hold. If there are constants $A, \alpha, \tau > 0$ and a parameter $\pi \in (0, 1]$ such that*

$$\max_{\ell \leq j < q} \frac{\mu_j}{\max\{\mu^{(q-j)/(q-\ell+1)}, 1\}} \leq A\pi^\alpha, \tag{70}$$

$$A\mu^{1/(q-\ell+1)} \geq \mathbb{1}_{\{\pi > N^{-\tau}\}} \log N, \tag{71}$$

then for $\varepsilon > 0$ we have

$$\mathbb{P}(X(\mathcal{H}) \geq (1+\varepsilon)\mu) \leq (1 + bN^{-\ell}) \exp\left(-c\min\left\{\varphi(\varepsilon)\mu,\ \min\{\varepsilon^2, 1\}\mu^{1/(q-\ell+1)}\log(e/\pi)\right\}\right)$$

$$\leq (1 + bN^{-\ell}) \exp\left(-d\min\{\varepsilon^2, 1\}\min\{\mu,\ \mu^{1/(q-\ell+1)}\log(e/\pi)\}\right), \tag{72}$$

where $b = 3q$, $c = c(\ell, q, k, L, D, A, \alpha, \tau) > 0$ and $d = c/3$.

Remark 31. If $\pi = N^{-1}$, then (71) is trivially satisfied for $\tau = 1/2$, and $\log(e/\pi) \geq \log N$ holds in (72).

Simple applications of the inductive approaches [19,15,32] often implicitly assume (70) with $\pi = 1$, and replace (71) by the stronger assumption $\min\{\varepsilon^2, 1\}\mu^{1/(q-\ell+1)} = \omega(\log N)$, say (see, e.g., the proof of Corollary 6.3 in [32] or Theorem 2.1 in [31]). Their conclusion is then of the form $\mathbb{P}(X(\mathcal{H}) \geq (1+\varepsilon)\mu) \leq \exp(-a\min\{\varepsilon^2, 1\}\mu^{1/(q-\ell+1)})$, where $\mu^{1/(q-\ell+1)} = \min\{\mu, \mu^{1/(q-\ell+1)}\log(e/\pi)\}$ holds by assumption. In other words,

our inequality (72) yields an extra logarithmic factor when $\pi \in \{N^{-1}, p\}$ in (70). To illustrate this, for subgraph counts in $G_{n,p}$ the setup of Example 21 (with $\ell = 1$, $q = k = e$ and $N = n^2$) naturally yields

$$\max_{\ell \le j < q} \frac{\mu_j}{\mu^{(q-j)/(q-\ell+1)}} \le \max_{1 \le j < e} \frac{O(\sum_{J \subseteq H : e_J = j} n^{v-v_J} p^{e-j})}{\Theta((n^v p^e)^{(e-j)/e})} \le O\Big(\sum_{J \subseteq H : 1 \le e_J < e} n^{e_J v/e - v_J}\Big),$$

which is well-known to be $O(n^{-\beta})$ for so-called 'strictly balanced' graphs and $O(1)$ for 'balanced' graphs (the details are deferred to (104) and (115) in Section 6.2; see also Section 6.3 in [32]).

The next upper tail result assumes that all the parameters μ_j are decaying polynomially in N, which typically requires that $\mu = \mathbb{E} X(\mathcal{H})$ is small (as $v(\mathcal{H}) \le N$). On first reading of Theorem 32 the reader may wish to set $\ell = 1$, $q = k$ and $K = 1$, so that (74) is of form $\mathbb{P}(X(\mathcal{H}) \ge \mu + t) \le \exp(-a\min\{t^2/\mu, t^{1/k}\log N\})$ when $t \in [1, \mu]$. Here our main novelty is the $t^{1/k}\log N$ term, which is key for the applications in Section 6.2.1.

Theorem 32 *(Easy-to-apply upper tail inequality: the small expectations case). Given $\mathcal{H}$ with $1 \le \ell \le q \le k$, assume that (Hℓ), (P) and (Pq) hold. If there are constants $A, \alpha > 0$ such that*

$$\max_{\ell \le j < q} \mu_j \le AN^{-\alpha}, \tag{73}$$

then for $t, K > 0$ we have

$$\mathbb{P}(X(\mathcal{H}) \ge \mu + t) \le (1 + bN^{-q}) \exp\Big(-\min\big\{c\varphi(t/\mu)\mu, \ \max\{ct^{1/(q-\ell+1)}, K\}\log N\big\}\Big)$$
$$\le (1 + bN^{-q}) \exp\Big(-\min\big\{dt^2/\mu, \ dt, \ \max\{ct^{1/(q-\ell+1)}, K\}\log N\big\}\Big), \tag{74}$$

where $b = 2q$, $c = c(\ell, q, k, L, D, A, \alpha, K) > 0$ and $d = c/3$.

The inductive approaches [30,15] yield variants of (74) where $\max\{ct^{1/(q-\ell+1)}, K\}$ is qualitatively replaced by K (see, e.g., Corollary 4.10 in [15]). For K large enough this gives bounds of the form $\mathbb{P}(X(\mathcal{H}) \ge (1 + \varepsilon)\mu) \le N^{-\beta}$ for $\mu \ge C(\varepsilon, d, \beta)\log n$, and $\mathbb{P}(X(\mathcal{H}) \ge (1 + \varepsilon)\mu) \le \exp(-d\varepsilon^2\mu)$ for $\mu \le \log n$ and $\varepsilon \le 1$, say (see, e.g., Corollaries 4.11–4.12 in [15]). To illustrate assumption (73), for subgraph counts in $G_{n,p}$ with $p = O(n^{-v/e+\sigma})$, the setup of Example 21 (with $\ell = 1$, $q = k = e$ and $N = n^2$) yields $\mu = O(n^{e\sigma})$ and

$$\max_{\ell \le j < q} \mu_j \le O\Big(\sum_{J \subseteq H : 1 \le e_J < e} n^{v-v_J} p^{e-e_J}\Big) \le O\Big(\sum_{J \subseteq H : 1 \le e_J < e} n^{e_J v/e - v_J + \sigma(e - e_J)}\Big),$$

which for 'strictly balanced' graphs is well-known to be $O(n^{-\sigma/2})$ for sufficiently small $\sigma > 0$ (the details are deferred to (104) and (107) in Section 6.2; see also Claim 6.2 in [32]).

5.2. More general tail inequalities

In this section we state some more general upper tail inequalities which (i) mimic the heuristic discussion of Section 3.1.2, and (ii) are easier to compare with the work of Kim–Vu/Janson–Ruciński [19,30,32,15]; the proofs are deferred to Section 5.3. Readers primarily interested in applications may proceed to Section 6.

We start with a rigorous analogue of the basic upper tail inequality (23) from Section 3.1.2, which is inspired by very similar classical results for the special case $\mathcal{G} = \mathcal{H}_p$ with $\Delta_q(\mathcal{H}) \leq D$ (see, e.g., Theorem 3.10 in [15] and Theorem 4.2 in [32]). In applications convenient choices of the parameters $(R_j)_{\ell \leq j < q}$ and D are often of form $D = \Theta(1)$, $R_j = \lambda^{q-j}D$ and $\lambda = B \max\{\mu^{1/(q-\ell+1)}, 1\}$, so that in (76) we have $\min\{\mu/R_\ell = \Theta(\lambda)$ and $R_j/R_{j+1} = \lambda$ when $\mu \geq 1$ (see, e.g., the proof of Corollary 6.3 in [32] or Theorem 2.1 in [31]).

Claim 33 *(Basic upper tail inequality). Given $\mathcal{H}$ with $1 \leq \ell \leq q \leq k$, assume that (H$\ell$) and (P) hold. Suppose that $t > 0$. Given positive $(R_j)_{\ell \leq j < q}$ and D, let $R_q = D$. If inequality*

$$\left(\frac{e\mu_j}{R_j}\right)^{R_j/R_{j+1}} \leq N^{-4kj} \tag{75}$$

holds for all $\ell \leq j < q$, then there are $a, b > 0$ (depending only on ℓ, k, L) such that

$$\mathbb{P}(X(\mathcal{G}) \geq \mu + t \text{ and } \Delta_q(\mathcal{G}) \leq D \text{ for some } \mathcal{G} \subseteq \mathcal{H}_p)$$

$$\leq \exp\left(-\frac{a\varphi(t/\mu)\mu}{R_\ell}\right) + \sum_{\ell \leq j < q} N^{-j}\left(\frac{e\mu_j}{R_j}\right)^{bR_j/R_{j+1}}. \tag{76}$$

To familiarize the reader with the form of assumption (75) and inequality (76), it is instructive to briefly relate them to work of Kim and Vu [19,31,32]. Theorem 4.2 in [32] qualitatively sets $t = \sqrt{\lambda \mu R_\ell}$, and (in our notation) its parametrization assumes roughly $\Delta_q(\mathcal{H}) \leq D = R_q$, $\mu/R_\ell \geq \lambda = \omega(\log N)$, as well as $R_j \geq 2e\mu_j$ and $R_j/R_{j+1} \geq \lambda$ for all $\ell \leq j < q$, say. In this case $(e\mu_j/R_j)^{R_j/R_{j+1}} \leq 2^{-\lambda} = N^{-\omega(1)}$ follows, so assumption (75) holds. We also have $t = \mu\sqrt{\lambda R_\ell/\mu} \leq \mu$, so that Remark 11 yields $\varphi(t/\mu)\mu/R_\ell \geq t^2/(3\mu R_\ell) = \lambda/3$, say. Recalling $\Delta_q(\mathcal{H}) \leq D$, for suitable $C = C(q)$ and $c = c(a, b)$ it follows that (76) yields

$$\mathbb{P}(X(\mathcal{H}_p) \geq \mu + t) \leq \exp(-a\lambda/3) + \mathbb{1}_{\{q>\ell\}}qN^{-\ell}2^{-b\lambda} \leq C\exp(-c\lambda), \tag{77}$$

which is of similar form as (24) or Theorem 4.2 [32].

We now state our improved variant[6] of Claim 33, which corresponds to a rigorous analogue of the upper tail inequality (26) from Section 3.1.2. Convenient choices of the parameters $(R_j)_{\ell \leq j < q}$ and $(D_j)_{\ell \leq j \leq q}$ are often of form $D_j = B^{q-j} D_q = \Theta(1)$, $R_j = \lambda^{q-j} D_q$ and $\lambda = B \max\{\mu^{1/(q-\ell+1)}, 1\}$, so that in (80) we have $R_j/R_{j+1} = \lambda$ and $t/R_\ell = \Theta(\lambda)$ when $t = \Theta(\mu)$ and $\mu \geq 1$. One key novelty of (80) is the $\mu/Q_\ell = \min\{\mu s/R_\ell, \mu/D_\ell\}$ term, which intuitively allows us to sharpen inequality (76) whenever $R_j = \omega(\mu_j)$ holds (by using $s = \omega(1)$ in (78), so that usually $\mu/Q_\ell = \omega(\mu/R_\ell)$ in (80), say).

Theorem 34 *(Extended upper tail inequality). Given $\mathcal{H}$ with $1 \leq \ell \leq q \leq k$, assume that (H$\ell$) and (P) hold. Suppose that $s \geq 1$ and $t > 0$. Given positive $(R_j)_{\ell \leq j < q}$ and $(D_j)_{\ell \leq j \leq q}$ with $R_j \geq D_j$, define*

$$Q_j = \max\{R_j/s, D_j\} \tag{78}$$

for $\ell \leq j < q$, and $R_q = Q_q = D_q$. If inequality

$$\max\left\{\left(\frac{e\mu_j}{Q_j}\right)^{R_j/R_{j+1}}, \; \mathbb{1}_{\{Q_j < R_j \text{ and } Q_{j+1}=D_{j+1}\}}\left(\frac{e\mu_j}{Q_j}\right)^{Q_j/D_{j+1}}\right\} \leq N^{-4kj} \tag{79}$$

holds for all $\ell \leq j < q$, then for $a = 1/\left(4L\binom{k}{\ell}\right)$, $b = 1/(2k)$ and $d = 1/\left(4Lqk\binom{k}{\ell}\right)$ we have

$$\mathbb{P}(X(\mathcal{G}) \geq \mu + t \text{ and } \Delta_q(\mathcal{G}) \leq D_q \text{ for some } \mathcal{G} \subseteq \mathcal{H}_p)$$
$$\leq \exp\left(-\frac{a\varphi(t/\mu)\mu}{Q_\ell}\right) + 2\sum_{\ell \leq j < q} N^{-j}\left(\frac{e\mu_j}{Q_j}\right)^{bR_j/R_{j+1}}$$
$$+ \sum_{\ell \leq j < q} \mathbb{1}_{\{Q_j < R_j \text{ and } Q_{j+1}=D_{j+1}\}} N^{-j}\left(\frac{e\mu_j}{Q_j}\right)^{\max\{dt/(R_\ell D_{j+1}), \, bQ_j/D_{j+1}\}}. \tag{80}$$

To illustrate Theorem 34, in the applications of Sections 5.3.2 and 6.1 we have $e\mu_j/R_j \leq p^\alpha/e$ with $p \in (0,1]$, in which case $s = \log(e/p^{\alpha/2})$ is a convenient choice. Indeed, $x\log(e/x) \leq 1$ then implies $e\mu_j/Q_j \leq e\mu_j s/R_j \leq p^{\alpha/2}/e = e^{-s}$. We thus think of the (79) as a minor variant of the assumption (75) from Claim 33 (note that $e\mu_j/R_j \leq e^{-s}$ holds, and that $Q_j < R_j$ implies $Q_j = R_j/s$). Using $D_j = \Theta(1)$ and the additional Kim–Vu type assumptions discussed below Claim 33, we now review inequality (80) of Theorem 34. Since $1/Q_\ell = \min\{s/R_\ell, 1/D_\ell\}$, using $t/R_\ell = \sqrt{\lambda\mu/R_\ell} \geq \lambda$ we obtain analogous to (77) an estimate of the form

---

$$\mathbb{P}(X(\mathcal{H}_p) \geq \mu + t) \leq \exp\left(-\tilde{a}\min\{t^2/\mu, \lambda s\}\right) + \mathbb{1}_{\{q>\ell\}} 3qN^{-\ell} e^{-\tilde{d}\lambda s}$$

$$\leq C\exp\left(-c\min\{t^2/\mu, \lambda \log(e/p)\}\right). \tag{81}$$

If $q > \ell$ then $t^2/\mu = \lambda R_\ell \geq \lambda^{q-\ell+1} R_q = \omega(\lambda \log N)$, so (81) usually decays like $C\exp(-c\lambda \log(e/p))$. When $\lambda \approx \mu^{1/(q-\ell+1)}$ or $t = \varepsilon\mu$ we similarly see that (81) decays like $C\exp(-c\min\{\mu, \lambda \log(e/p)\})$. In all these cases we thus improve the exponential decay of the classical bound (77) by an extra logarithmic factor.

The following upper tail inequality for polynomially small μ_j is a minor extension of Theorem 32. Note that (82) decays exponentially in $\min\{t^2/\mu, t^{1/(q-\ell+1)} \log N\}$ for $1 \leq t \leq O(\mu)$, which seems quite informative when $\mu = \Theta(\text{Var } X(\mathcal{H}))$ holds (i.e., in the Poisson range).

Theorem 35 *(Upper tail inequality: the small expectations case). Given $\mathcal{H}$ with $1 \leq \ell \leq q \leq k$, assume that (H$\ell$) and (P) hold. If there are $A, \alpha > 0$ such that inequality (73) holds, then for $t, K > 0$ we have*

$$\mathbb{P}(X(\mathcal{G}) \geq \mu + t \text{ and } \Delta_q(\mathcal{G}) \leq D \text{ for some } \mathcal{G} \subseteq \mathcal{H}_p)$$

$$\leq \exp\left(-a\varphi(t/\mu)\mu\right) + \mathbb{1}_{\{q>\ell\}} 2qN^{-q} \exp\left(-\max\{bt^{1/(q-\ell+1)}, K\} \log N\right), \tag{82}$$

where $a, b > 0$ depend only on $\ell, q, k, L, D, A, \alpha, K$.

5.3. Proofs

5.3.1. Proofs of Claim 33 and Theorems 34–35

Combining Theorem 15 and 18, by setting $S_j = R_j/s$ the proof of Theorem 34 is straightforward.

Proof of Theorem 34. We first consider the special case $q = \ell$. Since $R_q = D_q$, using $s \geq 1$ we thus infer $\max\{R_\ell/s, D_\ell\} = D_\ell = R_\ell$. Hence (64) of Theorem 26 readily implies (80).

In the remainder we focus on the more interesting case $q > \ell$. Analogous to the proof of Theorem 18, inequality (80) is trivial when $N < 1$ (the left hand side is zero). So we henceforth may assume $N \geq 1$, and using the assumption (79) it follows that $Q_j \geq e\mu_j$. Let $S_j = R_j/s$, and recall that $Q_j = \max\{S_j, D_j\}$ in Theorem 23. Note that $s \geq 1$ and $R_j \geq D_j$ imply $Q_j \leq R_j$. In view of (60) and (64) of Theorem 23 and 26, it remains to estimate $\mathbb{P}_{j,1}$, $\mathbb{P}_{j,2}$ and $\mathbb{P}_{j,3,\ell}$ defined in (34)–(36). Starting with $\mathbb{P}_{j,1}$ and $\mathbb{P}_{j,2}$, using (65) together with $R_j \geq Q_j$, $Q_j/S_{j+1} \geq S_j/S_{j+1} = R_j/R_{j+1}$ and the assumption (79) we infer

$$\mathbb{P}_{j,1} + \mathbb{P}_{j,2} \leq N^j \left(\frac{e\mu_j}{R_j}\right)^{R_j/(kR_{j+1})} + N^j \left(\frac{e\mu_j}{Q_j}\right)^{Q_j/(kS_{j+1})}$$

$$\leq 2N^j \left(\frac{e\mu_j}{Q_j}\right)^{R_j/(kR_{j+1})} \leq 2N^{-j} \left(\frac{e\mu_j}{Q_j}\right)^{R_j/(2kR_{j+1})}. \tag{83}$$

Finally, for $\mathbb{P}_{j,3,\ell}$ of (36) we henceforth tacitly assume $Q_j < R_j$ and $Q_{j+1} = D_{j+1}$. With an eye on (66), using $Q_j \geq e\mu_j$ and the assumption (79) we then (with foresight) similarly deduce

$$\Pi := N^j \left(\frac{e\mu_j}{\lceil Q_j \rceil}\right)^{\lceil Q_j \rceil/(kD_{j+1})} \leq N^j \left(\frac{e\mu_j}{Q_j}\right)^{\lceil Q_j \rceil/(kD_{j+1})} \leq N^{-j}\left(\frac{e\mu_j}{Q_j}\right)^{\lceil Q_j \rceil/(2kD_{j+1})}.$$

Since $\lceil x \rceil \geq \max\{x, 1\}$, by applying (66) with $(x, r, y, z) = (Q_j, t/(2qL), D_{j+1}, R_\ell)$ it follows that

$$\mathbb{P}_{j,3,\ell} \leq (\Pi)^{\left\lceil t/(2Lq\binom{k}{\ell}\lceil Q_j \rceil R_\ell)\right\rceil} \leq N^{-j}\left(\frac{e\mu_j}{Q_j}\right)^{\max\{dt/(R_\ell D_{j+1}),\, bQ_j/D_{j+1}\}}.$$

Recalling our tacit assumption for $\mathbb{P}_{j,3,\ell}$, this completes the proof in view of (60), (64) and (83). □

The details of the similar but simpler proof of Claim 33 are omitted (the above proof carries over by setting $s = 1$ and $D_j = R_j$, since $Q_j = \max\{R_j/s, D_j\} = R_j$ implies $\mathbb{P}_{j,2} = \mathbb{P}_{j,3,\ell} = 0$).

For the proof of Theorem 35 we need to define the parameters $(R_j)_{\ell \leq j \leq q}$ and $(D_j)_{\ell \leq j \leq q}$ of Theorem 15 and 18 in a suitable way. Intuitively, we shall set $R_j = \lambda^{q-j}D$, $\lambda = \max\{t^{1/(q-\ell+1)}, B\}$ and $D_j = Q_j = B^{q-j}D = \Theta(1)$, and the crux is that the assumption (73) eventually yields $e\mu_j/x \leq N^{-\Theta(1)}$ in (65)–(66). We shall also exploit the indicators in Theorem 23 for estimating t/R_ℓ in (80), see (86) below.

Proof of Theorem 35. With foresight, let $B = \max\{4qk/\alpha, 2kK/\alpha, Ae/D, 1\}$ and $\lambda = \max\{t^{1/(q-\ell+1)}, B\}$. Define $D_j = S_j = B^{q-j}D$ and $R_j = \lambda^{q-j}D$ for all $\ell \leq j \leq q$. Note that $Q_j = \max\{S_j, D_j\} = D_j$ and $\min\{Q_j, R_j\} = D_j$, so that $\mathbb{P}_{j,2} = 0$ in (35). Combining (60) and (64) of Theorem 23 and 26, we obtain

$$\mathbb{P}(X(\mathcal{G}) \geq \mu + t \text{ and } \Delta_q(\mathcal{G}) \leq D \text{ for some}$$

$$\mathcal{G} \subseteq \mathcal{H}_p) \leq \exp\left(-\frac{\varphi(t/\mu)\mu}{4L\binom{k}{\ell}D_\ell}\right) + \sum_{\ell \leq j < q}\left[\mathbb{P}_{j,1} + \mathbb{P}_{j,3,\ell}\right]. \tag{84}$$

Tacitly assuming $q > \ell$, it remains to estimate $\mathbb{P}_{j,1}$ and $\mathbb{P}_{j,3,\ell}$ defined in (34) and (36). Starting with $\mathbb{P}_{j,1}$, by inserting (73) into (65), using $R_j \geq DB \geq Ae$ and $R_j/R_{j+1} = \lambda \geq B \geq 4qk/\alpha$ we infer

$$\mathbb{P}_{j,1} \leq N^j \left(\frac{e\mu_j}{R_j}\right)^{R_j/(kR_{j+1})} \leq N^q\left(\mu_j/A\right)^{\lambda/k} \leq N^{q-\alpha\lambda/k} \leq N^{-q-\alpha\lambda/(2k)}. \tag{85}$$

For $\mathbb{P}_{j,3,\ell}$, using $\lceil Q_j \rceil \geq Ae$ and $Q_j/D_{j+1} \geq B \geq 4qk/\alpha$ we (with foresight) similarly deduce

$$\Pi := N^j \left(\frac{e\mu_j}{\lceil Q_j \rceil} \right)^{\lceil Q_j \rceil / (kD_{j+1})} \leq N^{-q - \alpha \lceil Q_j \rceil / (2kD_{j+1})}.$$

Note that $\lambda = B$ implies $R_j = D_j = Q_j$. Hence $Q_j < R_j$ ensures $\lambda = t^{1/(q-\ell+1)}$, so that $t/R_\ell = t^{1/(q-\ell+1)}/D$. Recalling $\lceil Q_j \rceil / D_{j+1} \geq B$, by applying (66) with $(x, r, y, z) = (Q_j, t/(2qL), D_{j+1}, R_\ell)$ we thus infer

$$\mathbb{P}_{j,3,\ell} \leq \mathbb{1}_{\{Q_j < R_j\}} (\Pi)^{\lceil t/(2Lq\binom{k}{\ell}\lceil Q_j \rceil R_\ell) \rceil} \leq N^{-q - \max\{\beta t^{1/(q-\ell+1)}/D_{j+1}, \, \alpha B/(2k)\}}, \quad (86)$$

with $\beta = \alpha/(4Lqk\binom{k}{\ell}D)$. With the above estimates (85) and (86) for $\mathbb{P}_{j,1}$ and $\mathbb{P}_{j,3,\ell}$ in hand, using $B \geq 2kK/\alpha$ and $D_{j+1} \leq D_\ell$ it follows by definition of $\lambda = \max\{t^{1/(q-\ell+1)}, B\}$ that

$$\sum_{\ell \leq j < q} [\mathbb{P}_{j,1} + \mathbb{P}_{j,3,\ell}] \leq \mathbb{1}_{\{q > \ell\}} 2qN^{-q} \exp\left(-\max\{bt^{1/(q-\ell+1)}, \, K\} \log N \right),$$

with $b = \min\{\alpha/(2k), \beta/D_\ell\}$. Recalling (84), this establishes (82) with $a = 1/(4L\binom{k}{\ell}D_\ell)$. $\quad\square$

5.3.2. *Proofs of Theorem 30 and 32*

The 'easy-to-apply' inequalities from Section 5.1 are convenient corollaries of Theorems 34–35. Indeed, Remark 11 implies $\varphi(t/\mu)\mu \geq \min\{t^2/\mu, t\}/3$, so Theorem 32 follows readily from Theorem 35. For Theorem 30 the basic strategy is to apply Theorem 34 with $s = \log(e/\pi^{\alpha/2})$, $R_j = \lambda^{q-j}D$, $\lambda = B\max\{\mu^{1/(q-\ell+1)}, 1\}$ and $D_j = B^{q-j}D = \Theta(1)$. The crux is that the assumption (70) eventually yields $e\mu_j/Q_j \leq \pi^{\alpha/2}/e = e^{-s}$ in (79)–(80). As before, the indicators in Theorem 34 facilitate estimating t/R_ℓ in (80), see (89) below.

Proof of Theorem 30. The proof is naturally divided into four parts: (i) introducing definitions, (ii) estimating $e\mu_j/Q_j$, (iii) applying inequality (80) of Theorem 34, and (iv) verifying assumption (79).

Analogous to the proof of Theorem 18 and 34, we may henceforth assume $N \geq 1$. Furthermore, by increasing A or D if necessary, we may of course assume $A, D \geq 1$. With foresight, let $\beta = \alpha/2$ and $s = \log(e/\pi^\beta)$. Set $B = \max\{e^2A/D, 4k^2/(\tau\beta), 4k^2(4A)^q, 1\}$ and $\lambda = B\max\{\mu^{1/(q-\ell+1)}, 1\}$. Define $R_j = \lambda^{q-j}D$ and $D_j = B^{q-j}D$, so that $R_j \geq D_j$ and $R_q = D_q = D$.

Next we estimate $e\mu_j/Q_j$, where $Q_j \geq R_j/s$. Using assumption (70) and $\alpha = 2\beta$, for $\ell \leq j < q$ we have

$$\frac{e\mu_j}{Q_j} \leq \frac{e\mu_j s}{R_j} = \frac{e\mu_j s}{DB^{q-j} \max\{\mu^{(q-j)/(q-\ell+1)}, 1\}} \leq \frac{eA\pi^{2\beta} \log(e/\pi^\beta)}{DB} \leq \frac{\pi^\beta}{e} = e^{-s}, \quad (87)$$

where we tacitly used $\pi \in (0, 1]$ and $x\log(e/x) \leq 1$ for all $x \in [0, 1]$.

We now apply inequality (80) of Theorem 34, deferring the proof of the claim that assumption (79) holds. Using (87) and $R_j/R_{j+1} = \lambda$, note that $X(\mathcal{H}) = X(\mathcal{H}_p)$ and $\Delta_q(\mathcal{H}) \leq D = D_q$ yield

$$\mathbb{P}(X(\mathcal{H}) \geq (1+\varepsilon)\mu) \leq \mathbb{P}(X(\mathcal{G}) \geq \mu + \varepsilon\mu \text{ and } \Delta_q(\mathcal{G}) \leq D_q \text{ for some } \mathcal{G} \subseteq \mathcal{H}_p)$$
$$\leq \exp\left(-\frac{a\varphi(\varepsilon)\mu}{\max\{R_\ell/s, D_\ell\}}\right) \tag{88}$$
$$+ qN^{-\ell}\left[2e^{-b\lambda s} + \max_{\ell \leq j < q} \mathbb{1}_{\{Q_j < R_j\}} e^{-d\varepsilon\mu s/(R_\ell D_{j+1})}\right].$$

Note that $\lambda = B$ implies $R_j = D_j$, in which case $s \geq 1$ yields $Q_j = D_j = R_j$. Hence $Q_j < R_j$ ensures $\lambda = B\mu^{1/(q-\ell+1)}$, so that $R_\ell = (B\mu^{1/(q-\ell+1)})^{q-\ell}D$. Noting $D_{j+1} \leq D_\ell$, it follows that

$$\max_{\ell \leq j < q} \mathbb{1}_{\{Q_j < R_j\}} e^{-d\varepsilon\mu s/(R_\ell D_{j+1})} \leq \exp\left(-\frac{d}{D_\ell B^{q-\ell}D} \cdot \varepsilon\mu^{1/(q-\ell+1)}s\right). \tag{89}$$

Similarly, using $s \geq 1$ we also see that $R_\ell/s > D_\ell$ implies $R_\ell = (B\mu^{1/(q-\ell+1)})^{q-\ell}D$. Hence

$$\exp\left(-\frac{a\varphi(\varepsilon)\mu}{\max\{R_\ell/s, D_\ell\}}\right) \leq \exp\left(-\min\left\{\frac{a}{D_\ell} \cdot \varphi(\varepsilon)\mu, \ \frac{a}{B^{q-\ell}D} \cdot \varphi(\varepsilon)\mu^{1/(q-\ell+1)}s\right\}\right). \tag{90}$$

Remark 11 implies $\min\{\varphi(\varepsilon), 1, \varepsilon\} \geq \min\{\varepsilon^2, 1\}/3$. So, combining (88)–(90), using $s \geq \min\{1, \beta\}\log(e/\pi)$ and $\lambda \geq B\mu^{1/(q+\ell-1)}$ our findings thus establish (72) for suitable $c = c(\varepsilon, k, q, D, L, \alpha) > 0$.

In the following we verify assumption (79), i.e., the claim omitted above. Note that $R_j/R_{j+1} = \lambda \geq B$ and $Q_j/D_{j+1} \geq D_j/D_{j+1} = B$. Using (87), for $\pi \leq N^{-\tau}$ the left hand side of (79) can thus be bounded by

$$\left(\frac{e\mu_j}{Q_j}\right)^B \leq \pi^{\beta B} \leq N^{-\tau\beta B} \leq N^{-4k^2} \leq N^{-4kj}. \tag{91}$$

For $\pi > N^{-\tau}$ we defer the proof of the claim that for $\ell \leq j < q$ we have

$$\min\{\lambda, R_j/D_{j+1}\} \geq 4k^2 \log N. \tag{92}$$

Using (87), $s \geq 1$, $Q_j \geq R_j/s$ and (92) we see that the left hand side of (79) can be bounded by

$$\max\left\{\left(e^{-1}\right)^{R_j/R_{j+1}}, \left(e^{-s}\right)^{R_j/(sD_{j+1})}\right\} \leq \max\left\{e^{-\lambda}, e^{-R_j/D_{j+1}}\right\} \leq N^{-4k^2} \leq N^{-4kj}.$$

To sum up, we have verified (79), assuming that (92) holds for $\pi > N^{-\tau}$. Turning to the remaining claim (92), using assumption (71) we see that $\pi > N^{-\tau}$ implies

$$\lambda \geq B\mu^{1/(q-\ell+1)} \geq B(\log N)/A \geq 4k^2 \log N.$$

Similarly, $\pi > N^{-\tau}$, $\ell \leq j < q$ and $N \geq 1$ imply

$$R_j/D_{j+1} = \lambda^{q-j}/B^{q-j-1} \geq \big(B\mu^{1/(q-\ell+1)}\big)^{q-j}/B^{q-j-1} \geq B\big((\log N)/A\big)^{q-j} \geq 4k^2 \log N,$$

establishing (92). As discussed, this completes the proof of (72). $\quad\square$

6. Applications

In this section we illustrate our concentration techniques, by applying the basic inequalities from Section 5.1 to several pivotal examples. In Section 6.1 we improve previous work of Janson and Ruciński [16] on random induced subhypergraphs, and derive sharp upper tail inequalities for several quantities of interest in additive combinatorics. In Section 6.2 we answer a question of Janson and Ruciński [13] on subgraph counts in binomial random graphs, and improve the main applications of Wolfovitz [37] and Šileikis [25].

6.1. Random induced subhypergraphs

In probabilistic combinatorics, random induced subhypergraphs $\mathcal{H}_p$ are a standard test-bed for upper tail inequalities (see, e.g., Section 3 in the survey [14]). Janson and Ruciński studied the number of randomly induced edges in [16], and one of their principle results concerns k-uniform hypergraphs with $v(\mathcal{H}) = N$ vertices, $e(\mathcal{H}) \geq \gamma N^q$ edges and $\Delta_q(\mathcal{H}) \leq D$ (for easier comparison with Theorem 2.1 in [16], note that $\Delta_j(\mathcal{H}) \leq N^{\max\{q-j,0\}}\Delta_q(\mathcal{H})$ holds). Writing $X = e(\mathcal{H}_p)$ and $\mu = \mathbb{E}X$, they obtained bounds of form

$$\exp\big(-C(\varepsilon)\mu^{1/q}\log(1/p)\big) \leq \mathbb{P}(X \geq (1+\varepsilon)\mu) \leq \exp\big(-c(\varepsilon)\mu^{1/q}\big), \qquad (93)$$

determining $\log \mathbb{P}(X \geq (1+\varepsilon)\mu)$ up to a missing logarithmic factor (in fact, their lower bound needs an extra assumption). For $2 \leq q < k$ the following corollary of Theorem 30 improves the exponential rate of decay of (93) in the more general weighted case. Noteworthily, inequality (94) below closes the $\log(1/p)$ gap left open by Janson and Ruciński [16] (for the special case $q = 2$ this was already resolved in [35]).

Theorem 36 (*Weighted edge-count of random induced subhypergraphs*). *Let $1 \leq q < k$ and $\gamma, D, a, L > 0$. Assume that $\mathcal{H}$ is a k-uniform hypergraph with $v(\mathcal{H}) \leq N$, $e(\mathcal{H}) \geq \gamma N^q$, $\Delta_q(\mathcal{H}) \leq D$, and $w_f \in [a, L]$ for all $f \in \mathcal{H}$. Set $X = w(\mathcal{H}_p)$ and $\mu = \mathbb{E}X$. For any $\varepsilon > 0$ there is $c = c(\varepsilon, k, \gamma, D, a, L) > 0$ such that for all $p \in (0, 1]$ we have*

$$\mathbb{P}(X \geq (1+\varepsilon)\mu) \leq \exp\Big(-c\min\{\mu,\ \mu^{1/q}\log(e/p)\}\Big). \qquad (94)$$

Remark 37. Setting $p = m/v(\mathcal{H})$, inequality (94) also carries over to $\mathcal{H}_m$ as defined in Section 3.5.

Inequality (94) does not always hold in the excluded case $q = k$. A concrete counterexample is the complete k-uniform hypergraph $\mathcal{H} = \mathcal{H}_N$ with $V(\mathcal{H}) = [N]$ and $w_f = 1$. Then $q = k$, $X = \binom{|[N]_p|}{k} \approx |[N]_p|^k/k!$ and $\mu = \binom{N}{k}p^k \approx (Np)^k/k!$. For $\mu = \omega(1)$, $p \leq 1/2$ and $\varepsilon = \Theta(1)$ it is routine to see that $\mathbb{P}(w(\mathcal{H}_p) \geq (1+\varepsilon)\mu) = \exp(-\Theta(Np)) = \exp(-\Theta(\mu^{1/q}))$ holds, i.e., that there is no logarithmic term.

Concerning sharpness of (94), in applications we usually do not consider a single hypergraph $\mathcal{H}$, but sequences of hypergraph $(\mathcal{H}_N)_{N \in \mathbb{N}}$ which are nearly monotone, i.e., where $\mathcal{H}_N \subseteq \mathcal{H}_{N+1}$ holds up to some minor 'defects' (arising, e.g., due to boundary effects). The following remark states that, in this frequent case, the upper tail inequality (94) is best possible up to the value of the parameter c (for $2 \leq q < k$).

Remark 38 *(Matching lower bound)*. Let $2 \leq q < k$ and $\gamma, D, a, L, n_1, n_2 > 0$. Let $(\mathcal{H}_N)_{N \geq n_1}$ be a sequence of k-uniform hypergraphs such that all $\mathcal{H} = \mathcal{H}_N$ satisfy the assumptions of Theorem 36. Assume that there is $\beta \in (0,1]$ such that $e(\mathcal{H}_N \cap \mathcal{H}_M) \geq \beta e(\mathcal{H}_N)$ for all $M \geq N \geq n_2$. Then for all $\varepsilon > 0$ there are $n_0 = n_0(k, \gamma, D, a, L, \beta, n_1, n_2) > 0$ and $C = C(\varepsilon, \gamma, k, q, D, a, L, \beta, n_1, n_2,) > 0$ such that for all $\mathcal{H} = \mathcal{H}_N$ with $N \geq n_0$, setting $X = w(\mathcal{H}_p)$ and $\mu = \mathbb{E}X$, for all $p \in (0,1]$ we have

$$\mathbb{P}(X \geq (1+\varepsilon)\mu) \geq \mathbb{1}_{\{1 \leq (1+\varepsilon)\mu \leq w(\mathcal{H})\}} \exp\left(-C \min\{\mu, \, \mu^{1/q} \log(1/p)\}\right). \tag{95}$$

We omit the proof of Remark 38, which mimics the lower bound techniques from [35] in a routine way.

Proof of Theorem 36. Let $\delta = a\gamma$, and note that $\mu \geq e(\mathcal{H})p^k \cdot \min_{f \in \mathcal{H}} w_f \geq \delta N^q p^k$ (we never use $w_f \geq a$ again, i.e., we could weaken our assumptions). Inequality (94) holds trivially whenever $N < k$ (since then $0 \leq w(\mathcal{H}_p) \leq L \cdot e(\mathcal{H}) = 0$), so we may henceforth assume $N \geq k$. Our main task is to verify the assumptions of Theorem 30. Let $\ell = 1$ and $\tau = q/(2k)$. As $N^{1/2} \geq \log N$ for all $N > 0$, for $p \geq N^{-\tau}$ we have

$$\mu^{1/(q-\ell+1)} = \mu^{1/q} \geq \delta^{1/q} N p^{k/q} \geq \delta^{1/q} N^{1-k\tau/q} \geq \delta^{1/q} N^{1/2} \geq \delta^{1/q} \log N. \tag{96}$$

As discussed in Example 20, using (57) and $|\Gamma_U(\mathcal{H})| \leq v(\mathcal{H})^{q-j} \cdot \Delta_q(\mathcal{H})$, for $1 \leq j < q$ we thus have

$$\mu_j \leq N^{q-j} \cdot D \cdot p^{k-j}. \tag{97}$$

Recalling $\ell = 1$, (96) and $q < k$, there thus is a constant $A = A(D, \delta) > 0$ such that for $1 \leq j < q$ we have

$$\frac{\mu_j}{\mu^{(q-j)/(q-\ell+1)}} \leq \frac{DN^{q-j}p^{k-j}}{(\mu^{1/q})^{q-j}} \leq D\delta^{j/q-1}p^{j(k/q-1)} \leq Ap^{1/q}. \tag{98}$$

Hence assumptions (70)–(71) hold with $\pi = p$ and $\alpha = 1/q$. Using (72) of Theorem 30 it follows that

$$\mathbb{P}(w(\mathcal{H}_p) \geq (1+\varepsilon)\mu) \leq (1 + 3qN^{-1})e^{-\Pi}, \tag{99}$$

where $\Pi = c' \min\{\varepsilon^2, 1\} \min\{\mu, \ \mu^{1/(q-\ell+1)} \log(e/p)\}$ and $c' = c'(\ell, q, k, L, D, A, \delta) > 0$.

The author finds (99) quite satisfactory, but in the literature the usually irrelevant prefactor $1 + 3qN^{-1}$ is often suppressed for cosmetic reasons. Below we shall achieve this by inflating the constant in the exponent (*without* assuming that n, p or Π are large). If $\Pi \geq 6$, then $N \geq k \geq q$ implies $3qN^{-1} \leq 3 \leq \Pi/2$, so that

$$\mathbb{P}(w(\mathcal{H}_p) \geq (1+\varepsilon)\mu) \leq e^{-\Pi + 3qN^{-1}} \leq e^{-\Pi/2}.$$

Otherwise $1 \geq \Pi/6$ holds, in which case $\varepsilon/(1+\varepsilon) \geq \min\{1, \varepsilon\}/2$ and Markov's inequality yield

$$\mathbb{P}(w(\mathcal{H}_p) \geq (1+\varepsilon)\mu) \leq \frac{1}{1+\varepsilon} = 1 - \frac{\varepsilon}{1+\varepsilon} \leq e^{-\varepsilon/(1+\varepsilon)} \leq e^{-\min\{1,\varepsilon\}\Pi/12},$$

establishing (94) for suitable $c = c(\varepsilon, c') > 0$. $\square$

Combining Theorem 36 and Remark 38, we obtain the following convenient upper tail result (see [35] for a similar result in the special case $q = 2$). It applies to many widely-studied objects in additive combinatorics and Ramsey theory, each time closing the logarithmic gap present in previous work, see (93) and [16].

Corollary 39. *Let $2 \leq q < k$ and $\gamma, D, a, L, n_1 > 0$. Let $(\mathcal{H}_n)_{n \geq n_1}$ be k-uniform hypergraphs such that $\mathcal{H}_n \subseteq \mathcal{H}_{n+1}$, $v(\mathcal{H}_n) \leq n$, $e(\mathcal{H}_n) \geq \gamma n^q$, $\Delta_q(\mathcal{H}_n) \leq D$, and $w_f \in [a, L]$ for all $f \in \mathcal{H}_n$. Then for all $\varepsilon > 0$ there are $n_0 = n_0(k, \gamma, D, a, L, n_1) > 0$ and $c, C > 0$ (depending only on $\varepsilon, k, \gamma, D, a, L, n_1$) such that for all $\mathcal{H} = \mathcal{H}_n$ with $n \geq n_0$, setting $X = w(\mathcal{H}_p)$ and $\mu = \mathbb{E}X$, for all $p \in (0, 1]$ we have*

$$\mathbb{1}_{\{1 \leq (1+\varepsilon)\mu \leq w(\mathcal{H})\}} \exp\left(-C\Psi_{q,\mu}\right) \leq \mathbb{P}(X \geq (1+\varepsilon)\mu) \leq \exp\left(-c\Psi_{q,\mu}\right), \tag{100}$$

where $\Psi_{q,\mu} = \min\{\mu, \ \mu^{1/q} \log(1/p)\}$.

In particular, letting the edges of the k-uniform hypergraphs $\mathcal{H}_n$ with vertex-set $V(\mathcal{H}) = [n]$ encode the relevant objects, it is not difficult to check that Corollary 39 with uniform weights $w_f = 1$ implies[7] all the upper tail bounds presented in Examples 2–5 of Section 1.1.1 (using $q = 2$ for k-term arithmetic progressions, $(k, q) = (3, 2)$ for Schur triples,

---

[7] Note that using weights $w_f = 1$ we count unordered objects, i.e., treat the objects as k-sets (if desired, we could also treat them as ordered k-vectors by using non-uniform weights $w_f > 0$, say).

$(k, q) = (4, 3)$ for additive quadruples, and $(k, q) = (r + s, r + s - 1)$ for (r, s)-sums). Motivated by Section 2.1 in [16], we now record a further common generalization of these examples.

Example 40 *(Integer solutions of linear homogeneous systems).* Let $1 \le r \le k - 2$. Let A be a $r \times k$ integer matrix. Following [16], we assume that every $r \times r$ submatrix B of A has full rank, i.e., $\text{rank}(B) = r = \text{rank}(A)$. We also assume that there exists a distinct-valued positive integer solution to $Ax = \underline{0}$, where $x = (x_1, \dots, x_k)$ is a column vector and $\underline{0} = (0, \dots, 0)$ is an r-dimensional column vector. Let the edges of the k-uniform hypergraph $\mathcal{H}_n$ with $V(\mathcal{H}_n) = [n]$ encode solutions $\{x_1, \dots, x_k\} \subseteq [n]$ of the system $Ax = \underline{0}$ with distinct x_i. The discussion of Section 2.1 in [16] implies that $(\mathcal{H}_n)_{n \ge n_1}$ satisfies the assumptions of Corollary 39 with $q = k - r$, so the upper tail inequality (100) holds for $X = e(\mathcal{H}_p)$, say.

6.1.1. Small expectations case

Note that inequality (100) does not guarantee a similar dependence of $c, C > 0$ on ε. Of course, we can also ask for finer results, which determine how the exponential decay of the upper tail depends on ε. The following corollary of Theorem 32 provides a partial answer for small p (see [35] for results which for $q = 2$ cover all p).

Theorem 41. *Let $k \ge 2$. Let $1 \le q \le k$ and $D, L > 0$. Assume that $\mathcal{H}$ is a k-uniform hypergraph with $v(\mathcal{H}) \le N$, $\Delta_q(\mathcal{H}) \le D$ and $\max_{f \in \mathcal{H}} w_f \le L$, where $N \ge 1$. Set $X = w(\mathcal{H}_p)$ and $\mu = \mathbb{E}X$. For all $\sigma, \Lambda > 0$ there are $c = c(\sigma, \Lambda, k, D, L) > 0$ and $d = d(q) \ge 1$ such that for all $p \le \Lambda N^{-(q-1)/(k-1)-\sigma}$ and $t > 0$ we have*

$$\mathbb{P}(X \ge \mu + t) \le d \exp\left(-c \min\left\{ \varphi(t/\mu)\mu, \ t^{1/q} \log N \right\} \right). \tag{101}$$

Furthermore, setting $p = m/v(\mathcal{H})$, inequality (101) also holds with $\mathcal{H}_p$ replaced by $\mathcal{H}_m$.

Assume that $\mathcal{H} = \mathcal{H}_N$ also satisfies $e(\mathcal{H}_N) \ge \gamma N^q$, the monotonicity conditions of Remark 38, $w_f = 1$ and $2 \le q < k$. Mimicking the lower bound arguments from [35], inequality (101) can then shown to be best possible up to the values of d, c for some range of small p (we leave the details to the interested reader).

Proof of Theorem 41. Our main task is to verify assumption (73) of Theorem 32. To this end we exploit that

$$\frac{q - 1}{k - 1} = \max_{1 \le j < q} \frac{q - j}{k - j}.$$

Indeed, using (97) and $N \ge 1$ there thus is a constant $A = A(D, \Lambda) > 0$ such that we have

$$\max_{1\leq j<q} \mu_j \leq \sum_{1\leq j<q} DN^{q-j}p^{k-j} \leq D\sum_{1\leq j<q} \Lambda^{k-j} N^{(q-j)-(k-j)(q-1)/(k-1)-(k-j)\sigma} \leq AN^{-\sigma}.$$

Applying Theorem 32 (with $\sigma = \alpha$ and $K = 1$) now readily establishes inequality (101). □

6.2. Subgraph counts in random graphs

In this section we consider subgraph counts in the binomial random graph $G_{n,p}$, which are pivotal examples for illustrating various concentration methods (see, e.g., [19, 31,32,14,15,12] and Examples 21–22 in Section 4.1.1). We shall discuss two qualitatively different upper tail bounds in Sections 6.2.1 and 6.2.2.

We henceforth tacitly write $X = X_H$ for the number of copies of H in $G_{n,p}$, and set $\mu = \mathbb{E}X = \Theta(n^{v_H}p^{e_H})$. Let us recall some definitions from random graph theory. Writing $d(J) = e_J/v_J$, a graph H is called *balanced* if $e_H \geq 1$ and $d(H) \geq d(J)$ for all $J \subsetneq H$ with $v_J \geq 1$. If this holds with $d(H) > d(J)$, then H is called *strictly balanced*. Writing $d_2(J) = (e_J - 1)/(v_J - 2)$, a graph H is called 2-*balanced* if $e_H \geq 2$ and $d_2(H) \geq d_2(J)$ for all $J \subsetneq H$ with $v_J \geq 3$. If this holds with $d_2(H) > d_2(J)$, then H is called *strictly* 2-*balanced*.

6.2.1. Small deviations: sub-Gaussian type bounds

We first consider *sub-Gaussian type* $\mathbb{P}(X \geq \mu + t) \leq C\exp(-ct^2/\operatorname{Var}X)$ upper tail inequalities. Our main focus is on the Poisson range, where $\operatorname{Var}X \sim \mathbb{E}X = \mu$ holds, which according to Kannan [18] is the more difficult range. For small p the following simple corollary of Theorem 32 extends/sharpens several results from [30,15,25,37, 18,36], and implies Theorem 6. (For balanced and 2-balanced graphs H it is folklore that $\delta_H \geq 1$. Furthermore, with the exception of perfect matchings, all 2-balanced graphs are strictly balanced.)

Theorem 42 *(Subgraph counts in random graphs: small expectations case). Let H be a graph with $v = v_H$ vertices, $e = e_H$ edges and minimum degree $\delta = \delta_H$. Let $X = X_H$ and $\mu = \mathbb{E}X$. Define $s = \min\{v - 1, e - \delta + 1\}$. If H is strictly balanced, then for every $\Lambda > 0$ there are $c = c(\Lambda, H) > 0$ and $C = C(H) \geq 1$ such that for all $n \geq v$, $\varepsilon \in (0, \Lambda]$ and $p \in [0, 1]$ satisfying $\mu^{(s-1)/s} \leq \Lambda \log n$ we have*

$$\mathbb{P}(X \geq (1 + \varepsilon)\mu) \leq C\exp\Big(-c\varepsilon^2\mu\Big). \tag{102}$$

If H is 2-balanced, then for all $\sigma, \Lambda > 0$ there are $c = c(\sigma, \Lambda, H) > 0$ and $C = C(H) \geq 1$ such that for all $n \geq v$, $0 \leq p \leq \Lambda n^{-(v-2)/(e-1)-\sigma}$ and $0 < t \leq \Lambda\min\{(\mu\log n)^{1/(2-1/s)}, \mu\}$ we have

$$\mathbb{P}(X \geq \mu + t) \leq C\exp\Big(-ct^2/\mu\Big). \tag{103}$$

Remark 43. It is well-known that in (102)–(103) we have $\mu = \mathbb{E}X \sim \operatorname{Var} X$ when $p = o(1)$. The proof shows that the constants C can be replaced by $1 + o(1)$, and that (102)–(103) both carry over to $G_{n,m}$. Furthermore, [26] demonstrates that the sub-Gaussian type tail inequality (102) can already fail for balanced graphs H.

To put Theorem 42 into context, in the year 2000 Vu [30] showed that the sub-Gaussian inequality (102) holds for strictly balanced graphs as long as $\varepsilon = O(1)$ and $\mu \leq \log n$ (note that $\varepsilon^2 \mu \sim (\varepsilon \mu)^2 / \operatorname{Var} X$ by Remark 43). Shortly afterwards, this result was reproved via a different method by Janson and Ruciński [15], who also raised the question whether the restriction $\mu = O(\log n)$ is necessary (see Section 6 in [13]). For the special case $\varepsilon = \Theta(1)$ the aforementioned results were yet again reproved by Šileikis [25] in 2012. Our methods allow us (i) to go beyond all these three approaches from 2000–2012, and (ii) to answer the aforementioned question of Janson and Ruciński: inequality (102) still holds in the wider range $\mu = O((\log n)^{1+\xi})$.

Wolfovitz demonstrated the applicability of his sub-Gaussian concentration result [37] via the complete graph K_r and the complete bipartite graph $K_{r,r}$, showing that inequality (103) holds for both strictly 2-balanced graphs in certain ranges of the parameters p, t. Theorem 42 generalizes these main applications from [37] to all 2-balanced graphs (for a slightly wider parameter range). For $n^{-1} \leq p \leq n^{-1/2-\sigma}$ inequality (103) also slightly extends the t–range of two K_3-specific results of Kannan [18] and Wolfovitz [36].

Proof of Theorem 42. The proofs of (102)–(103) are very similar: each time we shall apply Theorem 32 twice, using the two different setups of Examples 21–22. Hence our main task is to check assumption (73).

For (102) we assume that H is strictly balanced, in which case $\delta = \delta_H \geq 1$ is folklore. By assumption there is a constant $\beta = \beta(H) > 0$ such that for all subgraphs $J \subsetneq H$ with $v_J \geq 1$ we have

$$v_J \cdot \frac{e}{v} \geq e_J + \beta \quad \text{and} \quad e_J \cdot \frac{v}{e} \leq v_J - \beta. \tag{104}$$

Using the setup of Example 21, by (58) there is a constant $B_1 > 0$ such that the corresponding μ_j satisfy

$$\max_{1 \leq j < e - \delta + 1} \mu_j \leq B_1 \sum_{J \subseteq H : 1 \leq e_J < e - \delta + 1} n^{v - v_J} p^{e - e_J}. \tag{105}$$

Similarly, using the setup of Example 22, by (59) there is a constant $B_2 > 0$ such that

$$\max_{2 \leq j < v} \mu_j \leq B_2 \sum_{J \subseteq H : 2 \leq v_J < v} n^{v - v_J} p^{e - e_J}. \tag{106}$$

Recalling $s = \min\{v - 1, e - \delta + 1\}$, in our further estimates of (105)–(106) we may assume $s > 1$ (otherwise $H = K_2$ and (105)–(106) are both equal to zero). Recalling

$\mu = \Theta(n^v p^e)$, we now pick $S = S(\Lambda, H) \geq 1$ large enough such that the assumption $\mu^{(s-1)/s} \leq \Lambda \log n$ implies $p \leq S n^{-v/e + \beta/(2e)}$ for all $n \geq v$. Using $\delta = \delta_H \geq 1$ and the density condition (104), it follows that there are constants $B_3, B_4, B_5 > 0$ such that

$$
\begin{aligned}
(105) + (106) &\leq B_3 \sum_{J \subseteq H : v_J \geq 2, e_J < e} n^{v - v_J} p^{e - e_J} \\
&\leq B_4 \sum_{J \subseteq H : v_J \geq 2, e_J < e} n^{e_J v/e - v_J + \beta/2} \leq B_5 n^{-\beta/2}.
\end{aligned}
\tag{107}
$$

Armed with (107), we now apply Theorem 32 with $K = 1$, $A = B_5$ and $\alpha = \beta/4$, using the setup of Example 21 (with $\ell = 1$, $k = e$, $q = e - \delta + 1$ and $N = n^2$) and Example 22 (with $\ell = 2$, $k = q = v$ and $N = n$). So, applying (74) twice, there is a constant $c_1 > 0$ such that for $t = \varepsilon \mu$ we have

$$
\mathbb{P}(X \geq \mu + t) \leq \left(1 + 2 \max\{v_H, e_H\} n^{-1}\right) \exp\left(-c_1 \min\left\{t^2/\mu,\ t,\ t^{1/s} \log n\right\}\right). \tag{108}
$$

Since $t = \varepsilon \mu \leq \Lambda \mu$, we infer $t \geq t^2/(\Lambda \mu)$. Hence, after adjusting the constant c_1, the t-term is irrelevant for the exponent of (108). As $t^{2 - 1/s} \leq (\Lambda \mu)^{1 + (s-1)/s} = O(\mu \log n)$ by assumption, this establishes (102).

For (103) we proceed similarly, assuming that H is 2-balanced. In this case, for all subgraphs $J \subsetneq H$ with $2 \leq v_J < v$, the assumption that H is 2-balanced (and noting that (109) is trivial when $v_J = 2$) implies

$$
\frac{e - e_J}{v - v_J} = \frac{(e - 1) - (e_J - 1)}{(v - 2) - (v_J - 2)} \geq \frac{e - 1}{v - 2}. \tag{109}
$$

Analogous to (107), in Examples 21 and 22 (with $1 \leq j < e - \delta + 1$ and $2 \leq j < v$) the assumption $p \leq \Lambda n^{-(v-2)/(e-1) - \sigma}$ and the density result (109) entail existence of constants $B_6, B_7 > 0$ such that

$$
\mu_j \leq B_6 \sum_{J \subseteq H : v_J \geq 2, e_J < e} n^{(v - v_J) - (e - e_J)(v-2)/(e-1) - (e - e_J)\sigma} \leq B_7 n^{-\sigma}. \tag{110}
$$

Armed with (110), we now obtain (108) by applying Theorem 32 twice (with $A = B_7$ and $\alpha = \sigma/2$) analogous to the proof of (102). Noting $t \leq \Lambda \mu$ and $t^{2 - 1/s} = O(\mu \log n)$ then readily completes the proof of (103). $\square$

Parts of Theorem 42 can be proved in a simpler/more direct way, but in view of the previous work [30,15,25,37,18,36] here the main point is to illustrate that (102)–(103) follow *routinely* from our general bounds.

6.2.2. Large deviations: upper tail problem

Next we consider the classical upper tail problem for subgraph counts, which concerns $\mathbb{P}(X \geq (1 + \varepsilon)\mu)$ for constant $\varepsilon > 0$. Here our general methods usually give much

weaker estimates than modern specialized approaches such as [12,7,6], but it turns out that our methods can *routinely* sharpen results based on classical inductive approaches (which might potentially be useful in other contexts). Indeed, for balanced graphs Kim and Vu used two different inductions (see Sections 6.3 and 6.6 in [32]), which together establish the following tail estimate: if $\varepsilon \leq C$ and $\varepsilon^2 \max\{\mu^{1/(v-1)}, \mu^{1/e}\} = \omega(\log n)$, then

$$\mathbb{P}(X \geq (1+\varepsilon)\mu) \leq \exp\left(-c\varepsilon^2 \max\{\mu^{1/(v-1)}, \mu^{1/e}\}\right). \tag{111}$$

This inequality was reproved by Janson and Ruciński [15] via their alternative inductive method. Using Theorem 30, we shall go beyond both approaches for strictly balanced graphs: (i) we improve the exponential rate of decay by an extra logarithmic factor, and (ii) we remove the restriction to 'large' expectations μ.

Theorem 44. *Let H be a strictly balanced graph with $v = v_H$ vertices and $e = e_H$ edges. Let $X = X_H$ and $\mu = \mathbb{E}X$. For any $\varepsilon > 0$ there is $c = c(\varepsilon, H) > 0$ such that for all $n \geq v$ and $p \in [0, 1]$ we have*

$$\mathbb{P}(X \geq (1+\varepsilon)\mu) \leq \exp\left(-c \min\left\{\mu, \ \max\{\mu^{1/(v-1)}, \mu^{1/e}\} \log n\right\}\right). \tag{112}$$

Remark 45. Writing the exponent of (112) in the form $\exp(-c\Psi)$, the proof shows that $c = c' \min\{\varepsilon^2, 1\}$ with $c' = c'(H) > 0$ suffices when $\min\{\varepsilon^2, 1\}\Psi \geq 1$. Furthermore, inequality (112) also carries over to $G_{n,m}$.

Remark 46. For balanced graphs H, the proof yields the following variant: for all $n \geq v$, $p \geq \xi n^{-v/e+\sigma}$ and $\varepsilon > 0$ we have $\mathbb{P}(X \geq (1+\varepsilon)\mu) \leq \exp(-c\mu^{1/(v-1)} \log n)$, where $c = c(\sigma, \xi, \varepsilon, H) > 0$.

For r-armed stars $H = K_{1,r}$ inequality (112) yields an $\exp\left(-\Omega(\min\{\mu, \mu^{1/r} \log n\})\right)$ exponential decay, which by [27] is best possible for $p \leq n^{-1/r}$ and $\varepsilon = \Theta(1)$. However, for general graphs H other approaches such as [12,7,6] yield better estimates (as mentioned before), so we defer the proof of Theorem 44 to Appendix A.

Acknowledgments

Appendix A. Proofs omitted from Section 6.2.2

In this appendix we give the proof of Theorem 44, which proceeds similar to Theorem 36 and 42. Namely, we prove (112) by two applications of Theorem 30 and Remark 31 (using the setups of Examples 21–22).

Proof of Theorem 44. We first use the setup of Example 21 with $\ell = 1$, $q = k = e$ and $N = n^2$. Using the bound (58) for μ_j, the expectation $\mu = \Theta(n^v p^e)$ and the density result (104), for $1 \le j < e = e_H$ we infer

$$\frac{\mu_j}{\mu^{(q-j)/(q-\ell+1)}} \le \frac{B \sum_{J \subseteq H: e_J = j} n^{v-v_J} p^{e-j}}{(\mu^{1/e})^{e-j}} \le B_1 \sum_{J \subseteq H: e_J = j} n^{e_J v/e - v_J} \le B_2 n^{-\beta}. \quad (113)$$

Applying Theorem 30 and Remark 31 with $A = B_2$ and $\alpha = \beta/2$, there thus is $c_1 > 0$ such that

$$\mathbb{P}(X \ge (1+\varepsilon)\mu) \le (1 + 3e_H n^{-2}) \exp\left(-c_1 \min\{\varepsilon^2, 1\} \min\{\mu, \mu^{1/e} \log n\}\right). \quad (114)$$

Next we use the setup of Example 22 with $\ell = 2$, $k = q = v$ and $N = n$. We distinguish several cases. If $p \le n^{-v/e}$, then using the bound (58) for μ_j and the density result (104), we infer for $2 \le j < v = v_H$ that

$$\mu_j \le B \sum_{J \subseteq H: v_J = j} n^{v-v_J} p^{e-e_J} \le B \sum_{J \subseteq H: 2 \le v_J < v_H} n^{e_J v/e - v_J} \le B_3 n^{-\beta}. \quad (115)$$

Otherwise $p \ge n^{-v/e}$, so $n^v p^e \ge 1$. Note that for $j < v$ we have $(v-j)/(v-1) \ge (v-j)/v + 1/v^2$. Recalling $\ell = 2$ and $q = v$, using (59), $\mu = \Theta(n^v p^e)$ and (104) we infer for $2 \le j < v = v_H$ that

$$\frac{\mu_j}{\mu^{(q-j)/(q-\ell+1)}} \le \frac{\mu_j}{B_4 (n^v p^e)^{(v-j)/v + 1/v^2}} \le \frac{B_5 \sum_{J \subseteq H: v_J = j} p^{v_J e/v - e_J}}{(n^v p^e)^{1/v^2}} \le \frac{B_6 p^\beta}{(n^v p^e)^{1/v^2}}. \quad (116)$$

Distinguishing $n^{-v/e} \le p \le n^{-v/(2e)}$ and $n^{-v/(2e)} \le p \le 1$, we see that

$$\frac{\mu_j}{\mu^{(q-j)/(q-\ell+1)}} \le B_6 \max\{n^{-\beta v/(2e)}, n^{-1/(2v)}\}. \quad (117)$$

Applying Theorem 30 and Remark 31 with $A = \max\{B_3, B_6\}$ and $\alpha = \min\{\beta, \beta v/(2e), 1/(2v)\}$, we deduce

$$\mathbb{P}(X \ge (1+\varepsilon)\mu) \le (1 + 3v_H n^{-1}) \exp\left(-c_2 \min\{\varepsilon^2, 1\} \min\{\mu, \mu^{1/(v-1)} \log n\}\right). \quad (118)$$

Finally, we combine the two upper bounds (114) and (118), and then remove (for cosmetic reasons) the multiplicative prefactor $1 + O(n^{-1})$ analogous to the proof of Theorem 36, which establishes (112). $\square$

For Remark 46 the point is that for balanced graphs H the density condition (104) only holds with $\beta = 0$, so in (116) we need $p \ge \xi n^{-v/e+\sigma}$ to establish (117) with $\le O(n^{-e\sigma/v^2})$, say.

References

[1] A. Baltz, P. Hegarty, J. Knape, U. Larsson, T. Schoen, The structure of maximum subsets of $\{1, \ldots, n\}$ with no solutions to $a + b = kc$, Electron. J. Combin. 12 (2005) 19.

[2] M. Bateman, N.H. Katz, New bounds on cap sets, J. Amer. Math. Soc. 25 (2012) 585–613.

[3] J. van den Berg, H. Kesten, Inequalities with applications to percolation and reliability, J. Appl. Probab. 22 (1985) 556–569.

[4] J. van den Berg, J. Jonasson, A BK inequality for randomly drawn subsets of fixed size, Probab. Theory Related Fields 154 (2012) 835–844.

[5] T. Bloom, A quantitative improvement for Roth's theorem on arithmetic progressions, J. Lond. Math. Soc. 93 (2016) 643–663.

[6] S. Chatterjee, The missing log in large deviations for triangle counts, Random Structures Algorithms 40 (2012) 437–451.

[7] B. DeMarco, J. Kahn, Tight upper tail bounds for cliques, Random Structures Algorithms 41 (2012) 469–487.

[8] R. Graham, V. Rödl, A. Ruciński, On Schur properties of random subsets of integers, J. Number Theory 61 (1996) 388–408.

[9] B. Green, The Cameron–Erdős conjecture, Bull. Lond. Math. Soc. 36 (2004) 769–778.

[10] S. Janson, Poisson approximation for large deviations, Random Structures Algorithms 1 (1990) 221–229.

[11] S. Janson, New versions of Suen's correlation inequality, Random Structures Algorithms 13 (1998) 467–483.

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

[13] S. Janson, A. Ruciński, The deletion method for upper tail estimates, preprint, http://www2.math.uu.se/~svante/papers/sj135_ppt.pdf, 2000.

[14] S. Janson, A. Ruciński, The infamous upper tail, Random Structures Algorithms 20 (2002) 317–342.

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

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

[17] S. Janson, L. Warnke, The lower tail: Poisson approximation revisited, Random Structures Algorithms 48 (2016) 219–246.

[18] R. Kannan, Two new probability inequalities and concentration results, preprint, arXiv:0809.2477v4, 2010.

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

[20] D. Reimer, Proof of the van den Berg-Kesten conjecture, Combin. Probab. Comput. 9 (2000) 27–32.

[21] O. Riordan, L. Warnke, The Janson inequalities for general up-sets, Random Structures Algorithms 46 (2015) 391–395.

[22] V. Rödl, A. Ruciński, Random graphs with monochromatic triangles in every edge coloring, Random Structures Algorithms 5 (1994) 253–270.

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

[24] M. Schacht, Extremal results for random discrete structures, Ann. of Math. 184 (2016) 333–365.

[25] M. Šileikis, On the upper tail of counts of strictly balanced subgraphs, Electron. J. Combin. 19 (2012), 4.

[26] M. Šileikis, L. Warnke, A counterexample to the DeMarco-Kahn upper tail conjecture, Random Structures Algorithms (2019), in press, arXiv:1809.09595.

[27] M. Šileikis, L. Warnke, Upper tail bounds for stars, preprint, arXiv:1901.10637, 2019.

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

[29] R. Spöhel, A. Steger, L. Warnke, General deletion lemmas via the Harris inequality, J. Comb. 4 (2013) 251–271.

[30] V.H. Vu, On the concentration of multivariate polynomials with small expectation, Random Structures Algorithms 16 (2000) 344–363.

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

[32] V.H. Vu, Concentration of non-Lipschitz functions and applications, Random Structures Algorithms 20 (2002) 262–316.

[33] L. Warnke, When does the K_4-free process stop?, Random Structures Algorithms 44 (2014) 355–397.
[34] L. Warnke, On the method of typical bounded differences, Combin. Probab. Comput. 25 (2016) 269–299.
[35] L. Warnke, Upper tails for arithmetic progressions in random subsets, Israel J. Math. 221 (2017) 317–365.
[36] G. Wolfovitz, Sub-Gaussian tails for the number of triangles in $G(n, p)$, Combin. Probab. Comput. 20 (1) (2011) 155–160.
[37] G. Wolfovitz, A concentration result with application to subgraph count, Random Structures Algorithms 40 (2012) 254–267.