THE DELETION METHOD FOR UPPER TAIL ESTIMATES

SVANTE JANSON AND ANDRZEJ RUCIŃSKI

Abstract. We present a new method to show concentration of the upper tail of random variables that can be written as sums of variables with plenty of independence. The method is formally different from a recent method by Kim and Vu, but in many cases it leads to the same results.

Several applications to random graphs are given. In particular, for X_{K_4} being the number of copies of K_4 in the random graph $\mathbb{G}(n,p)$, with $p \le n^{1/2}$, we almost precisely determine the asymptotics of $\ln \mathbb{P}(X_{K_4} \ge 2\,\mathbb{E}\,X_{K_4})$.

1. Introduction

Kim and Vu [6] and Vu [12, 14, 15] have developed a very interesting new method to show concentration of certain random variables, i.e. to obtain upper bounds (typically exponentially small) of the probabilities $\mathbb{P}(X \le \mu - t)$ and $\mathbb{P}(X \ge \mu + t)$, where X is the random variable, $\mu = \mathbb{E}\,X$ and $t > 0$; see also the further references with various applications given in these papers. Two key features of their method are that a basic martingale inequality is used inductively, and that, when applied to a function of some underlying independent random variables, the obtained estimates use the *average* influence of one or several of the underlying variables, in contrast to e.g. Azuma's inequality where the *maximum* influence appears; the latter improvement is crucial for many applications.

In the present paper, we introduce another method, based on ideas by Rödl and Ruciński [8], to obtain bounds for the upper tail $\mathbb{P}(X \ge \mu + t)$. The new method, which we call the deletion method, see Remark 2.10, is formally quite different from the method of Kim and Vu; it is based on different ideas and the basic estimate differs from their results. Nevertheless, in many situations both methods naturally lead to induction yielding very similar estimates. Indeed, in the applications we have tried so far, we obtain, up to the values of inessential numerical constants, the same results as by the method of Kim and Vu. The only exception is Example 6.7 which gives a new and essentially sharp bound on the probability of having e.g. twice as many copies of K_4 as expected in a random graph, improving an earlier bound by Vu [14], but we guess that the new bound could be derived using Kim and Vu's method too.

There are several reasons for presenting the new method, even if we cannot claim that it produces new results. First, in some applications, although the methods yield the same final result, our method may be somewhat easier to

---

Date: October 26, 2000; revised December 18, 2000.

Research of the second author supported by KBN grant 2 P03A 032 16.

apply. In other applications, the required estimates are the same, and we invite the reader to form his or her own opinion by comparing the two methods on the examples in Section 6.

Secondly, the new method is stated in a different and more general setting than Kim and Vu's method, at least in current versions. Kim and Vu generally study variables that can be expressed as polynomials in independent random variables; we have no need for this constraint and instead use certain independence assumptions. Hence it is conceivable that applications will emerge where only the new method can be applied.

Thirdly, applications may emerge where the numerical constants in the results are important. In such cases, we do not know which of the methods can be trimmed to yield the best result.

Fourthly, we want to stimulate more research into these methods. Neither of the methods seems yet to be fully developed and in a final version, and it is likely that further versions will appear and turn out to be important for applications. It would be most interesting to find formal relations and implications between Kim and Vu's method and our new method, possibly by finding a third approach that encompasses both methods. Conversely, it would also be very interesting and illuminating to find applications where the methods yield different results.

Of course, our method has the drawback that it applies to the upper tail only, but this is not serious, since bounds for the lower tail easily are obtained by other well-known methods, see Janson [1], Suen [11] and Janson [2], or the survey in [3, Chapter 2]. As a complement to our estimates for the upper tail, we give in Section 7 as an appendix a new version of Suen's inequality that applies in the setting of our basic theorem. Note that the bounds for the lower tail obtained by these methods often are much better (i.e. show faster decay) than the bounds obtained for the upper tail by the deletion method. Indeed, it seems that in many applications, the lower tail really is much more concentrated than the upper tail, see Example 6.7 for an example and some explanation. Nevertheless, it is convenient to obtain estimates for both tails at the same time, as by Kim and Vu's method, so we leave the question whether the deletion method can be extended to the lower tail as an important open problem.

Problem 1.1. *Does the bound for $\mathbb{P}(X \geq \mu + t)$ in Theorem 2.1 below apply to $\mathbb{P}(X \leq \mu - t)$ too?*

The basic theorem is stated and proved in Section 2, together with some immediate consequences. These results are directly applicable in some situations. In other cases, the basic result may be used repeatedly with an inductive argument. We give in Section 3 several results obtained in that way for rather general situations. These theorems are still a bit technical, and we give in Section 4 several more easily applicable corollaries.

The results in this paper are to a large extent inspired by the results of Kim and Vu. In Sections 5 and 6 we give several examples where we rederive some of their results using our method. We give also a few other applications. For comparisons with other methods, we refer to [5].

We use ln for natural logarithms and lg for logarithms with base 2. If Γ is a set and $k \geq 1$ a natural number, then $[\Gamma]^k$ denotes the family of all subsets $I \subseteq \Gamma$ with $|I| = k$ and $[\Gamma]^{\leq k} := \bigcup_{j=0}^{k} [\Gamma]^j$ denotes the family of all subsets $I \subseteq \Gamma$ with $|I| \leq k$. We use c or C, sometimes with subscripts or superscripts, to denote various constants that may depend on the parameter k only, unless we explicitly give some parameters; we often give explicit values for these constants, but we have not tried to optimize them.

2. The basic theorem

We begin with a general theorem stated for sums of random variables with a dependency graph given for the summands. We need here only the weak version of dependency graphs with independence between a single vertex and the set of its non-neighbours. Cf. Theorem 7.1, where a stronger version is used. Note that, except in trivial cases, we demand $\alpha \sim \alpha$ in the theorem, because a non-constant random variable is not independent of itself; in other words, we define dependency graphs to have loops at every vertex except when the corresponding variable is constant.

Theorem 2.1. *Suppose that Y_α, $\alpha \in \mathcal{A}$, is a finite family of non-negative random variables and that $\sim$ is a symmetric relation on the index set $\mathcal{A}$ such that each Y_α is independent of $\{Y_\beta : \beta \not\sim \alpha\}$; in other words, the pairs (α, β) with $\alpha \sim \beta$ define the edge set of a (weak) dependency graph for the variables Y_α. Let $X := \sum_\alpha Y_\alpha$ and $\mu := \mathbb{E} X = \sum_\alpha \mathbb{E} Y_\alpha$. Let further, for $\alpha \in \mathcal{A}$, $\tilde{X}_\alpha := \sum_{\beta \sim \alpha} Y_\beta$ and*

$$X^* := \max_{\alpha \in \mathcal{A}} \tilde{X}_\alpha.$$

If $t > 0$, then for every real $r > 0$,

$$\begin{aligned} \mathbb{P}(X \geq \mu + t) &\leq \Bigl(1 + \frac{t}{2\mu}\Bigr)^{-r} + \mathbb{P}\Bigl(X^* > \frac{t}{2r}\Bigr) \\ &\leq \Bigl(1 + \frac{t}{\mu}\Bigr)^{-r/2} + \sum_{\alpha \in \mathcal{A}} \mathbb{P}\Bigl(\tilde{X}_\alpha > \frac{t}{2r}\Bigr). \end{aligned}$$

Remark 2.2. In applications, a suitable value of r has to be found that makes both terms in the estimate small; note that the first terms in the estimates decrease with r, while the second terms increase. Of course, the theorem is useless unless we can bound the probability that $\tilde{X}_\alpha$ is large. We will later see several ways of doing this.

For the first terms it is often convenient to use the estimate

$$\Bigl(1 + \frac{t}{\mu}\Bigr)^{-r/2} \leq \begin{cases} e^{-rt/3\mu}, & t \leq \mu; \\ e^{-r/3}, & t \geq \mu; \end{cases}$$

this follows since $\ln(1 + t/\mu) \geq \min(t/\mu, 1) \ln 2$ by concavity, and $\ln 2 > 2/3$.

Proof. Let $\sum^*_{\alpha_1,\dots,\alpha_m}$ denote the sum over all sequences of $\alpha_1,\dots,\alpha_m \in \mathcal{A}$ such that $\alpha_j \not\sim \alpha_j$ for $1 \le i < j \le m$. We first show that for every integer $m \ge 1$, letting $X_+ := \max(X,0)$,

$$\sum_{\alpha_1,\dots,\alpha_m}^{*} Y_{\alpha_1}\cdots Y_{\alpha_m} \ge \prod_{j=0}^{m-1}(X-jX^*)_+ \ge \bigl(X-(m-1)X^*\bigr)_+^m. \tag{2.1}$$

To verify this, suppose that $\alpha_1,\dots,\alpha_{m-1}$ are given. Then

$$\sum_{\alpha\sim\alpha_i \text{ for some } i<m} Y_\alpha \le \sum_{i=1}^{m-1} \tilde{X}_{\alpha_i} \le (m-1)X^*,$$

and consequently

$$\sum_{\alpha\not\sim\alpha_i,\, i\le m-1} Y_\alpha \ge \bigl(X-(m-1)X^*\bigr)_+.$$

Hence,

$$\sum_{\alpha_1,\dots,\alpha_m}^{*} Y_{\alpha_1}\cdots Y_{\alpha_m} \ge \bigl(X-(m-1)X^*\bigr)_+ \sum_{\alpha_1,\dots,\alpha_{m-1}}^{*} Y_{\alpha_1}\cdots Y_{\alpha_{m-1}}$$

and (2.1) follows by induction.

Next, taking the expectations in (2.1) and observing that the factors in each term in the sum are assumed to be independent,

$$\mathbb{E}\bigl(X-(m-1)X^*\bigr)_+^m \le \sum_{\alpha_1,\dots,\alpha_m}^{*} \mathbb{E}\,Y_{\alpha_1}\cdots \mathbb{E}\,Y_{\alpha_m} \le \Bigl(\sum_\alpha \mathbb{E}\,Y_\alpha\Bigr)^m = \mu^m. \tag{2.2}$$

Now, take $m = \lceil r\rceil$. If $X \ge \mu + t$ and $X^* \le t/2r$, then $(m-1)X^* \le t/2$ and thus $X-(m-1)X^* \ge \mu + t/2$. Consequently, using Markov's inequality and (2.2),

$$\begin{aligned}
\mathbb{P}(X \ge \mu+t) &\le \mathbb{P}(X^* > t/2r) + \mathbb{P}\bigl(X-(m-1)X^* \ge \mu + t/2\bigr)\\
&\le \mathbb{P}(X^* > t/2r) + (\mu+t/2)^{-m}\,\mathbb{E}\bigl(X-(m-1)X^*\bigr)_+^m\\
&\le \mathbb{P}\Bigl(X^* > \frac{t}{2r}\Bigr) + \Bigl(\frac{\mu}{\mu+t/2}\Bigr)^m\\
&\le \mathbb{P}\Bigl(X^* > \frac{t}{2r}\Bigr) + \Bigl(1+\frac{t}{2\mu}\Bigr)^{-r}.
\end{aligned}$$

This shows the first inequality in the statement. The second follows easily, using $(1+x/2)^2 > 1+x$ and thus $(1+x/2)^{-1} < (1+x)^{-1/2}$ for $x>0$. ∎

In combinatorial applications, the variables Y_α usually are indexed by subsets of some index set Γ. We then obtain the following estimate.

Theorem 2.3. *Suppose that $\mathcal{H} \subseteq [\Gamma]^{\le k}$ for an integer $k \ge 1$, and that Y_I, $I \in \mathcal{H}$, is a family of non-negative random variables such that each Y_I is independent of $\{Y_J : J\cap I = \emptyset\}$. Let $X := \sum_I Y_I$ and $\mu := \mathbb{E}\,X = \sum_I \mathbb{E}\,Y_I$. Let further, for $I \subseteq \Gamma$, $X_I := \sum_{J\supseteq I} Y_J$ and*

$$X_1^* := \max_{i\in\Gamma} X_{\{i\}}.$$

If $t > 0$, then for every real $r > 0$,

$$\mathbb{P}(X \geq \mu + t) \leq \Bigl(1 + \frac{t}{2\mu}\Bigr)^{-r} + \mathbb{P}\Bigl(X_1^* > \frac{t}{2kr}\Bigr) \leq \Bigl(1 + \frac{t}{\mu}\Bigr)^{-r/2} + \sum_{i \in \Gamma} \mathbb{P}\Bigl(X_{\{i\}} > \frac{t}{2kr}\Bigr).$$

Proof. We apply Theorem 2.1 with $\mathcal{A} = \mathcal{H}$ and $I \sim J$ if $I \cap J \neq \emptyset$, and note that

$$\tilde{X}_I = \sum_{J \cap I \neq \emptyset} Y_J \leq \sum_{i \in I} X_{\{i\}} \leq k X_1^*. \qquad \square$$

In some applications, the summands Y_I satisfy a stronger independence assumption: two common elements are needed for dependence. For example, this is the case for variables that are indexed by subsets of vertices of the random graph $\mathbb{G}(n,p)$, but are functions of edge indicators. (See e.g. [3] for definition of $\mathbb{G}(n,p)$.) In this case, we have the following alternative to Theorem 2.3, which usually gives stronger bounds.

Theorem 2.4. *Suppose that $\mathcal{H} \subseteq [\Gamma]^{\leq k}$ for an integer $k \geq 2$, and that Y_I, $I \in \mathcal{H}$, is a family of non-negative random variables such that each Y_I is independent of $\{Y_J : |J \cap I| \leq 1\}$. Let $X := \sum_I Y_I$ and $\mu := \mathbb{E}\, X = \sum_I \mathbb{E}\, Y_I$. Let further, for $I \subseteq \Gamma$, $X_I := \sum_{J \supseteq I} Y_J$ and*

$$X_2^* := \max_{i \neq j \in \Gamma} X_{\{i,j\}}.$$

If $t > 0$, then for every real $r > 0$,

$$\mathbb{P}(X \geq \mu + t) \leq \Bigl(1 + \frac{t}{2\mu}\Bigr)^{-r} + \mathbb{P}\Bigl(X_2^* > \frac{t}{k(k-1)r}\Bigr) \leq \Bigl(1 + \frac{t}{\mu}\Bigr)^{-r/2} + \sum_{\{i,j\} \in [\Gamma]^2} \mathbb{P}\Bigl(X_{\{i,j\}} > \frac{t}{k(k-1)r}\Bigr).$$

Proof. This time we apply Theorem 2.1 with $I \sim J$ if $|I \cap J| \geq 2$, and note that

$$\tilde{X}_I = \sum_{|J \cap I| \geq 2} Y_J \leq \sum_{\{i,j\} \in [I]^2} X_{\{i,j\}} \leq \binom{k}{2} X_2^*. \qquad \square$$

Remark 2.5. For random graphs, another possibility leading to the same bounds is to use Theorem 2.3 with Γ being the set of edges of the complete graph; nevertheless, Theorem 2.4 is often more convenient and will be useful in Section 3.

As remarked in Remark 2.2, there are several ways to bound the term $\mathbb{P}(\tilde{X}_\alpha > t/2r)$ in Theorem 2.1 and the corresponding terms in Theorems 2.3 and 2.4. It seems that this problem has to be approached on a case to case basis, and that there is room for ingenuity and ad hoc arguments.

In some cases, these terms can be estimated directly, for example using a Chernoff bound for sums of independent variables as in Example 6.6 below. We use a related but more complicated argument, using two Chernoff estimates, in Example 6.7.

In other applications, the terms are naturally estimated by induction; we explore this in detail in Section 3.

The simplest possibility to estimate these probabilities is to choose r so small that they trivially vanish, as in the following corollaries.

Corollary 2.6. *Let the assumptions of Theorem 2.1 hold. Suppose further that M is a number such that $0 \le Y_\alpha \le M$ for each α, and let $\Delta := \max_i |\{j : j \sim i\}|$, the maximum degree of the dependency graph (with loops contributing 1). Then*

$$\mathbb{P}(X \ge \mu + t) \le \Big(1 + \frac{t}{\mu}\Big)^{-t/(4M\Delta)}.$$

Proof. Take $r = t/(2M\Delta)$ in Theorem 2.1 and observe that then $\tilde{X}_\alpha \le \Delta M = t/2r$. □

Corollary 2.7. *Let the assumptions of Theorem 2.3 hold. Suppose further that M is a number such that $0 \le Y_I \le M$ for each I, and let $N := |\Gamma|$ and $\Delta_1 := \max_{i \in \Gamma} |\{J \in \mathcal{H} : i \in J\}|$. Then*

$$\mathbb{P}(X \ge \mu + t) \le \Big(1 + \frac{t}{\mu}\Big)^{-t/(4kM\Delta_1)} \le \Big(1 + \frac{t}{\mu}\Big)^{-t/(4kMN^{k-1})}.$$

Proof. Take $r = t/(2kM\Delta_1)$ in Theorem 2.3 and observe that $\Delta_1 \le N^{k-1}$. □

Corollary 2.8. *Let the assumptions of Theorem 2.4 hold. Suppose further that M is a number such that $0 \le Y_I \le M$ for each I, and let $N := |\Gamma|$ and $\Delta_2 := \max_{i \ne j \in \Gamma} |\{J \in \mathcal{H} : i, j \in J\}|$. Then*

$$\mathbb{P}(X \ge \mu + t) \le \Big(1 + \frac{t}{\mu}\Big)^{-t/(2k(k-1)M\Delta_2)} \le \Big(1 + \frac{t}{\mu}\Big)^{-t/(2k^2MN^{k-2})}.$$

Proof. Take $r = t/(k(k-1)M\Delta_2)$ in Theorem 2.4 and observe that $\Delta_2 \le N^{k-2}$. □

These corollaries yield essentially the same estimate as the one obtained (for a special case) in [3, Proposition 2.44] by another method, based on another idea by Rödl and Ruciński [7]. See also [5].

Note further that for the case of independent summands ($\Delta = 1$ in Corollary 2.6, $k = 1$ in Corollary 2.7 or $k = 2$ in Corollary 2.8), we obtain, at least for $t = O(\mu)$, up to a constant in the exponent, the well-known Chernoff bound, see e.g. [3, Chapter 2].

Remark 2.9. Sometimes, for example when studying random hypergraphs, even stronger independence properties than in Theorem 2.4 may hold; for instance that Y_I is independent of $\{Y_J : |J \cap I| < 3\}$. All such cases are easily handled by Theorem 2.1, and we leave the formulation of analogues of Theorem 2.4 to the reader.

Remark 2.10. The reason that we call this approach "the deletion method" is that the original version stated roughly, in the setting of Theorem 2.3, that with probability at least $1-(1+t/\mu)^{-r}$, it is possible to find a subset E of Γ of order at most rk such that if we delete all Y_I with $I \cap E \neq \emptyset$, then the sum of the remaining Y_I's is at most $\mu + t$, see [8] and [3, Lemma 2.51]. The theorems above combine this with trivial estimates of the deleted terms.

3. Induction

In many cases, Theorem 2.3 can be used inductively. A general setting where this is possible is described by the following set of assumptions, which will be used throughout this section and the next one.

(H1) Let, as above, $X := \sum_I Y_I$, where Y_I, $I \in \mathcal{H} \subseteq [\Gamma]^{\leq k}$ for some finite index set Γ and an integer $k \geq 1$, is a family of non-negative random variables.

Suppose further that $\mathcal{A}$ is another index set and that there is a family ξ_α, $\alpha \in \mathcal{A}$, of independent random variables and a family of subsets $\mathcal{A}_I \subseteq \mathcal{A}$, $I \in [\Gamma]^{\leq k}$, such that each Y_I is a function of $\{\xi_\alpha : \alpha \in \mathcal{A}_I\}$ and, further, $\mathcal{A}_\emptyset = \emptyset$ and $\mathcal{A}_I \cap \mathcal{A}_J = \mathcal{A}_{I \cap J}$ for all $I, J \in [\Gamma]^{\leq k}$.

Let $\mu := \mathbb{E}\, X$ and $N := |\Gamma|$. To avoid trivialities, assume $N > 1$.

Note that although Y_I is defined for $I \in \mathcal{H}$ only, we want $\mathcal{A}$ to be defined for all $I \in [\Gamma]^{\leq k}$. Actually, we can without loss of generality assume that Y_I is defined for all $I \in [\Gamma]^{\leq k}$ too, by setting $Y_I = 0$ for $I \notin \mathcal{H}$, but this is slightly inconvenient in applications.

It is easily seen that the assumptions of Theorem 2.3 hold under (H1). The situation studied here is more special than in Theorem 2.3, but applications are usually of this type. The conditions (H1) are a bit technical, and we give some examples.

Example 3.1. In many applications we simply take $\mathcal{A} = \Gamma$ and $\mathcal{A}_I = I$. In other words, ξ_i, $i \in \Gamma$, are independent random variables and Y_I is a function of $\{\xi_i : i \in I\}$.

Example 3.2. An important special case of Example 3.1 is when each ξ_i is an indicator random variable, i.e. attains the values 0 and 1 only, and $Y_I = \prod_{i \in I} \xi_i$. In other words, the indicator random variables ξ_i describe a random (Bernoulli) subset $\Gamma_{\mathbf{p}}$ of Γ, $\mathbf{p} = (p_1, \dots, p_N)$, where $p_i = \mathbb{P}(\xi_i = 1)$, and X is the number of elements of $\mathcal{H}$ that are contained in $\Gamma_{\mathbf{p}}$.

Example 3.3. We may treat subgraph counts in the random graph $\mathbb{G}(n,p)$ as in Example 3.2, letting Γ be the set of all edges in the complete graph K_n, $\mathcal{H}$ the family of edge sets of copies of a given graph G assumed to have no isolated vertices, and ξ_i the indicator that edge i is present in $\mathbb{G}(n,p)$; we thus take k to be the number of edges in G. (See e.g. [3] for various properties of subgraph counts of $\mathbb{G}(n,p)$.)

Example 3.4. To treat the number of *induced* copies in $\mathbb{G}(n,p)$ of a given graph G with $v(G)$ vertices, we may again let Γ, $\mathcal{A}$, $\mathcal{A}_I$ and ξ_i be as in Examples

3.3 and 3.1, but now letting $\mathcal{H}$ be the family of edge sets of copies of $K_{v(G)}$ and Y_I the indicator of the event that the subgraph of $\mathbb{G}(n,p)$ defined by I is isomorphic to G. Here $k = \binom{v(G)}{2}$.

We now consider some examples where $\mathcal{A}$ and Γ are not the same.

Example 3.5. Subgraph counts can also be treated as follows. Let $\Gamma = V(K_n)$ be the vertex set of the complete graph K_n and let $\mathcal{A} = [\Gamma]^2$ be its edge set. Let ξ_α be the indicator variable showing whether the edge α is present or not in $\mathbb{G}(n,p)$, and let, for $I \subseteq \Gamma$, $\mathcal{A}_I = [I]^2$, the set of all edges in K_n with both endpoints in I. Again, let G be a fixed graph, and let Y_I be the number of copies of G in $\mathbb{G}(n,p)$ that have vertex set I; this time we thus take k to be the number of vertices of G and $\mathcal{H} = [\Gamma]^k$. Induced copies of G can be treated in exactly the same way.

Example 3.6. For substructure counts in random ℓ-uniform hypergraphs, we similarly may take $\mathcal{A} = [\Gamma]^\ell$. Here ℓ can be any positive integer.

Example 3.7. For an example with $\mathcal{A} = \bigcup_{j=1}^2 [\Gamma]^j$ and $\mathcal{A}_I = [I]^{\le 2} \cap \mathcal{A}$, suppose that the vertices in the random graph $\mathbb{G}(n,p)$ are randomly coloured using 7 different colours. Then the number of rainbow 7-cycles, i.e. cycles containing exactly one vertex of each colour, is a sum X of this type; we let ξ_i, $i \in [\Gamma]^1 = \Gamma$, be the colour of vertex i, and ξ_α, $\alpha \in [\Gamma]^2$, be the indicator of edge α. Further examples with such $\mathcal{A}$ are given in [4].

Example 3.8. More generally, we can take any $\mathcal{A} \subseteq \bigcup_{j=1}^\ell [\Gamma]^j = [\Gamma]^{\le \ell} \setminus \{\emptyset\}$, for some ℓ, and $\mathcal{A}_I = [I]^{\le \ell} \cap \mathcal{A}$. For another example with $\ell = 2$ and $\mathcal{A} = \bigcup_{j=1}^2 [\Gamma]^j$, consider the number of extensions of a given type in $\mathbb{G}(n,p)$ with fixed roots $\{1, \dots, r\}$; we take $\Gamma = \{r+1, \dots, n\}$, let $\xi_{\{i,j\}}$, $\{i,j\} \in [\Gamma]^2$, be the random indicator of the edge ij and let ξ_i, $i \in [\Gamma]^1 = \Gamma$, be the random vector of edge indicators $(\xi_{i1}, \dots, \xi_{ir})$.

Subgraph counts in random graphs can thus be treated in two different ways; this is similar to the choice between vertex exposure and edge exposure in martingale arguments. It turns out that in many cases, the approach in Example 3.5 yields better results with the theorems below, although we do not know whether that always holds. One reason why the latter approach is better is that it usually gives a lower value of k; another is that it exhibits the stronger independence assumption in Theorem 2.4.

In order to formulate our results, we need some more notation. Let as above $X_I := \sum_{J \supseteq I} Y_J$ and consider $\mathbb{E}(X_I \mid \xi_\alpha,\, \alpha \in \mathcal{A}_I)$, the conditional expectation of X_I when we fix the values of ξ_α for $\alpha \in \mathcal{A}_I$ (i.e. taking the expectation over ξ_α, $\alpha \notin \mathcal{A}_I$). This is a function of ξ_α, $\alpha \in \mathcal{A}_I$, and we define μ_I to be its maximum (or, in general, supremum):

$$\mu_I := \sup \mathbb{E}(X_I \mid \xi_\alpha,\, \alpha \in \mathcal{A}_I). \tag{3.1}$$

Further let, for $l \le k$,

$$\mu_l := \max_{|I|=l} \mu_I. \tag{3.2}$$

In other words, μ_l is the smallest number such that $\mathbb{E}(X_I \mid \xi_\alpha,\, \alpha \in \mathcal{A}_I) \le \mu_l$ for every $I \in \mathcal{H}$ with $|I| = l$ and every choice of values of ξ_α, $\alpha \in \mathcal{A}_I$.

Note that if $|I| = k$, then $X_I = Y_I$, which is a function of ξ_α, $\alpha \in \mathcal{A}_I$, and consequently, $\mathbb{E}(X_I \mid \xi_\alpha,\, \alpha \in \mathcal{A}_I) = Y_I$ and $\mu_I = \sup Y_I$. Hence,

$$\mu_k = \max_{|I|=k} \sup X_I = \max_{|I|=k} \sup Y_I. \tag{3.3}$$

Moreover, trivially $\mu_0 = \mu = \mathbb{E}\, X$.

Example 3.9. In Example 3.2, μ_I is the expected number of elements $J \in \mathcal{H}$ such that $I \subseteq J \subseteq \Gamma_{\mathbf{p}}$, given that $I \subseteq \Gamma_{\mathbf{p}}$. In the special case $\mathbb{P}(\xi_i = 1) = p$ for all i, we obtain $\mu_I = \sum_{J\in\mathcal{H},\, J\supseteq I} p^{|J|-|I|}$.

We now can state one of our principal results.

Theorem 3.10. *Assume* (H1). *With notation as above, for every $t > 0$ and $r_1, \dots, r_k$ such that*

$$r_1 \cdots r_j \cdot \mu_j \le t, \qquad j = 1, \dots, k, \tag{3.4}$$

we have, with $c = 1/8k$,

$$\mathbb{P}(X \ge \mu + t) \le \Bigl(1 + \frac{t}{\mu}\Bigr)^{-cr_1} + \sum_{j=1}^{k-1} N^j \Bigl(1 + \frac{t}{r_1 \cdots r_j\, \mu_j}\Bigr)^{-cr_{j+1}}. \tag{3.5}$$

Proof. We apply Theorem 2.3 with $r = r_1/4k$ and obtain, letting $t_1 = t/r_1 = t/4kr$,

$$\mathbb{P}(X \ge \mu + t) \le \Bigl(1 + \frac{t}{\mu}\Bigr)^{-r/2} + \sum_{i\in\Gamma} \mathbb{P}(X_{\{i\}} > 2t_1). \tag{3.6}$$

If $k = 1$, we have by (3.3) and (3.4), for every $i \in \Gamma$, $X_{\{i\}} \le \mu_1 \le t/r_1 = t_1$, and the result follows by (3.6). (Alternatively, use Corollary 2.7 with $M = \mu_1$ and $\Delta_1 = 1$).

If $k \ge 2$ we use induction, assuming the theorem to hold for $k - 1$. Fix $i \in \Gamma$ and let $\widetilde{\Gamma} = \Gamma \setminus \{i\}$. Then $X_{\{i\}} = \sum_{I\in\widetilde{\mathcal{H}}} \widetilde{Y}_I$, with $\widetilde{Y}_I = Y_{I\cup\{i\}}$ and $\widetilde{\mathcal{H}} = \{I \subseteq \widetilde{\Gamma} : I \cup \{i\} \in \mathcal{H}\} \subseteq [\widetilde{\Gamma}]^{\le k-1}$. Conditioned on ξ_α, $\alpha \in \mathcal{A}_{\{i\}}$, the random variables $\widetilde{Y}_I$ satisfy the assumptions (H1), with $\widetilde{\mathcal{A}} = \mathcal{A} \setminus \mathcal{A}_{\{i\}}$ and $\widetilde{\mathcal{A}}_J = \mathcal{A}_{J\cup\{i\}} \setminus \mathcal{A}_{\{i\}}$; the numbers defined by (3.1) and (3.2) become $\widetilde{\mu}_I \le \mu_{I\cup\{i\}}$ and $\widetilde{\mu}_l \le \mu_{l+1}$. Note further that, by (3.1), (3.2) and (3.4),

$$\mathbb{E}(X_{\{i\}} \mid \xi_\alpha,\, \alpha \in \mathcal{A}_{\{i\}}) \le \mu_{\{i\}} \le \mu_1 \le t/r_1 = t_1.$$

Consequently, still conditioning on ξ_α, $\alpha \in \mathcal{A}_{\{i\}}$, we can apply the induction hypothesis, with r_j replaced by $\tilde{r}_j = r_{j+1}$ and t replaced by t_1, noting that (3.4) holds for these numbers because

$$\tilde{r}_1 \cdots \tilde{r}_j \cdot \widetilde{\mu}_j \le r_2 \cdots r_{j+1} \cdot \mu_{j+1} \le t/r_1 = t_1.$$

This yields

$$\begin{aligned}\mathbb{P}(X_{\{i\}} > 2t_1) &\le \mathbb{P}\bigl(X_{\{i\}} \ge \mathbb{E}(X_{\{i\}} \mid \xi_\alpha,\ \alpha\in\mathcal{A}_{\{i\}}) + t_1\bigr)\\ &\le \Bigl(1+\frac{t_1}{\mu_1}\Bigr)^{-cr_2} + \sum_{j=1}^{k-2} N^j\Bigl(1+\frac{t_1}{r_2\cdots r_{j+1}\,\mu_{j+1}}\Bigr)^{-cr_{j+2}}\\ &= \Bigl(1+\frac{t}{r_1\mu_1}\Bigr)^{-cr_2} + \sum_{j=2}^{k-1} N^{j-1}\Bigl(1+\frac{t}{r_1\cdots r_j\,\mu_j}\Bigr)^{-cr_{j+1}}. \qquad (3.7)\end{aligned}$$

The same estimate then holds unconditionally, and the result follows from (3.6) and (3.7). ∎

We still have the freedom, and burden, of choosing suitable values of $r_1, \dots r_k$ when applying Theorem 3.10. In the next section, we give several corollaries that are suitable for immediate application, and the impatient reader may proceed there directly.

In the remainder of this section we give some variants of Theorem 3.10 that yield better results under some circumstances.

Stronger independence. In the case of random graphs treated as in Example 3.5, we have the stronger independence property of Theorem 2.4, since we need a common edge, i.e. two common vertices, to get dependence between two variables Y_I (or families of such variables). This is expressed by the following property.

(H2) $\mathcal{A}_I = \emptyset$ when $|I| \le 1$.

In such cases, we can improve the estimate above. Note that there is no r_1 in the following statement.

Theorem 3.11. *Assume* (H1) *and* (H2). *Then, with notation as above, for every $t > 0$ and $r_2, r_3, \dots, r_k$ such that*

$$r_2 r_3 \cdots r_j \cdot \mu_j \le t, \qquad j = 2, \dots, k, \qquad (3.8)$$

we have, with $c = 1/4k^2$,

$$\mathbb{P}(X \ge \mu + t) \le \Bigl(1+\frac{t}{\mu}\Bigr)^{-cr_2} + \sum_{j=2}^{k-1} N^j\Bigl(1+\frac{t}{r_2 r_3\cdots r_j\,\mu_j}\Bigr)^{-cr_{j+1}}. \qquad (3.9)$$

Proof. We apply Theorem 2.4 with $r = r_2/2k^2$ and obtain, letting $t_1 = t/r_2 = t/2k^2 r$,

$$\mathbb{P}(X \ge \mu + t) \le \Bigl(1+\frac{t}{\mu}\Bigr)^{-r/2} + \sum_{\{i,j\}\in[\Gamma]^2} \mathbb{P}\Bigl(X_{\{i,j\}} > 2t_1\Bigr). \qquad (3.10)$$

Each term in the sum is estimated as in the proof of Theorem 3.10; this time we fix two indices $i, j \in \Gamma$, let $\widetilde{\Gamma} = \Gamma \setminus \{i, j\}$ and have $X_{\{i,j\}} = \sum_{I\in\widetilde{\mathcal{H}}} \widetilde{Y}_I$ with $\widetilde{Y}_I = Y_{I\cup\{i,j\}}$ and $\widetilde{\mathcal{H}} = \{I \subseteq \widetilde{\Gamma} : I \cup \{i, j\} \in \mathcal{H}\} \subseteq [\widetilde{\Gamma}]^{\le k-2}$. Conditioned on ξ_α, $\alpha \in \mathcal{A}_{\{i,j\}}$, the random variables $\widetilde{Y}_I$ satisfy (H1), with $\widetilde{\mathcal{A}} = \mathcal{A} \setminus \mathcal{A}_{\{i,j\}}$

and $\widetilde{\mathcal{A}}_J = \mathcal{A}_{J\cup\{i,j\}} \setminus \mathcal{A}_{\{i,j\}}$; the numbers defined by (3.1) and (3.2) become $\widetilde{\mu}_I \le \mu_{I\cup\{i,j\}}$ and $\widetilde{\mu}_l \le \mu_{l+2}$. Moreover, by (3.1), (3.2) and (3.8),

$$\mathbb{E}(X_{\{i,j\}} \mid \xi_\alpha,\, \alpha \in \mathcal{A}_{\{i,j\}}) \le \mu_{\{i,j\}} \le \mu_2 \le t/r_2 = t_1.$$

Consequently, still conditioning on ξ_α, $\alpha \in \mathcal{A}_{\{i,j\}}$, we obtain by Theorem 3.10 with k replaced by $k-2$, r_j replaced by $\tilde{r}_j = r_{j+2}$ and t replaced by t_1,

$$\begin{aligned}\mathbb{P}(X_{\{i,j\}} > 2t_1) &\le \mathbb{P}(X_{\{i,j\}} \ge \mu_{\{i,j\}} + t_1)\\ &\le \Bigl(1+\frac{t_1}{\mu_2}\Bigr)^{-cr_3} + \sum_{j=1}^{k-3} N^j\Bigl(1+\frac{t_1}{r_3\cdots r_{j+2}\,\mu_{j+2}}\Bigr)^{-cr_{j+3}}. \qquad (3.11)\end{aligned}$$

The same estimate then holds unconditionally, and the result follows from (3.10) and (3.11). □

Note that unlike the proof of Theorem 3.10, this proof does not use induction, since the additional independence hypothesis (H2) does not have to be satisfied by the variables $\widetilde{Y}_I$. Instead, we combine Theorem 2.4 and Theorem 3.10, i.e. we combine one application of Theorem 2.4 and repeated applications of Theorem 2.3. This is thus a kind of combination of edge exposure and vertex exposure.

As remarked in Remark 2.9, we sometimes may have even stronger independence properties. For example, for random hypergraphs as in Example 3.6, we need ℓ common vertices to get dependence; more precisely, the following generalization of (H2) holds. (Here ℓ is any integer with $2 \le \ell \le k$.)

(Hℓ) $\mathcal{A}_I = \emptyset$ when $|I| \le \ell - 1$.

We then have the following generalization of Theorem 3.11.

Theorem 3.12. *Assume* (H1) *and* (Hℓ)*, for some $\ell \ge 2$. Then, with notation as above, for every $t > 0$ and $r_\ell, \dots, r_k$ such that*

$$r_\ell \cdots r_j \cdot \mu_j \le t, \qquad j = \ell, \dots, k, \qquad (3.12)$$

we have, with $c = c(k,\ell)$,

$$\mathbb{P}(X \ge \mu + t) \le \Bigl(1+\frac{t}{\mu}\Bigr)^{-cr_\ell} + \sum_{j=\ell}^{k-1} N^j\Bigl(1+\frac{t}{r_\ell\cdots r_j\,\mu_j}\Bigr)^{-cr_{j+1}}. \qquad (3.13)$$

Proof. We apply Theorem 2.1 with $I \sim J$ when $|I \cap J| \ge \ell$, estimate $\tilde{X}_I \le \sum_{J\in[I]^\ell} X_J$ and use conditioning and Theorem 3.10 as in the proof of Theorem 3.11 to estimate $\mathbb{P}(X_J > t/2r\binom{k}{\ell})$ for $|J| = \ell$; we omit the details. □

Further refinements. We define, for $1 \le j \le k$,

$$M_j := \max_{|J|=j} \sup X_J. \qquad (3.14)$$

Hence $M_k = \mu_k$ by (3.3). We then have the following extension of Theorem 3.10 (which is the case $k_0 = k$). It sometimes yields better bounds, but often there is no advantage in taking $k_0 < k$ because typically then M_{k_0} is much larger than μ_{k_0}.

Theorem 3.13. *Assume* (H1)*, and let k_0 be an integer with $1 \le k_0 \le k$. Then, with notation as above, for every $t > 0$ and $r_1, \dots, r_{k_0}$ such that*

$$\begin{aligned} &r_1 \cdots r_j \cdot \mu_j \le t, \qquad j = 1, \dots, k_0 - 1, \\ &r_1 \cdots r_{k_0} \cdot M_{k_0} \le t, \end{aligned} \tag{3.15}$$

we have, with $c = 1/8k$,

$$\mathbb{P}(X \ge \mu + t) \le \Bigl(1 + \frac{t}{\mu}\Bigr)^{-cr_1} + \sum_{j=1}^{k_0-1} N^j \Bigl(1 + \frac{t}{r_1 \cdots r_j\, \mu_j}\Bigr)^{-cr_{j+1}}.$$

Proof. If $k_0 = 1$, we have by (3.15), for every $i \in \Gamma$, $X_{\{i\}} \le M_1 \le t/r_1$, and the result follows by taking $r = r_1/2k$ in Theorem 2.3.

If $k_0 \ge 2$ we use induction; this time on k_0. The same argument as in the proof of Theorem 3.10 completes the proof; we leave the verification to the reader. □

With the stronger independence property (H2), or more generally (Hℓ), we similarly get the following extension of Theorem 3.12.

Theorem 3.14. *Assume* (H1) *and* (Hℓ)*, for some $\ell \ge 2$. Then, with notation as above, for every $t > 0$, $\ell \le k_0 \le k$ and $r_\ell \dots, r_{k_0}$ such that*

$$\begin{aligned} &r_\ell \cdots r_j \cdot \mu_j \le t, \qquad j = \ell, \dots, k_0 - 1, \\ &r_\ell \cdots r_{k_0} \cdot M_{k_0} \le t, \end{aligned}$$

we have for some $c > 0$,

$$\mathbb{P}(X \ge \mu + t) \le \Bigl(1 + \frac{t}{\mu}\Bigr)^{-cr_\ell} + \sum_{j=\ell}^{k_0-1} N^j \Bigl(1 + \frac{t}{r_\ell \cdots r_j\, \mu_j}\Bigr)^{-cr_{j+1}}. \qquad \square$$

Remark 3.15. In most applications, all summands Y_I have $|I| = k$, but we allow the possibility that different cardinalities occur. In that case, we can make another improvement of the estimates above.

Let $X'_I := \sum_{J \supsetneq I} Y_J$, thus omitting the term Y_I, and define

$$\begin{aligned} \mu'_I &:= \sup \mathbb{E}(X'_I \mid \xi_\alpha,\, \alpha \in \mathcal{A}_I), \\ \mu'_l &:= \max_{|I|=l} \mu'_I. \end{aligned}$$

Conditioned on ξ_α, $\alpha \in \mathcal{A}_I$, the difference $X_I - X'_I = Y_I$ is a constant, and thus we can in the induction step (3.7) in the proof of Theorem 3.10 use $X'_{\{i\}}$ instead of $X_{\{i\}}$. This leads to the following result; we omit the details: *We may replace μ_j by μ'_j in* (3.5) (*keeping μ_j in* (3.4))*, and similarly in Theorems 3.11, 3.12, 3.13 and 3.14.*

4. Corollaries

We give in this section several corollaries of the theorems in the preceding section, obtained by suitable choices of r_i. These corollaries are more convenient for applications, and are often as powerful as the theorems. They have, however, more restricted applicability, so we give several different versions to cover different situations. We continue with the notation of Section 3.

We begin with a consequence of Theorem 3.10. The following explicit bounds are widely applicable and form one of our principal results.

Corollary 4.1. *Assume* (H1). *With notation as above, and* $c = 1/12k$, *for every* $t > 0$,

$$\mathbb{P}(X \geq \mu + t) \leq 2N^{k-1} \exp\Bigl(-c \min_{1\leq j\leq k} \Bigl(\frac{t \lg(1+t/\mu)}{\mu_j}\Bigr)^{1/j}\Bigr)$$
$$\leq \begin{cases} 2N^{k-1} \exp\Bigl(-c \min_{1\leq j\leq k} \Bigl(\frac{t^2}{\mu\mu_j}\Bigr)^{1/j}\Bigr), & t \leq \mu; \\ 2N^{k-1} \exp\Bigl(-c \min_{1\leq j\leq k} \Bigl(\frac{t}{\mu_j}\Bigr)^{1/j}\Bigr), & t \geq \mu. \end{cases} \tag{4.1}$$

Proof. We estimate the terms in the sum in (3.5) using (3.4), which implies

$$1 + \frac{t}{r_1 \cdots r_j \, \mu_j} \geq 2. \tag{4.2}$$

Hence, (3.5) yields, writing $\tau = \lg(1 + t/\mu)$ and $c_1 = 1/8k$

$$\mathbb{P}(X \geq \mu + t) \leq 2^{-c_1 r_1 \tau} + \sum_{j=2}^{k} N^{j-1} 2^{-c_1 r_j}. \tag{4.3}$$

We choose $r_1 = r/\tau$ and $r_2, \dots, r_k = r$, where r is the largest number that makes (3.4) hold, i.e.

$$r = \min_{1\leq j\leq k} \Bigl(\frac{t\tau}{\mu_j}\Bigr)^{1/j}.$$

This makes all exponents of 2 in (4.3) equal to $-c_1 r$, and the right hand side of (4.2) can be bounded by

$$2^{-c_1 r} \sum_{j=1}^{k} N^{j-1} < e^{-(c_1 \ln 2) r} (2N^{k-1}).$$

The first estimate follows using $c_1 \ln 2 > 2c_1/3 = 1/12k = c$. The second estimate follows because $\lg(1 + t/\mu) \geq \min(1, t/\mu)$ by concavity. $\square$

Remark 4.2. It is easily seen that the choice of r_j in the proof of Corollary 4.1 is essentially optimal in (4.3); any other choice would make one of the exponents of 2 smaller in absolute value, and thus the corresponding term larger; hence the resulting estimate differs from the optimum in (4.3) by at most the factor $2N^{k-1}$.

When the stronger independence hypothesis (H2) holds, we obtain a stronger result using Theorem 3.11. This is another of our principal results.

Corollary 4.3. *Suppose that* (H1) *and* (H2) *hold. With notation as above, and $c = 1/6k^2$, for every $t > 0$,*

$$\mathbb{P}(X \geq \mu + t) \leq 2N^{k-1} \exp\Bigl(-c \min_{2 \leq j \leq k} \Bigl(\frac{t \lg(1 + t/\mu)}{\mu_j}\Bigr)^{1/(j-1)}\Bigr)$$
$$\leq \begin{cases} 2N^{k-1} \exp\Bigl(-c \min_{2 \leq j \leq k} \Bigl(\frac{t^2}{\mu \mu_j}\Bigr)^{1/(j-1)}\Bigr), & t \leq \mu; \\ 2N^{k-1} \exp\Bigl(-c \min_{2 \leq j \leq k} \Bigl(\frac{t}{\mu_j}\Bigr)^{1/(j-1)}\Bigr), & t \geq \mu. \end{cases} \tag{4.4}$$

Proof. We use Theorem 3.11 with (4.2), now without r_1, choosing $r_2 = r/\tau$ and $r_3, \ldots, r_k = r$, where

$$r = \min_{2 \leq j \leq k} \Bigl(\frac{t\tau}{\mu_j}\Bigr)^{1/(j-1)}. \qquad \square$$

More generally, we similarly obtain from Theorem 3.12 the following. (As above, we can replace $t \lg(1 + t/\mu)$ by t^2/μ when $t \geq \mu$.)

Corollary 4.4. *Assume* (H1) *and* (Hℓ)*, for some $\ell \geq 2$. With notation as above, for every $t > 0$,*

$$\mathbb{P}(X \geq \mu + t) \leq 2N^{k-1} \exp\Bigl(-c \min_{\ell \leq j \leq k} \Bigl(\frac{t \lg(1 + t/\mu)}{\mu_j}\Bigr)^{1/(j-\ell+1)}\Bigr). \qquad \square$$

If we compare Corollaries 4.1 and 4.3, we see that the power in the exponent in Corollary 4.3 is larger. For example, it is often the case that the terms with $j = k$ are the minimum ones; if, for simplicity, further $t = \mu$ and $\mu_k = 1$, then the estimates are, ignoring the factor $2N^{k-1}$, $\exp(-c\mu^{1/k})$ and $\exp(-c\mu^{1/(k-1)})$, respectively. The difference between the two corollaries stems from the fact that the basic estimate Theorem 2.1 is used (unravelling the induction) k times in the proof of Theorem 3.10 and thus of Corollary 4.1, but only $k - 1$ times in the proof of Theorem 3.11 and Corollary 4.3, since we there jump by two in the first step. (Corollary 4.4 with $\ell > 2$ is even better.)

This is typical for this kind of induction; if we apply the basic estimate inductively m times, and want a final estimate of $\exp(-\lambda)$, we need to choose $r_1, \ldots, r_m$ roughly equal to λ, at least, and for the final step we need something like $t/(r_1 \cdots r_m) \geq 1$; hence, again for $t = \mu$, typically $\lambda^m \leq \mu$. Although this is not completely rigorous, it shows that often it is advantageous to avoid too many induction steps.

One way to cut down the number of induction steps is to use Theorems 3.13 and 3.14. Again using (4.2) and choosing r_j as in the proofs above, for the largest r now allowed, we obtain the following corollaries. They are sometimes better than Corollaries 4.1, 4.3 and 4.4, but as remarked above, the advantage gained by taking $k_0 < k$ (and thus reducing the number of induction steps) is often lost because M_{k_0} may be much larger than μ_{k_0}. We omit the proofs.

Corollary 4.5. *Assume* (H1). *With notation as above, and $c = 1/12k$, for every $k_0 \le k$ and $t > 0$,*

$$\mathbb{P}(X \ge \mu + t) \\ \le 2N^{k_0-1} \exp\Big(-c \min\Big(\min_{1\le j\le k_0-1} \Big(\frac{t \lg(1+t/\mu)}{\mu_j}\Big)^{1/j}, \Big(\frac{t\lg(1+t/\mu)}{M_{k_0}}\Big)^{1/k_0}\Big)\Big).$$

□

Corollary 4.6. *Assume* (H1) *and* (Hℓ), *for some $\ell \ge 2$. With notation as above and some $c > 0$, for every $t > 0$ and $\ell \le k_0 \le k$,*

$$\mathbb{P}(X \ge \mu + t) \\ \le 2N^{k_0-1} \exp\Bigg(-c \min\Big(\min_{\ell\le j\le k_0-1} \Big(\frac{t \lg(1+t/\mu)}{\mu_j}\Big)^{1/(j-\ell+1)}, \\ \Big(\frac{t\lg(1+t/\mu)}{M_{k_0}}\Big)^{1/(k_0-\ell+1)}\Big)\Bigg).$$

□

All the corollaries above are useful only when the exponents in them are large. Consider, for simplicity, the case $t \le \mu$. The factor N^{k-1} in Corollary 4.1 is harmless when the exponent is much larger than $(k-1)\ln N$, i.e. if $t^2/\mu \ge C\mu_j \ln^j N$ for some large constant C and all $1 \le j \le k$. On the other hand, the corollary is useless if $t^2/\mu \le c\mu_j \ln^j N$ for some small constant c and some $j \le k$. In such cases, the following version is better; it yields non-trivial results when $t^2/\mu \ge C\mu_j \ln^{j-1} N$, $1 \le j \le k$.

Corollary 4.7. *Assume* (H1). *With notation as above and some $c > 0$, for every $t > 0$,*

$$\mathbb{P}(X \ge \mu + t) \le 2\exp\Bigg(-c\min\Bigg(\min_{1\le j\le k}\Big(\frac{t\lg(1+t/\mu)}{\mu_j}\Big)^{1/j}, \min_{2\le j\le k}\frac{t\lg(1+t/\mu)}{\mu_j \ln^{j-1} N}\Bigg)\Bigg).$$

Proof. As in the proof of Corollary 4.1, we use (4.3), where $\tau = \lg(1+t/\mu)$ and $c_1 = 1/8k$, but now choose $r_1 = r/\tau$ and $r_j = r + kc_1^{-1}\lg N$, $j \ge 2$, with

$$r = \min\Big(\tfrac12 \min_{1\le j\le k}\Big(\frac{t\tau}{\mu_j}\Big)^{1/j}, \min_{1\le j\le k}\frac{c_1^{j-1} t\tau}{\mu_j (2k)^{j-1}\lg^{j-1} N}\Big).$$

(This yields $c = 2^{-(4k-1)}k^{-(2k-1)}\ln^{k-1} 2$ for $k > 1$, which certainly can be improved.) □

Again we obtain a stronger result when (Hℓ) holds.

Corollary 4.8. *Assume* (H1) *and* (Hℓ), *for some $\ell \ge 2$. With notation as above and some $c > 0$, for every $t > 0$,*

$$\mathbb{P}(X \ge \mu + t) \\ \le 2\exp\Bigg(-c\min\Bigg(\min_{\ell\le j\le k}\Big(\frac{t\lg(1+t/\mu)}{\mu_j}\Big)^{1/(j-\ell+1)}, \min_{\ell+1\le j\le k}\frac{t\lg(1+t/\mu)}{\mu_j \ln^{j-\ell} N}\Bigg)\Bigg).$$

Proof. We choose $r_\ell = r/\tau$ and $r_j = r + C \lg N$, $j > \ell$, in Theorem 3.12 and optimize r; we leave the details as an exercise. □

Similarly, we obtain from Theorems 3.13 and 3.14 the following more general results, here condensed into one statement; we omit the proof.

Corollary 4.9. *Let $1 \le \ell \le k_0 \le k$. If $\ell = 1$, assume* (H1), *and if $\ell \ge 2$, assume* (H1) *and* (Hℓ). *With notation as above, let $\bar\mu_j = \mu_j$ for $j < k_0$ and $\bar\mu_{k_0} = M_{k_0}$. Then, for every $t > 0$,*

$$\mathbb{P}(X \ge \mu + t) \le 2 \exp\Bigg(-c \min\bigg(\min_{\ell \le j \le k_0} \Big(\frac{t \lg(1 + t/\mu)}{\bar\mu_j}\Big)^{1/(j-\ell+1)}, \min_{\ell+1 \le j \le k_0} \frac{t \lg(1 + t/\mu)}{\bar\mu_j \ln^{j-\ell} N}\bigg)\Bigg).$$

□

We have so far used (4.2) and (4.3), and the corresponding estimates obtained from the other theorems, but in some situations with t^2/μ small, the full strength of (3.5) etc. is needed. In the following result, we assume that $\mu_1, \dots, \mu_{k-1}$ are small, while μ_k may be 1.

Corollary 4.10. *Assume* (H1). *For every $\alpha, \beta > 0$, there is a constant $c = c(k, \alpha, \beta) > 0$ such that, with notation as above, if $\mu_j \le N^{-\alpha}$ for $1 \le j \le k-1$ and $\mu_k \le 1$, then for $0 < t \le \mu$,*

$$\mathbb{P}(X \ge \mu + t) \le e^{-ct^2/\mu} + N^{-\beta}.$$

Proof. Let $A \ge 1$ be a constant, and choose $r_2, \dots, r_k = A$ and $r_1 = A^{1-k} t$. Then (3.4) is satisfied, and Theorem 3.10 yields

$$\mathbb{P}(X \ge \mu + t) \le \Big(1 + \frac{t}{\mu}\Big)^{-cA^{1-k}t} + \sum_{j=1}^{k-1} N^j (N^\alpha)^{-cA}.$$

The result follows by choosing A so that $c\alpha A = \beta + k$. □

We obtain two immediate corollaries by letting one of the terms on the right hand side dominate the other.

Corollary 4.11. *Assume* (H1). *For every $\alpha, \beta, \varepsilon > 0$, there is a constant $Q = Q(k, \alpha, \beta, \varepsilon) > 0$ such that, with notation as above, if $\mu_j \le N^{-\alpha}$ for $1 \le j \le k-1$, $\mu_k \le 1$, and $\mu \ge Q \ln N$, then*

$$\mathbb{P}\big(X \ge (1 + \varepsilon)\mu\big) \le N^{-\beta}.$$ □

Corollary 4.12. *Assume* (H1). *For every $\alpha > 0$, there is a constant $c = c(k, \alpha) > 0$ such that, with notation as above, if $\mu_j \le N^{-\alpha}$ for $1 \le j \le k-1$, $\mu_k \le 1$, $0 < \varepsilon \le 1$ and $\mu \le \ln N$, then*

$$\mathbb{P}\big(X \ge (1 + \varepsilon)\mu\big) \le 2e^{-c\varepsilon^2\mu}.$$ □

Remark 4.13. Remark 3.15 implies that in Corollaries 4.10–4.12, the assumptions on μ_j may be weakened to $\mu'_j \le N^{-\alpha}$, $1 \le j \le k-1$, and $Y_I \le 1$, $I \in \mathcal{H}$.

5. Relations with Kim and Vu's results

As said earlier, the results in Sections 3 and 4 are inspired by the results and methods in Kim and Vu [6] and Vu [12, 14, 15], where similar induction arguments are used.

The general setting of Kim and Vu is to consider a random variable X which is a polynomial $X(\xi_1, \dots, \xi_N)$ of degree k in N independent random variables $\xi_1, \dots, \xi_N$. (We change their notation to correspond to ours.) It is furthermore assumed that the polynomial has only non-negative coefficients and that $0 \le \xi_i \le 1$; sometimes it is further assumed that the variables ξ_i are binary, i.e. $\xi_i \in \{0, 1\}$.

Let, for a multi-set $A = \{i_1, \dots, i_j\}$, $\partial_A X$ denote the partial derivative $\frac{\partial^j}{\partial \xi_{i_1} \cdots \partial \xi_{i_j}} X(\xi_1, \dots, \xi_N)$, and define $E'_j := \max_{|A|=j} \mathbb{E}(\partial_A X)$ and

$$E_j := \max_{|A| \ge j} \mathbb{E}(\partial_A X) = \max_{l \ge j} E'_l.$$

In particular, $E_0 \ge E'_0 = \mathbb{E}\, X = \mu$.

This setting is an instance of (H1) as in Example 3.1; we take $\Gamma = \mathcal{A} = \{1, \dots, N\}$ and let Y_I be the sum of all terms $a_{i_1 \cdots i_j} \xi_{i_1} \cdots \xi_{i_j}$ in X such that $\{i_1, \dots, i_j\} = I$. (If no variable occurs to higher power than 1, Y_I is just a single term. Example 3.2 is a special case of this case.)

If no variable occurs to a higher power than 1 in X, then $\partial_A X = 0$ if A contains any repetition, and otherwise $\partial_A X$ equals our X_A with ξ_i replaced by 1 for $i \in A$; hence (assuming $\sup \xi_i = 1$ for each i) $\mathbb{E}\, \partial_A X = \mu_A$ and $\mu_j = E'_j$. In general, it is easily seen that

$$\mu_j \le c \sum_{l=j}^{k} E'_l \le C E_j. \tag{5.1}$$

Let us begin with the first estimate by Kim and Vu's method, the main theorem of [6].

Example 5.1. Let $\lambda \ge 1$, and take $t = \sqrt{E_0 E_1} \lambda^k$, $r_1 = c_1 \lambda \sqrt{E_0 / E_1}$ and $r_i = \lambda$ for $i \ge 2$ in Theorem 3.10. By (5.1), $\mu_j \le C E_1$ for every $j \ge 1$; hence it is easily checked that if $c_1 = 1/C$, then (3.4) is satisfied. Moreover, $t/\mu \ge t/E_0 \ge (E_1/E_0)^{1/2}$ and thus

$$r_1 \lg(1 + t/\mu) \ge r_1 \lg(1 + (E_1/E_0)^{1/2}) \ge r_1 (E_1/E_0)^{1/2} = c_1 \lambda.$$

Hence Theorem 3.10 yields the estimate

$$\mathbb{P}\big(X \ge \mu + (E_0 E_1)^{1/2} \lambda^k\big) \le 2 N^{k-1} e^{-c_2 \lambda}, \qquad \lambda \ge 1,$$

which is equivalent to the upper tail part of the main theorem in Kim and Vu [6], see also [15, Theorem 3.1], apart from the numerical value of the constants.

Corollary 4.1 easily yields the somewhat better estimate

$$\mathbb{P}\big(X \ge \mu + (E_0 E_1)^{1/2} \lambda^k\big) \le 2 N^{k-1} \exp\big(-c \min(\lambda^2, \lambda (E_0/E_1)^{1/2k})\big), \qquad \lambda \ge 1.$$

These estimates trivially hold for $\lambda \le 1$ too.

This first result by Kim and Vu is superseded by later results by Vu; we begin with a simplified version.

Example 5.2. Suppose that $\lambda \geq 1$ and that $\mathcal{E}_0, \dots, \mathcal{E}_k$ are numbers such that

$$\mathcal{E}_j \geq E_j, \qquad 0 \leq j \leq k, \tag{5.2}$$

$$\mathcal{E}_j / \mathcal{E}_{j+1} \geq \lambda, \qquad 0 \leq j \leq k-1. \tag{5.3}$$

Take $t = (\lambda \mathcal{E}_0 \mathcal{E}_1)^{1/2}$. Then $t/\mu = t/E'_0 \geq t/\mathcal{E}_0$ by (5.2) and, by (5.3), $t \leq \mathcal{E}_0$; hence $t \lg(1 + t/\mu) \geq t \lg(1 + t/\mathcal{E}_0) \geq t^2/\mathcal{E}_0 = \lambda \mathcal{E}_1$. Consequently, using (5.1) together with (5.2) and (5.3) again,

$$\frac{t \lg(1 + t/\mu)}{\mu_j} \geq \frac{\lambda \mathcal{E}_1}{C \mathcal{E}_j} \geq C^{-1} \lambda^j. \tag{5.4}$$

Hence, Corollary 4.1 yields

$$\mathbb{P}\big(X \geq \mu + (\lambda \mathcal{E}_0 \mathcal{E}_1)^{1/2}\big) \leq 2N^{k-1} e^{-c\lambda}. \tag{5.5}$$

The result in (5.5) is useful when $\lambda \geq C \ln N$; then we can remove the factor N^{k-1}. For smaller λ, the same can be done with a stronger hypothesis.

Example 5.3. If we strengthen (5.3) to

$$\mathcal{E}_j / \mathcal{E}_{j+1} \geq \lambda + j \ln N, \qquad 0 \leq j \leq k-1, \tag{5.6}$$

we see by (5.4) also

$$\frac{t \lg(1 + t/\mu)}{\mu_j} \geq \frac{\lambda \mathcal{E}_1}{C \mathcal{E}_j} \geq C^{-1} \lambda (\ln N)^{j-1}.$$

Corollary 4.7 thus implies

$$\mathbb{P}\big(X \geq \mu + (\lambda \mathcal{E}_0 \mathcal{E}_1)^{1/2}\big) \leq 2e^{-c\lambda}.$$

This result yields the upper tail parts of Theorem 2.3 in [13] and of the case $\tilde{k} = k$ in Theorem 3.2 in [15]. In general, the upper tail part of the latter theorem follows by Corollary 4.9 (with $\ell = 1$). Indeed, Vu [15] inspired both Theorem 3.13 and Corollaries 4.7–4.9.

The results discussed so far in this section all use, in our version, $\mathcal{A} = \Gamma$ as in Example 3.1. The more general setting in (H1) is inspired by Vu [14]. In particular, our Theorems 3.10 and 3.11 and the corresponding Corollaries 4.1 and 4.3, owe much to Theorems 2 and 1, respectively, in [14]. The upper tail parts of these theorems by Vu follow from our Corollaries 4.1 and 4.3 as shown below.

Example 5.4. Vu [14] studies the subgraph count X_G in $\mathbb{G}(n,p)$, where G is a fixed graph with k vertices; see Section 6 below where we do the same in detail. He defines F_j as the minimum of $\mathbb{E}\, X_H$ over all subgraphs H of G with at least j vertices, and $E_j := \mathbb{E}(X_G)/F_j$. Thus $E_0 \geq \mu$ and $E_k = 1$.

Vu further studies the corresponding, more general, problem of counting extensions with a fixed set of roots; we denote this random number by X_L, where L is a given rooted graph with k non-root vertices. Vu defines M_j

corresponding to, and generalizing, E_j; see [14] for details. Again, $M_0 \geq \mu$ and $M_k = 1$.

Note that the subgraph case is an instance of our Example 3.5 where (H2) holds, while the extension case is an instance of Example 3.8 where (H2) fails; thus we (and Vu) obtain better bounds for the subgraph case.

In the subgraph case, it is easily seen that $\mu_j \leq CE_j$, $0 \leq j \leq k$, see (6.1) below; similarly, in the extension case $\mu_j \leq CM_j$.

In his Theorem 2, Vu [14] assumes that $\lambda \geq C \ln n$ and $\mathcal{M}_0, \dots, \mathcal{M}_k$ are numbers, with $\mathcal{M}_k = 1$, such that

$$\begin{aligned} \mathcal{M}_j &\geq M_j, & j &= 0, \dots, k, \\ \mathcal{M}_j / \mathcal{M}_{j+1} &\geq \lambda, & j &= 0, \dots, k-1. \end{aligned}$$

It then follows from Corollary 4.1, exactly as in Example 5.2 above, that

$$\mathbb{P}\big(X_L \geq \mu + (\lambda \mathcal{M}_0 \mathcal{M}_1)^{1/2}\big) \leq 2n^{k-1} e^{-c_1 \lambda} \leq e^{-c\lambda},$$

which is the upper tail part of Vu's result.

In Theorem 1, for subgraphs, Vu [14] assumes, more weakly, that $\lambda \geq C \ln n$ and $\mathcal{E}_0, \mathcal{E}_2, \dots, \mathcal{E}_k$ are numbers, with $\mathcal{E}_k = 1$, such that

$$\begin{aligned} \mathcal{E}_j &\geq E_j, & j &= 0, 2, \dots, k, \\ \mathcal{E}_j / \mathcal{E}_{j+1} &\geq \lambda, & j &= 2, \dots, k, \\ \mathcal{E}_0 / \mathcal{E}_2 &\geq \lambda. \end{aligned}$$

Now take $t = (\lambda \mathcal{E}_0 \mathcal{E}_2)^{1/2}$. Then, as in Example 5.2, $t \leq \mathcal{E}_0$ and, for $2 \leq j \leq k$,

$$\frac{t \lg(1 + t/\mu)}{\mu_j} \geq \frac{t^2/\mathcal{E}_0}{C\mathcal{E}_j} = \frac{\lambda \mathcal{E}_2}{C\mathcal{E}_j} \geq C^{-1} \lambda^{j-1}.$$

Corollary 4.3 yields

$$\mathbb{P}\big(X \geq \mu + (\lambda \mathcal{E}_0 \mathcal{E}_2)^{1/2}\big) \leq 2n^{k-1} e^{-c_1 \lambda} \leq e^{-c\lambda},$$

which is the upper tail part of Vu's result.

Similarly, the upper tail part of Theorem 6 in [14] follows from our Corollary 4.4.

Corollaries 4.10–4.12 are inspired by and strongly related to results in Vu [12]; we have not been able to derive Theorem 1.3 in [12] by our method, but the upper tail parts of its corollaries Theorem 1.2 and Theorem 1.4 follow immediately from our Corollaries 4.11 and 4.12, respectively, together with Remark 4.13. Also Remark 3.15 is inspired by Vu [12].

6. Applications to random graphs

We give in this section several applications of the general results above to random graphs. In order to compare our method with the method of Kim and Vu, we mainly consider applications treated by Vu [12, 14, 15], and rederive several of his results. Many other applications from [6, 12, 14, 15] could be handled similarly. See further [5], where also other methods are considered.

Several of the arguments used below are similar to the arguments of Vu, but sometimes there are differences; we invite the reader to compare the details.

We denote the numbers of vertices and edges of a graph G by $v(G)$ and $e(G)$, respectively.

Let G be a fixed graph, and let X_G be the number of copies of G in the random graph $\mathbb{G}(n,p)$. As explained in Example 3.5, we have $X_G = \sum_{I\in[\Gamma]^k} Y_I$, where $k = v(G)$ and Y_I is the number of copies of G in $\mathbb{G}(n,p)$ with vertex set I, so we are in the setting of Section 3. We have (for $n \ge k$)

$$\mathbb{E}\, X_G = \frac{k!}{\operatorname{aut}(G)}\binom{n}{k} p^{e(G)} \asymp n^k p^{e(G)},$$

where $\operatorname{aut}(G)$ is the number of automorphisms of G and $\asymp$ means that the quotient of the two sides is bounded from above and below by positive constants.

We split the contents of this section into two part according to how large the deviation t is.

Large deviations. Throughout this subsection we write $t = \varepsilon\mu$; the reader may concentrate on the typical case when ε is a constant, but the results cover also cases when ε depends on n. Corollary 4.3 yields the following estimate.

Theorem 6.1. *For any graph G, there exists a constant $c > 0$ such that, for every $\varepsilon > 0$, n and p,*

$$\mathbb{P}\big(X_G \ge (1+\varepsilon)\,\mathbb{E}\, X_G\big) \le 2n^{v(G)-1} \exp\Big(-c \min_{H\subseteq G: v(H)\ge 2} \big(\lg(1+\varepsilon)\varepsilon\, \mathbb{E}\, X_H\big)^{1/(v(H)-1)}\Big).$$

If further $\lg(1+\varepsilon)\varepsilon\,\mathbb{E}\,X_H \ge C \ln^{v(H)-1} n$ for some large constant C and every $H \subseteq G$ with $v(H) \ge 2$, then

$$\mathbb{P}\big(X_G \ge (1+\varepsilon)\,\mathbb{E}\, X_G\big) \le \exp\Big(-c \min_{H\subseteq G: v(H)\ge 2} \big(\lg(1+\varepsilon)\varepsilon\, \mathbb{E}\, X_H\big)^{1/(v(H)-1)}\Big).$$

Proof. The first estimate follows directly by Corollary 4.3, since if $1 \le j \le k = v(G)$, then

$$\mu_j \asymp n^{k-j} \max_{H\subseteq G,\, v(H)=j} p^{e(G)-e(H)} \asymp \max_{H\subseteq G,\, v(H)=j} \frac{\mathbb{E}\, X_G}{\mathbb{E}\, X_H}. \tag{6.1}$$

The second follows (with a smaller c) since the exponent now is at least $k \log n$. ∎

Recall that the graph G is said to be *balanced* if $e(H)/v(H) \le e(G)/v(G)$ for every $H \subseteq G$, see [3].

Corollary 6.2. *If G is a balanced graph, there exist constants $c, C > 0$ such that, for every $\varepsilon > 0$, n and p with $\min(\varepsilon^2, \varepsilon^{1/(v(G)-1)})(\mathbb{E}\,X_G)^{1/(v(G)-1)} \ge C \ln n$,*

$$\mathbb{P}\big(X_G \ge (1+\varepsilon)\,\mathbb{E}\, X_G\big) \le \begin{cases} \exp\big(-c\varepsilon^2(\mathbb{E}\, X_G)^{1/(v(G)-1)}\big), & \varepsilon \le 1,\\ \exp\big(-c(\varepsilon\,\mathbb{E}\, X_G)^{1/(v(G)-1)}\big), & \varepsilon \ge 1.\end{cases}$$

Proof. Since G is balanced, $(\mathbb{E}\, X_H)^{1/v(H)} \asymp np^{e(H)/v(H)} \geq c_1(\mathbb{E}\, X_G)^{1/v(G)}$. It follows a fortiori that $(\mathbb{E}\, X_H)^{1/(v(H)-1)} \geq c_2(\mathbb{E}\, X_G)^{1/(v(G)-1)}$ when $\varepsilon \leq 1$ and thus $\mathbb{E}\, X_G > 1$, and $(\varepsilon\, \mathbb{E}\, X_H)^{1/(v(H)-1)} \geq c_2(\varepsilon\, \mathbb{E}\, X_G)^{1/(v(G)-1)}$ when $\varepsilon \geq 1$ and thus $\varepsilon\, \mathbb{E}\, X_G > 1$. The result follows from the theorem. □

For balanced graphs (and ε bounded), we have obtained the same bound as Vu [14, Theorem 3], up to the values of the constants.

Moreover, Theorem 6.1 applies also to unbalanced graphs. Note that by the argument in the proof of Corollary 6.2, if ε is constant, it suffices to consider $H = G$ and induced subgraphs H with density $e(H)/v(H) > e(G)/v(G)$ when taking the minimum. Note further that the minimum may be attained for $H = G$ also for an unbalanced G, at least for some ranges of p, so that we obtain the same estimate as for balanced graphs.

Example 6.3. Let G be K_4 with a pendant edge added; we have $v(G) = 5$ and $e(G) = 7$. This graph is not balanced, since the subgraph K_4 has density $6/4 > 7/5$. All other proper induced subgraphs have density at most 1, and thus Theorem 6.1 yields for $\varepsilon = 1$, provided the exponent is at least $C \ln n$,

$$\mathbb{P}(X_G \geq 2\, \mathbb{E}\, X_G) \leq \exp\bigl(-c \min(n^{5/4}p^{7/4}, n^{4/3}p^2)\bigr).$$

In particular, in the range $p \geq n^{-1/3}$, we obtain the same estimate $\exp(-c\mu^{1/(v(G)-1)})$ as for balanced graphs.

Vu [14] also gave a lower bound for the probability in Theorem 6.1. Let $\alpha^*(G)$ denote the fractional independence number of G, i.e. the maximum of $\sum_v a_v$ where a_v, $v \in V(G)$, are non-negative numbers such that $a_v + a_w \leq 1$ whenever v and w are two adjacent vertices. Vu [14] then showed that, provided $\mu \geq 1$ and $p \leq 1/2$, say,

$$\mathbb{P}\bigl(X_G \geq (1+\varepsilon)\, \mathbb{E}\, X_G\bigr) \geq \exp\bigl(-C(\varepsilon, G)(\mathbb{E}\, X_G)^{1/\alpha^*(G)} \ln(1/p)\bigr). \tag{6.2}$$

The proof of this bound is short and elementary. For instance, when $G = K_4$, when $\alpha^*(G) = 2$, consider the event that $\lceil (24(1+\varepsilon)\mu)^{1/4} \rceil + 3$ given vertices form a complete subgraph in $\mathbb{G}(n,p)$.

As observed by Vu [14], this lower bound shows that the estimate in Corollary 6.2 is sometimes sharp up to a logarithmic factor in the exponent; at least, this holds if G is a star. In general, however, the exact asymptotics are unknown, even ignoring such logarithmic factors.

Problem 6.4. *What are the asymptotics of* $-\ln \mathbb{P}\bigl(X_G \geq (1+\varepsilon)\, \mathbb{E}\, X_G\bigr)$*?*

When p is large, the bound of Theorem 6.1 is surpassed by a simple application of Corollary 2.8, which immediately yields the following.

Theorem 6.5. *For any graph G, there exists a constant $c > 0$ such that, for every $\varepsilon > 0$, n and p,*

$$\mathbb{P}\bigl(X_G \geq (1+\varepsilon)\, \mathbb{E}\, X_G\bigr) \leq \exp\bigl(-c\varepsilon \lg(1+\varepsilon) n^2 p^{e(G)}\bigr). \qquad \square$$

It is easily checked that for a constant ε, Corollary 6.2 and Theorem 6.5 both yield estimates $\exp(-c(\varepsilon)n)$ when $p = n^{-1/e(G)}$; for $p \gg n^{-1/e(G)}$ Theorem 6.5 is better than Corollary 6.2, and for $p \ll n^{-1/e(G)}$ Corollary 6.2 is the better.

We consider some examples, for simplicity taking $\varepsilon = 1$. Any constant $\varepsilon \le 1$ would give the same results with, at most, an extra factor ε^2 in the exponent. In particular, we will see how close our results are to the lower bound (6.2).

Example 6.6. A simple example is $G = K_3$. This graph is balanced, and Corollary 6.2 immediately yields, assuming $\mu \asymp n^3p^3 \ge C \ln^2 n$,

$$\mathbb{P}(X_{K_3} \ge 2\,\mathbb{E}\,X_{K_3}) \le \exp\bigl(-c\mu^{1/2}\bigr) \le \exp\bigl(-c' n^{3/2}p^{3/2}\bigr). \tag{6.3}$$

For $p \ge n^{-1/3}$, it is better to use Theorem 6.5, which yields

$$\mathbb{P}(X_{K_3} \ge 2\,\mathbb{E}\,X_{K_3}) \le \exp\bigl(-cn^2p^3\bigr).$$

In this simple case, we can also use Theorem 2.4. Note that $X_{\{i,j\}} \le Z$, where Z is the number of common neighbours of i and j. (More precisely, $X_{\{i,j\}} = Z$ if there is an edge between i and j, and $X_{\{i,j\}} = 0$ otherwise.) Z is a binomial random variable with mean $\nu := (n-2)p^2$, so we can use the Chernoff bound

$$\mathbb{P}(Z \ge z) \le e^{-z}, \qquad z \ge 7\nu \tag{6.4}$$

[3, Corollary 2.4]. Choosing $t = \mu$ and $r = \mu^{1/2}$ in Theorem 2.4, and applying (6.4) with $z = t/6r$, we again obtain (6.3).

We can improve this slightly, by using the sharper Chernoff bound

$$\mathbb{P}(Z \ge z) \le \exp\Bigl(-z \ln \frac{z}{e\nu}\Bigr), \tag{6.5}$$

again see [3, Corollary 2.4]. We choose $t = \mu$, $r = (\mu \ln \mu)^{1/2}$ and $z = t/6r = \frac{1}{6}\mu^{1/2} \ln^{-1/2} \mu$; then $z/e\nu \ge c_1 \mu^{1/6} \ln^{-1/2} \mu$ and thus $\ln(z/e\nu) \ge c_2 \ln \mu$ (for large n). Consequently, Theorem 2.4 and (6.5) yield, still assuming $\mu \asymp n^3p^3 \ge C \ln^2 n$,

$$\mathbb{P}(X_{K_3} \ge 2\,\mathbb{E}\,X_{K_3}) \le \exp\bigl(-c(\mu \ln \mu)^{1/2}\bigr) \le \exp\bigl(-c' n^{3/2}p^{3/2} \ln^{1/2}(np)\bigr). \tag{6.6}$$

On the other hand, we have $\alpha^*(K_3) = 3/2$, and thus (6.2) yields, for $\mu \ge 1$, the lower bound

$$\mathbb{P}\bigl(X_{K_3} \ge 2\,\mathbb{E}\,X_{K_3}\bigr) \ge \exp\bigl(-C\mu^{2/3} \ln(1/p)\bigr) \ge \exp\bigl(-Cn^2p^2 \ln(1/p)\bigr),$$

which for $p \to 0$ is not approached by any of the upper bounds above.

Example 6.7. Another example is $G = K_4$. This graph is balanced too. If we assume $\mu \asymp n^4p^6 \ge C \ln^3 n$, Corollary 6.2 yields

$$\mathbb{P}(X_{K_4} \ge 2\,\mathbb{E}\,X_{K_4}) \le \exp\bigl(-c\mu^{1/3}\bigr) \le \exp\bigl(-c' n^{4/3}p^2\bigr), \tag{6.7}$$

while Theorem 6.5 yields

$$\mathbb{P}(X_{K_4} \ge 2\,\mathbb{E}\,X_{K_4}) \le \exp\bigl(-cn^2p^6\bigr),$$

which is better when $p > n^{-1/6}$.

For some p, we can do substantially better by using Theorem 2.4 and the following argument to estimate the term $\mathbb{P}(X_{\{i,j\}} > t/12r)$.

Fix i and j, and let W be the number of subgraphs of $\mathbb{G}(n,p)$ on 4 vertices, including i and j, that are complete except possibly for the edge ij; such subgraphs are called extensions of type K_4 with roots i and j. Each such extension thus contains, besides i and j, two other vertices that are common neighbours of i and j, and further are joined by an edge. Clearly, $W \geq X_{\{i,j\}}$ ($W = X_{\{i,j\}}$ if i and j are adjacent, and $W = 0$ otherwise).

Expose first all edges in $\mathbb{G}(n,p)$ adjacent to i or j. Let, as in Example 6.6, $Z \sim \mathrm{Bi}(n-2, p^2)$ be the number of common vertices of i and j.

Then expose the remaining vertices. Conditioned on $Z = z$, there are $\binom{z}{2}$ possible edges that would complete an extension of the above type, so $W \sim \mathrm{Bi}\big(\binom{z}{2}, p\big)$.

Now, let $t = \mu$ and fix a large number a such that

$$\begin{aligned} a &\geq 7np^2, \\ \frac{t}{12r} &\geq 7\frac{a^2}{2}p. \end{aligned} \tag{6.8}$$

Then two applications of the Chernoff bound (6.4) and its analogue for W yield

$$\begin{aligned} \mathbb{P}(X_{\{i,j\}} > t/12r) &\leq \mathbb{P}(Z > a) + \mathbb{P}(W > t/12r,\, Z \leq a) \\ &\leq \mathbb{P}\big(\mathrm{Bi}(n-2, p^2) > a\big) + \mathbb{P}\left(\mathrm{Bi}\left(\binom{\lfloor a \rfloor}{2}, p\right) > t/12r\right) \\ &\leq e^{-a} + e^{-t/12r}, \end{aligned}$$

and consequently, by Theorem 2.4,

$$\mathbb{P}(X_{K_4} \geq 2\,\mathbb{E}\,X_{K_4}) \leq e^{-r} + n^2(e^{-a} + e^{-t/12r}). \tag{6.9}$$

We have to choose r and a so that (6.8) holds. If $n^2p^3 \geq C \ln n$ and $np^2 \leq 1$, we take $a = n^2p^3$ and $r = c_1 n^2 p^3$ for some small constant $c_1 > 0$, and obtain

$$\mathbb{P}(X_{K_4} \geq 2\,\mathbb{E}\,X_{K_4}) \leq e^{-cn^2p^3}. \tag{6.10}$$

Actually, (6.10) holds for all $p \leq n^{-1/2}$; the case $n^2p^3 \leq C \ln n$ and $\mu \geq C_1$ follows from Corollary 4.10, cf. Theorem 6.9 below and its proof, and the case $\mu \leq C_1$ is trivial by Markov's inequality; we omit the details.

By Vu's lower bound (6.2), we have (when $\mathbb{E}\,X_{K_4} \geq 1$ and $p \leq 1/2$)

$$\mathbb{P}(X_{K_4} \geq 2\,\mathbb{E}\,X_{K_4}) \geq e^{-c' n^2 p^3 \ln n}. \tag{6.11}$$

Hence, in the case $\mu \geq 1$ and $np^2 \leq 1$, we have found upper and lower bounds (6.10) and (6.11) that differ only by a logarithmic factor in the exponent.

If $p \leq n^{-1/2-\varepsilon}$, for some $\varepsilon > 0$, and $\mu \geq C \ln n$, we can improve the upper bound by taking $r = n^2p^3 \ln^{1/2} n$ instead and using (6.5) and its analogue for W. Since $a/np^2 = np > n^{1/3}$ and $t/12ra^2p > c_2 n^{\varepsilon}$, this yields, instead of (6.9),

$$\mathbb{P}(X_{K_4} \geq 2\,\mathbb{E}\,X_{K_4}) \leq e^{-r} + n^2(e^{-c_3 a \ln n} + e^{-c_4 \varepsilon (\ln n) t/12r})$$

and thus

$$\mathbb{P}(X_{K_4} \geq 2\,\mathbb{E}\,X_{K_4}) \leq e^{-c(\varepsilon) n^2 p^3 \ln^{1/2} n}, \tag{6.12}$$

getting even closer to the lower bound (6.2).

If $np^2 \geq 1$, the above choice of a and r is not allowed. Instead we take $a = r = c_1 n^{4/3} p^{5/3}$ for a small $c_1 > 0$ and find from (6.9)

$$\mathbb{P}(X_{K_4} \geq 2\,\mathbb{E}\,X_{K_4}) \leq e^{-cn^{4/3}p^{5/3}}, \tag{6.13}$$

which improves (6.7). It is still an open problem to find the correct asymptotics in this range of p.

Note that the probabilities of large deviations that we have found for the upper tail when $np^2 \leq 1$ are small but much larger than the corresponding results for the lower tail: for any ε with $0 < \varepsilon \leq 1$, $\mathbb{P}(X_{K_4} \leq (1-\varepsilon)\,\mathbb{E}\,X_{K_4}) \leq e^{-c\varepsilon^2\mu}$, see [3, Theorems 3.9 and 2.14]. This should not be surprising; a comparatively small number of clustered extra edges can create a large number of copies of K_4, see the argument after (6.2), but there is no comparable simple way to get substantially fewer copies than expected.

It is surprising that we are able to find (almost) the precise asymptotics for K_4, at least for some range of p, but not for the simpler case K_3.

The proof of Theorem 6.1 and Corollary 6.2, is, as the corresponding argument in [14], based on an inductive argument, adding one vertex at a time. (To be precise, the induction is for the more general problem of counting extensions, cf. the proof of Theorem 3.11.) The argument above yielding a better bound for K_4 suggests that it may for other graphs too be better to use another induction scheme, adding several vertices (or edges) each time according to some kind of "shell decompostion" of the graph, but we leave this possibility for future research.

Note that applying Corollary 4.1 with Γ the set of edges of K_n as in Example 3.3, i.e. adding one edge at a time ("edge exposure"), generally gives inferior results, even if we may improve them somewhat by using Corollary 4.5 with, for example, $k_0 = k+1-\delta(G)$, where $\delta(G)$ denotes the minimum degree. With this k_0, $M_{k_0} \leq C$ since this many edges determine the vertex set, and M_{k_0} is just the number of ways the remaining edges may be added to create a copy of G.

Small deviations. For small deviations t, i.e. for small ε, the estimate in Theorem 6.1 is useless because of the factor $n^{v(G)-1}$. We give some complementary results for small t; for simplicity we treat only some cases where we obtain bounds of the sub-Gaussian type $\mathbb{P}\bigl(X_G \geq \mu + t\bigr) \leq \exp\bigl(-ct^2/\operatorname{Var} X_G\bigr)$. Note that as soon as $\mathbb{P}(X_G = 0) \to 0$ and $n^2(1-p) \to \infty$, X_G is asymptotically normally distributed [9], [3, Theorem 6.5], and thus an estimate of this type holds if $t/(\operatorname{Var} X_G)^{1/2}$ is fixed, and by a continuity argument if this quantity is slowly increasing too; the problem is to find explicit ranges of t where this is true. Therefore, in this subsection, we write $\sigma^2 := \operatorname{Var} X_G$ and, instead of writing $t = \varepsilon\mu$, we compare t to the standard deviation σ. Theorem 6.8 below slightly extends the upper tail parts of results by Vu [14, Corollary 2], [15, Corollary 6.4].

To avoid trivialities, we assume $e(G) > 0$ and $p \le 1/2$. We begin by observing that then

$$\sigma^2 = \operatorname{Var} X_G \asymp \max_{H\subseteq G,\, v(H)\ge 2} \frac{(\mathbb{E}\, X_G)^2}{\mathbb{E}\, X_H} \asymp \max_{2\le j\le k} \mu\mu_j, \tag{6.14}$$

see [3, Lemma 3.5] and (6.1). We say that H is a *leading overlap* if it attains the maximum in (6.14), i.e. if $\mathbb{E}\, X_H$ is minimal, at least within a constant factor. (Formally this makes sense only if we let $n \to \infty$ and consider some given $p = p(n)$.) See [3, Section 3.2] for further information.

If $e(G) \ge 2$ we define

$$m^{(2)}(G) := \max\Bigl\{\frac{e(H)-1}{v(H)-2} : H \subseteq G \text{ with } v(H) \ge 3\Bigr\};$$

in the trivial case $e(G) = 1$ we set $m^{(2)}(G) := 1/2$. Thus, if $H \subseteq G$ with $v(H) \ge 2$, then

$$\mathbb{E}\, X_H \asymp n^{v(H)} p^{e(H)} \ge n^2 p \bigl(np^{m^{(2)}(G)}\bigr)^{v(H)-2}. \tag{6.15}$$

Note that $m^{(2)}(G) \ge 1/2$ for every G with $e(G) > 0$.

The significance of $m^{(2)}(G)$ comes partly from the fact that if $np^{m^{(2)}(G)} \ge 1$, then (6.14) and (6.15) imply that K_2 is a leading overlap, i.e.

$$\sigma^2 = \operatorname{Var} X_G \asymp \frac{(\mathbb{E}\, X_G)^2}{\mathbb{E}\, X_{K_2}} \asymp \frac{(\mathbb{E}\, X_G)^2}{n^2 p}. \tag{6.16}$$

Theorem 6.8. *For any graph G, there exist constants $c, C > 0$ such that, for every n and $p \le 1/2$ with $np^{m^{(2)}(G)} \ge \ln n$ and for every t with $C \le t/\sigma \le \bigl(np^{m^{(2)}(G)}\bigr)^{1/2}$,*

$$\mathbb{P}\bigl(X_G \ge \mu + t\bigr) \le \exp\bigl(-ct^2/\sigma^2\bigr).$$

Proof. Write $\omega = np^{m^{(2)}(G)}$. Since $m^{(2)}(G) \ge 1/2$, we have $\omega \le \omega^2 \le n^2 p$, and thus, using (6.16),

$$t^2 \le \omega\sigma^2 \le n^2 p \sigma^2 \le C_1 \mu^2. \tag{6.17}$$

By (6.1) and (6.15), for $2 \le j \le k$,

$$\mu_j \le C_2 \frac{\mathbb{E}\, X_G}{n^2 p\, \omega^{j-2}}$$

and thus, by (6.17) and (6.16),

$$\frac{t \lg(1 + t/\mu)}{\mu_j} \ge c_1 \frac{t^2}{\mu\mu_j} \ge c_2 \frac{t^2 n^2 p\, \omega^{j-2}}{(\mathbb{E}\, X_G)^2} \ge c_3 \frac{t^2}{\sigma^2} \omega^{j-2}.$$

By the assumptions, this is at least $c_3 (t^2/\sigma^2)^{j-1}$, and also at least $c_3 (t^2/\sigma^2) \ln^{j-2} n$, and the result follows by Corollary 4.8 with $\ell = 2$. □

Theorem 6.8 requires that p is so large that K_2 is the only leading overlap. Another result by Vu [12, Corollary 5.1] yields a similar bound in the opposite extreme case, viz. when p is small and G itself is the only leading overlap; it is further assumed that G is strictly balanced, i.e. $e(H)/v(H) < e(G)/v(G)$ for every proper subgraph H. We state his result as follows.

Theorem 6.9. *Suppose that G is a strictly balanced graph. Then there exists a constant $c > 0$ such that if $\mu := \mathbb{E}\, X_G \le \ln n$ and $0 < t \le \mu$, then*

$$\mathbb{P}(X_G \ge \mu + t) \le 2e^{-c_1 t^2/\mu} \le 2e^{-c_2 t^2/\sigma^2}.$$

Proof. The assumptions on G and μ imply that for every subgraph $H \subset G$ with $1 \le v(H) < k = v(G)$, and some $\alpha > 0$, we have $\mathbb{E}\, X_H \ge n^\alpha\, \mathbb{E}\, X_G$ (for large n, at least). By (6.1), $\mu_j = O(n^{-\alpha})$ for $1 \le j \le k-1$, while $\mu_k \le C$ for some C, and the first inequality follows by Corollary 4.12 (applied to X_G/C). The second follows because $\sigma^2 \asymp \mu$ by (6.14). (In this case, actually $\sigma^2 \sim \mu$.) □

The lower bound (6.11) shows that the result does not extend to $\mu \gg \ln^2 n$; we do not know whether $\mu = O(\ln n)$ is necessary.

We can obtain a similar estimate also in intermediate cases when some other subgraph is the only leading overlap, or when there are several leading overlaps but all have the same number of vertices.

Theorem 6.10. *Suppose that G is a graph and that $p \le 1/2$ is such that there exists a subgraph $F \subseteq G$ and a number $\gamma > 0$ with $\mathbb{E}\, X_H \ge n^\gamma\, \mathbb{E}\, X_F$ for every subgraph $H \subseteq G$ with $v(H) \ge 2$ and $v(H) \ne v(F)$. Then there is a constant $c = c(G, \gamma)$ such that if $0 < t/\sigma \le (\ln n)^{1/2}$ and $t \le \mu$, then*

$$\mathbb{P}(X_G \ge \mu + t) \le 2e^{-ct^2/\sigma^2}.$$

Proof. Let $f = v(F)$ and $k = v(G)$. By assumption and (6.1),

$$\mu_j \le Cn^{-\gamma}\mu_f, \qquad 2 \le j \le k \text{ and } j \ne f. \tag{6.18}$$

Moreover, by (6.14), $\sigma^2 \asymp \mu\mu_f$.

Let A be a large constant and apply Theorem 3.11 with $r_2 = A^{-k}t/\mu_f$, $r_3, \dots, r_f = A$ and $r_{f+1}, \dots, r_k = An^{\gamma/k}$. Using (6.18), it is easy to first verify (3.8) and then from (3.9) and (6.14) obtain

$$\begin{aligned}\mathbb{P}(X_G \ge \mu + t) &\le e^{-c_1 A^{-k} t^2/\mu\mu_f} + \sum_{j=2}^{f-1} n^j \Big(\frac{A\mu_f}{\mu_j}\Big)^{-cA} + \sum_{j=f}^{k-1} n^j 2^{-cAn^{\gamma/k}} \\ &\le e^{-c_2 A^{-k} t^2/\sigma^2} + n^{f-cA\gamma} + n^k 2^{-cAn^{\gamma/k}}.\end{aligned}$$

The result follows by choosing A large enough (depending on γ and k). □

Problem 6.11. *Can the range of t in Theorem 6.10 be extended?*

If we consider the special case $p = n^{-\alpha}$ for some fixed α, the assumptions of Theorem 6.10 are satisfied (for large n, at least) for all but a finite number of α; the exceptions are when there exist two leading overlaps with different numbers of vertices; cf. [3, Section 3.2]. We obtain the following corollary.

Corollary 6.12. *For any graph G and every $\alpha > 0$ except for a finite number of values (depending on G), there is a constant $c = c(G, \alpha)$ such that if $p = n^{-\alpha}$, $0 < t/\sigma \le (\ln n)^{1/2}$ and $t \le \mu$, then*

$$\mathbb{P}(X_G \ge \mu + t) \le 2e^{-ct^2/\sigma^2}.$$

□

The exceptional case when there are two leading overlaps with different numbers of vertices is more complicated; under suitable hypotheses it is still possible to obtain bounds from (2.4), but they will be weaker than in the theorems above.

Problem 6.13. *Does Corollary 6.12 extend to all α?*

We give here only a simple counterexample which shows that Theorem 6.9 does not extend to arbitrary balanced graphs.

Example 6.14. Let G consist of a triangle with 3 pendant edges attached to the same vertex of the triangle. (Equivalently, G is a star $K_{1,5}$ with an added edge.) Further, let np tend to ∞ very slowly, for example $p = \ln\ln n/n$. For convenience we assume that n is even and large.

We have $\mu := \mathbb{E}\, X_G \sim \frac{1}{12}(np)^6$ and $\sigma^2 := \operatorname{Var} X_G \asymp \frac{\mu^2}{(np)^3} \asymp (np)^9$ (K_3 is a leading overlap). Let $t = (np)^{11/2}$; thus $t/\sigma \asymp np \to \infty$ but $t \ll \mu$.

Expose the edges in two rounds, cf. [3, Section 1.1]: first select red edges with probabilities $p_1 = p(1-(np)^{-3/2})$ and then, independently, blue edges with probabilities $p_2 = (p-p_1)/(1-p_1) > p(np)^{-3/2}$; this makes $p_1 + p_2 - p_1p_2 = p$ and thus we obtain $\mathbb{G}(n,p)$ by ignoring colours and possible double edges.

Let X'_G be the number of red copies of G, and let X''_G be the number of copies with at least one blue edge. We have $\mathbb{E}\, X'_G = (p_1/p)^6\mu = \mu + O(\sigma)$ and $\operatorname{Var} X'_G \sim \operatorname{Var} X_G = \sigma^2$. Hence, by Chebyshev's inequality (or [9]), $\mathbb{P}(X'_G > \mu - t) \to 1$.

Suppose that $X'_G > \mu - t$ and select a vertex v in some red triangle. Let B be the number of blue edges adjacent to v, not counting those overlapping with a red edge; thus $B \sim \operatorname{Bi}(n-1-d_r(v), p_2)$, where $d_r(v)$ is the number of red neighbours of v. If $B \geq b := 3\lceil t^{1/3}\rceil$, then $X''_G \geq \binom{b}{3} > 2t$ and thus $X_G = X'_G + X''_G > \mu + t$. We may further suppose that $d_r(v) < n/2$, since otherwise $X_G \geq X'_G > \binom{n/2-2}{3} \gg \mu + t$. Consequently, conditioned on a red graph with $X'_G > \mu - t$,

$$\begin{aligned}\mathbb{P}(X_G > \mu + t) &\geq \mathbb{P}\Big(\operatorname{Bi}(n/2, p_2) = b\Big) \sim \frac{1}{b!}\Big(\frac{n}{2}p_2\Big)^b \geq \Big(\frac{np_2}{b}\Big)^b \\ &\geq \exp\big(-c_1 t^{1/3}\ln t\big) = \exp\big(-c_2(np)^{11/6}\ln(np)\big).\end{aligned}$$

Hence, unconditionally too, $\mathbb{P}(X_G > \mu + t) \geq \exp\big(-c_2(np)^{11/6}\ln(np)\big)$, which is larger than $e^{-ct^2/\sigma^2} \leq e^{-c'(np)^2}$.

Remark 6.15. The arguments and results in this section apply to counts of *induced* subgraphs too; note that our method does not require the summands Y_I to be increasing functions of the underlying variables ξ_α.

7. Appendix: The lower tail

As remarked in the introduction, estimates for the lower tail can be derived by other inequalities. In particular, we give here a new version of Suen's inequality that applies in the setting of Theorem 2.1. It is a slight extension of [3, Theorem 2.23], where it is assumed that the Y_α are indicator random variables (as is usually the case in applications). (Suen's original inequality [11] deals only with $\mathbb{P}(X = 0)$; see also [2, 10] for further related results.)

Note that the inequality in [3, Theorem 2.14] is slightly stronger when applicable (although the difference usually is insignificant), but it applies only under more restrictive assumptions.

Theorem 7.1. *Suppose that Y_α, $\alpha \in \mathcal{A}$, is a finite family of non-negative random variables and that $\sim$ is a symmetric relation on the index set $\mathcal{A}$ such that if A, B are two subsets of $\mathcal{A}$ such that $\alpha \not\sim \beta$ for $\alpha \in A$, $\beta \in B$, then the family $\{Y_\alpha : \alpha \in A\}$ is independent of $\{Y_\beta : \beta \in B\}$; in other words, the pairs (α, β) with $\alpha \sim \beta$ and $\alpha \neq \beta$ define the edge set of a strong dependency graph for the variables Y_α. Let $X := \sum_\alpha Y_\alpha$ and $\mu := \mathbb{E} X = \sum_\alpha \mathbb{E} Y_\alpha$. Let further, for $\alpha \in \mathcal{A}$, $\tilde{X}_\alpha := \sum_{\beta \sim \alpha} Y_\beta$ and let $\bar{\Delta} := \sum\sum_{\alpha \sim \beta} \mathbb{E}(Y_\alpha Y_\beta) = \sum_\alpha \mathbb{E}(\tilde{X}_\alpha Y_\alpha)$ and $\delta := \max_\alpha \mathbb{E} \tilde{X}_\alpha$. If $0 \le t \le \mu$, then*

$$\mathbb{P}(X \le \mu - t) \le \exp\Bigl(-\min\Bigl(\frac{t^2}{4\bar{\Delta}}, \frac{t}{6\delta}\Bigr)\Bigr). \tag{7.1}$$

Proof. The proof in [3, Theorem 2.23] holds with only notational changes (replace I_i by Y_α and p_i by $\mathbb{E} Y_\alpha$), since, for any $X_A \ge 0$,

$$(Y_\alpha - \mathbb{E} Y_\alpha)\bigl(1 - e^{-sX_A}\bigr) \le Y_\alpha\bigl(1 - e^{-sX_A}\bigr) \le Y_\alpha s X_A.$$

We omit the details. □

In the situation studied in Section 3, we derive the following corollary, which in many cases yields a stronger bound for the lower tail than the upper tail estimate in Corollary 4.1. (For $t = O(\mu)$, the bound is always at least as strong as (4.1) up to a constant factor in the exponent; typically, it is much stronger.)

Corollary 7.2. *Assume* (H1). *With notation as in Section 3, and $c = 2^{-k-2}$, for $0 \le t \le \mu$,*

$$\mathbb{P}(X \le \mu - t) \le \exp\Bigl(-c \min\Bigl(\min_{1 \le j \le k}\Bigl(\frac{t^2}{\mu\mu_j}\Bigr), \frac{t}{\mu_1}\Bigr)\Bigr). \tag{7.2}$$

Proof. In this case, cf. the proof of Theorem 2.3,

$$\delta = \max_I \mathbb{E} \sum_{J \cap I \neq \emptyset} Y_J \le \max_I \mathbb{E} \sum_{i \in I} X_{\{i\}} \le k\mu_1. \tag{7.3}$$

Moreover, given I, let for every non-empty $J \subseteq I$, $X^*_J := \sum_{K \cap I = J} Y_K \le X_J$. Conditioned on $\{\xi_\alpha,\, \alpha \in \mathcal{A}_J\}$, the random variables Y_I and X^*_J are independent, as a consequence of (H1). Hence, using (3.1),

$$\begin{aligned} \mathbb{E}(Y_I X^*_J \mid \xi_\alpha,\, \alpha \in \mathcal{A}_J) &= \mathbb{E}(Y_I \mid \xi_\alpha,\, \alpha \in \mathcal{A}_J)\, \mathbb{E}(X^*_J \mid \xi_\alpha,\, \alpha \in \mathcal{A}_J) \\ &\le \mathbb{E}(Y_I \mid \xi_\alpha,\, \alpha \in \mathcal{A}_J)\mu_J. \end{aligned} \tag{7.4}$$

and consequently $\mathbb{E}(Y_I X^*_J) \le \mathbb{E} Y_I \mu_J$. Hence,

$$\mathbb{E}(Y_I \tilde{X}_I) = \sum_{\emptyset \neq J \subseteq I} \mathbb{E}(Y_I X^*_J) \le \sum_{\emptyset \neq J \subseteq I} \mathbb{E} Y_I \mu_J \le \mathbb{E} Y_I \sum_{j=1}^{k} \binom{k}{j} \mu_j$$

and

$$\bar{\Delta} = \sum_{I} \mathbb{E}(Y_I \tilde{X}_I) \le \mu \sum_{j=1}^{k} \binom{k}{j} \mu_j \le 2^k \mu \max_{1 \le j \le k} \mu_j. \tag{7.5}$$

The result follows form Theorem 7.1, (7.3) and (7.5). □

Remark 7.3. We have throughout assumed that our summands Y_I be non-negative. In cases where the sign of Y_I may change, one may separate the positive and negative parts of Y_I and treat them separately, using a result from Section 2 or 3 for one part and Theorem 7.1 or Corollary 7.2 for the other.

Acknowledgement. This research was partly done during visits of the younger author to Uppsala and of both authors to the Department of Computer Science at Lund University; we thank Andrzej Lingas for providing the latter opportunity. We further thank Van Vu for sending us unpublished preprints and drafts.

References

[1] S. Janson, Poisson approximation for large deviations. *Random Struct. Alg.* **1** (1990), 221–230.
[2] S. Janson, New versions of Suen's correlation inequality. *Random Struct. Alg.* **13** (1998), 467–483.
[3] S. Janson, T. Łuczak & A. Ruciński, *Random Graphs*. Wiley, New York, 2000.
[4] S. Janson & K. Nowicki, The asymptotic distribution of generalized U-statistics with applications to random graphs. *Probab. Th. Rel. Fields* **90** (1991), 341-375.
[5] S. Janson & A. Ruciński, The infamous upper tail. In preparation.
[6] J.H. Kim & V.H. Vu, Concentration of multivariate polynomials and applications. *Combinatorica* **20** (2000), 417–434.
[7] V. Rödl & A. Ruciński, Random graphs with monochromatic triangles in every edge coloring. *Random Struct. Alg.* **5**:2 (1994), 253–270.
[8] V. Rödl & A. Ruciński, Threshold functions for Ramsey properties. *J. Amer. Math. Soc.* **8**:4 (1995), 917–942.
[9] A. Ruciński, When are small subgraphs of a random graph normally distributed? *Probab. Th. Rel. Fields* **78** (1988), 1–10.
[10] J. Spencer, A useful elementary correlation inequality, II. *J. Combin. Th. A* **84** (1998), 95–98.
[11] W.C.S. Suen, A correlation inequality and a Poisson limit theorem for nonoverlapping balanced subgraphs of a random graph. *Random Struct. Alg.* **1** (1990), 231–242.
[12] V.H. Vu, On the concentration of multivariate polynomials with small expectation. *Random Struct. Alg.* **16**:4 (2000), 344-363.
[13] V.H. Vu, New bounds on nearly perfect matchings in hypergraphs: higher codegrees do help. *Random Struct. Alg.* **17**:1 (2000), 29-63.
[14] V.H. Vu, A large deviation result on the number of small subgraphs of a random graph. Technical report MSR-TR-99-90. To appear in *Combin. Probab. Comput.*
[15] V.H. Vu, Concentration of non-Lipschitz functions and applications. Preprint, 2000.

Department of Mathematics, Uppsala University, PO Box 480, S-751 06 Uppsala, Sweden
E-mail address: svante.janson@math.uu.se

Department of Discrete Mathematics, Adam Mickiewicz Universiy, Poznań, Poland
E-mail address: rucinski@amu.edu.pl