COMBINATORICA
Bolyai Society – Springer-Verlag

COMBINATORICA **24** (4) (2004) 615–640

THE DELETION METHOD FOR UPPER TAIL ESTIMATES

SVANTE JANSON, ANDRZEJ RUCIŃSKI*

Received November 22, 2000
Revised August 25, 2003

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. We compare our method with the martingale method by Kim and Vu, which often leads to similar results.

Some applications are given to the number X_G of copies of a graph G in the random graph $\mathbb{G}(n,p)$. In particular, for $G=K_4$ and $G=C_4$ we improve the earlier known upper bounds on $-\ln \mathbb{P}(X_G \geq 2\,\mathbb{E}\,X_G)$ in some range of $p=p(n)$.

1. Introduction

Kim and Vu [8] and Vu [13,15,16] 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 $\mathbb{P}(X \leq \mu - t)$ and $\mathbb{P}(X \geq \mu + t)$, 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.

Mathematics Subject Classification (2000): 60F05

* Research of the second author supported by KBN grants 2 P03A 032 16 and 2 P03A 015 23.

In the present paper, we introduce another method, based on ideas by Rödl and Ruciński [11], to obtain bounds for the upper tail $\mathbb{P}(X \geq \mu + t)$. The new method, which we call the deletion method, see Remark 2.4, looks 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 6.2 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 [15] and the later bound in [5], but we guess that the new bound could be derived using Kim and Vu's method too, cf. [9] for the case of K_3.

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 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 various examples.

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. For these reason (and to give due credit), we specify some connections in detail in Section 5.

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 [12] and Janson [2], or the survey in [3, Chapter 2]. (See also the preprint version of the present paper [6] for 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. This is not necessarily due to a weakness of the method; it seems that in many applications, the lower tail really is much more concentrated than the upper tail, see for example [5]. 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; this is explaind in some detail in Section 5. In Section 6 we discuss some applications to subgraph counts in random graphs; two cases (K_4 and C_4) where we obtain new results are treated in detail. For comparisons with other methods, we refer to [7].

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. 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 $r > 0$,

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

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 term in the estimates decreases with r, while the second term increases. 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 term it is often convenient to use the estimate

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

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

Proof. For $J \subseteq \mathcal{A}$, we write $\alpha \sim J$ if $\alpha \sim \beta$ for some $\beta \in J$.

Let $\mathcal{E}_{r,t}$ be the event that for every set $J \subseteq \mathcal{A}$ with $|J| \leq r$, if we delete all Y_α with $\alpha \sim J$, then the sum of the remaining Y_α is at least $\mu + t$, i.e.

$$|J| \leq r \implies \sum_{\alpha\not\sim J} Y_\alpha \geq \mu + t.$$

We begin with a simple lemma.

Lemma 2.3 (Deletion lemma). *For all real r and $t > 0$, $\mathbb{P}(\mathcal{E}_{r,t}) \leq \big(1 + \frac{t}{\mu}\big)^{-r}$.*

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 $\alpha_i \not\sim \alpha_j$ for $1 \le i < j \le m$, and let $Z_m = \sum^*_{\alpha_1,\dots,\alpha_m} Y_{\alpha_1} \cdots Y_{\alpha_m}$. If $\mathcal{E}_{r,t}$ holds and $m \le r$, then

$$\begin{aligned} Z_{m+1} &= \sum^*_{\alpha_1,\dots,\alpha_m} Y_{\alpha_1} \cdots Y_{\alpha_m} \sum_{\alpha \not\sim \{\alpha_1,\dots,\alpha_m\}} Y_\alpha \\ &\ge \sum^*_{\alpha_1,\dots,\alpha_m} Y_{\alpha_1} \cdots Y_{\alpha_m} (\mu + t) = (\mu + t) Z_m, \end{aligned}$$

so by induction $Z_m \ge (\mu + t)^m$ for $m \le r + 1$.

On the other hand, by assumption, the factors Y_{α_i} in each term in Z_m are independent, and thus

$$\mathbb{E}\, Z_m = \sum^*_{\alpha_1,\dots,\alpha_m} \mathbb{E}\, Y_{\alpha_1} \cdots \mathbb{E}\, Y_{\alpha_m} \le \Big(\sum_\alpha \mathbb{E}\, Y_\alpha \Big)^m = \mu^m. \tag{2.1}$$

Now take $m = \lceil r \rceil$. Then, using Markov's inequality and (2.1),

$$\mathbb{P}(\mathcal{E}_{r,t}) \le \mathbb{P}\big(Z_m \ge (\mu + t)^m\big) \le \frac{\mathbb{E}\, Z_m}{(\mu + t)^m} \le \Big(\frac{\mu}{\mu + t} \Big)^m. \qquad \blacksquare$$

To complete the proof of Theorem 2.1, we note that

$$\sum_{\alpha \not\sim J} Y_\alpha \ge X - \sum_{\beta \in J} \tilde{X}_\beta \ge X - |J| X^*$$

and thus $\{X \ge \mu + t\} \cap \{X^* \le t/2r\} \subseteq \mathcal{E}_{r,t/2}$, so

$$\mathbb{P}(X \ge \mu + t) \le \mathbb{P}(\mathcal{E}_{r,t/2}) + \mathbb{P}(X^* > t/2r).$$

This and Lemma 2.3 show 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 $x > 0$. $\blacksquare$

Remark 2.4. The use of the event $\mathcal{E}_{r,t}$ above is the reason that we call our approach "the deletion method". For earlier versions, see [11] and [3, Lemma 2.51].

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

Theorem 2.5. *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* $r > 0$,

$$\begin{aligned}\mathbb{P}(X \ge \mu + t) &\le \Bigl(1 + \frac{t}{2\mu}\Bigr)^{-r} + \mathbb{P}\Bigl(X_1^* > \frac{t}{2kr}\Bigr)\\ &\le \Bigl(1 + \frac{t}{\mu}\Bigr)^{-r/2} + \sum_{i \in \Gamma} \mathbb{P}\Bigl(X_{\{i\}} > \frac{t}{2kr}\Bigr).\end{aligned}$$

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

$$\tilde{X}_I = \sum_{J \cap I \ne \emptyset} Y_J \le \sum_{i \in I} X_{\{i\}} \le k X_1^*. \qquad ∎$$

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 2.5, which usually gives stronger bounds.

Theorem 2.6. *Suppose that* $\mathcal{H} \subseteq [\Gamma]^{\le k}$ *for an integer* $k \ge 2$, *and that* Y_I, $I \in \mathcal{H}$, *is a family of non-negative random variables such that each* Y_I *is independent of* $\{Y_J : |J \cap I| \le 1\}$. *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*

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

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

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

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

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

Remark 2.7. For random graphs, another possibility leading to the same bounds is to use Theorem 2.5 with Γ being the set of edges of the complete graph; nevertheless, Theorem 2.6 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 2.5 and 2.6. 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 in Example 6.2 below where we use a Chernoff bound for sums of independent variables twice.

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

$$\mathbb{P}(X \geq \mu + t) \leq \left(1 + \frac{t}{\mu}\right)^{-t/(4M\Delta)}.$$

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

Corollary 2.9. *Let the assumptions of Theorem 2.5 hold. Suppose further that M is a number such that $0 \leq Y_I \leq M$ for each I, and let $N := |\Gamma|$ and $\Delta_1 := \max_{i \in \Gamma} |\{J \in \mathcal{H} : i \in J\}|$. Then*

$$\mathbb{P}(X \geq \mu + t) \leq \left(1 + \frac{t}{\mu}\right)^{-t/(4kM\Delta_1)} \leq \left(1 + \frac{t}{\mu}\right)^{-t/(4kMN^{k-1})}.$$

Proof. Take $r = t/(2kM\Delta_1)$ in Theorem 2.5 and observe that $\Delta_1 \leq N^{k-1}$. $\blacksquare$

Corollary 2.10. *Let the assumptions of Theorem 2.6 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*

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

Proof. Take $r = t/(k(k-1)M\Delta_2)$ in Theorem 2.6 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 [10]. See also [7].

Note further that for the case of independent summands ($\Delta = 1$ in Corollary 2.8, $k = 1$ in Corollary 2.9 or $k = 2$ in Corollary 2.10), 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 2.11. Sometimes, for example when studying random hypergraphs, even stronger independence properties than in Theorem 2.6 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 2.6 to the reader.

3. Induction

In many cases, Theorem 2.5 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, $I \in \mathcal{H} \subseteq [\Gamma]^{\le k}$ for some finite index set Γ and an integer $k \ge 1$, 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 $\mathcal{A}_I \subseteq \mathcal{A}$, $I \in [\Gamma]^{\le k}$, 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 $I, J \in [\Gamma]^{\le k}$.

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 $I \in [\Gamma]^{\le k}$. Actually, we can without loss of generality assume that Y_I is defined for all $I \in [\Gamma]^{\le k}$ 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 2.5 hold under (H1). The situation studied here is more special than in Theorem 2.5, 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 $\mathcal{A}_I = [I]^{\leq 2} \cap \mathcal{A}$, 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 $\mathcal{A} \subseteq \bigcup_{j=1}^{\ell}[\Gamma]^j = [\Gamma]^{\leq \ell} \setminus \{\emptyset\}$, for some ℓ, and $\mathcal{A}_I = [I]^{\leq \ell} \cap \mathcal{A}$. 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 2.6.

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 $l \leq k$,

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

In other words, μ_l is the smallest number such that $\mathbb{E}(X_I \mid \xi_\alpha, \alpha \in \mathcal{A}_I) \leq \mu_l$ 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* $r_1, \ldots, r_k$ *such that*

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

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 2.5 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 2.9 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

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

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.

3.1. Stronger independence

In the case of random graphs treated as in Example 3.5, we have the stronger independence property of Theorem 2.6, 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*

$$
r_2 r_3\cdots r_j\cdot \mu_j \le t, \qquad j=2,\dots,k,
\tag{3.8}
$$

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

$$
\mathbb{P}(X\ge \mu+t) \le \Big(1+\frac{t}{\mu}\Big)^{-cr_2} + \sum_{j=2}^{k-1} N^j \Big(1+\frac{t}{r_2r_3\cdots r_j\,\mu_j}\Big)^{-cr_{j+1}}.
\tag{3.9}
$$

Proof. We apply Theorem 2.6 with $r = r_2/2k^2$ and obtain, letting $t_1 = t/r_2 = t/2k^2 r$,

$$\mathbb{P}(X \geq \mu + t) \leq \Big(1 + \frac{t}{\mu}\Big)^{-r/2} + \sum_{\{i,j\} \in [\varGamma]^2} \mathbb{P}\Big(X_{\{i,j\}} > 2t_1\Big). \tag{3.10}$$

Each term in the sum is estimated as in the proof of Theorem 3.10; this time we fix two indices $i, j \in \varGamma$, let $\widetilde{\varGamma} = \varGamma \setminus \{i,j\}$ and have $X_{\{i,j\}} = \sum_{I \in \widetilde{\mathcal{H}}} \widetilde{Y}_I$ with $\widetilde{Y}_I = Y_{I \cup \{i,j\}}$ and $\widetilde{\mathcal{H}} = \{I \subseteq \widetilde{\varGamma} : I \cup \{i,j\} \in \mathcal{H}\} \subseteq [\widetilde{\varGamma}]^{\leq k-2}$. 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 $\widetilde{\mu}_I \leq \mu_{I \cup \{i,j\}}$ and $\widetilde{\mu}_l \leq \mu_{l+2}$. Moreover, by (3.1), (3.2) and (3.8),

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

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,

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

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 2.6 and Theorem 3.10, i.e. we combine one application of Theorem 2.6 and repeated applications of Theorem 2.5. This is thus a kind of combination of edge exposure and vertex exposure.

As remarked in Remark 2.11, 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 $2 \leq \ell \leq k$.)

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

We then have the following generalization of Theorem 3.11.

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

$$r_\ell \cdots r_j \cdot \mu_j \leq t, \qquad j = \ell, \dots, k, \tag{3.12}$$

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

$$\mathbb{P}(X \geq \mu + t) \leq \Big(1 + \frac{t}{\mu}\Big)^{-cr_\ell} + \sum_{j=\ell}^{k-1} N^j \Big(1 + \frac{t}{r_\ell \cdots r_j\, \mu_j}\Big)^{-cr_{j+1}}. \tag{3.13}$$

Proof. We apply Theorem 2.1 with $I \sim J$ when $|I \cap J| \geq \ell$, estimate $\tilde{X}_I \leq \sum_{J \in [I]^\ell} X_J$ 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. ∎

3.2. Further refinements

We define, for $1 \leq j \leq k$,

$$M_j := \max_{|J|=j} \sup X_J. \tag{3.14}$$

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 $1 \leq k_0 \leq k$. Then, with notation as above, for every $t > 0$ and $r_1, \dots, r_{k_0}$ such that*

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

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

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

Proof. If $k_0 = 1$, we have by (3.15), for every $i \in \Gamma$, $X_{\{i\}} \leq M_1 \leq t/r_1$, and the result follows by taking $r = r_1/2k$ in Theorem 2.5.

If $k_0 \geq 2$ 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* $\ell \geq 2$. *Then, with notation as above, for every* $t>0$, $\ell \leq k_0 \leq k$ *and* $r_\ell, \dots, r_{k_0}$ *such that*

$$r_\ell \cdots r_j \cdot \mu_j \leq t, \qquad j=\ell,\dots,k_0-1,$$
$$r_\ell \cdots r_{k_0} \cdot M_{k_0} \leq t,$$

we have for some $c>0$,

$$\mathbb{P}(X \geq \mu + t) \leq \Big(1+\frac{t}{\mu}\Big)^{-cr_\ell} + \sum_{j=\ell}^{k_0-1} N^j \Big(1+\frac{t}{r_\ell \cdots r_j\,\mu_j}\Big)^{-cr_{j+1}}. \qquad \blacksquare$$

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

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

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: *In Theorem 3.10 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* $t>0$,

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

(4.1)
$$\leq \begin{cases} 2N^{k-1}\exp\Big(-c\min_{1\leq j\leq k}\Big(\frac{t^2}{\mu\mu_j}\Big)^{1/j}\Big), & t\leq\mu;\\ 2N^{k-1}\exp\Big(-c\min_{1\leq j\leq k}\Big(\frac{t}{\mu_j}\Big)^{1/j}\Big), & t\geq\mu.\end{cases}$$

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 $c_1=1/8k$

$$\mathbb{P}(X\geq\mu+t)\leq 2^{-c_1r_1\tau}+\sum_{j=2}^{k}N^{j-1}2^{-c_1r_j}. \tag{4.3}$$

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.

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

This makes all exponents of 2 in (4.3) equal to $-c_1r$, and the right hand side of (4.2) can be bounded by

$$2^{-c_1r}\sum_{j=1}^{k}N^{j-1}<e^{-(c_1\ln 2)r}(2N^{k-1}).$$

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. ∎

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 $t>0$,*

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

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

$$r = \min_{2 \le j \le k} \Big(\frac{t\tau}{\mu_j}\Big)^{1/(j-1)}. \qquad \blacksquare$$

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 $t \ge \mu$.)

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

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

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 $r_1, \dots, r_m$ roughly equal to λ, at least, and for the final step we need something like $t/(r_1 \cdots r_m) \ge 1$; hence, again for $t = \mu$, typically $\lambda^m \le \mu$. 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* $t > 0$,

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

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* $\ell \le k_0 \le k$*,*

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

∎

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* $t>0$*,*

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

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

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

(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* $t>0$*,*

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

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 $t > 0$,*

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

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

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

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*

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

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*

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

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 [8] and Vu [13,15,16], where similar induction arguments are used. One difference, which does not matter for many applications, is that Kim and Vu study sums of variables of a special structure, while we focus on the independence properties of the summands.

The general setting of Kim and Vu in these papers 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. $\xi_i \in \{0,1\}$. 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.)

It is easily verified that a suitable choice of r_i in Theorem 3.10 yields the upper tail part of the main theorem in Kim and Vu [8], see also [16, Theorem 3.1], apart from the numerical value of the constants. (Corollary 4.1 yields a somewhat better estimate.) See [6] for detailed proofs of this and other claims in this section.

This result by Kim and Vu is superseded by later results by Vu [14,16]. Indeed, Vu [16] inspired us to both Theorem 3.13 and Corollaries 4.7–4.9. It can be checked that Corollary 4.7 yields the upper tail parts of Theorem 2.3 in [14] and, together with Corollary 4.9 (with $\ell = 1$), Theorem 3.2 in [16].

These results all use, in our version, $\mathcal{A} = \Gamma$ as in Example 3.1. The more general setting in (H1) is inspired by Vu [15], which studies the subgraph count X_G in $\mathbb{G}(n,p)$, where G is a fixed graph with k vertices (see Section 6 below) and the corresponding, more general, problem of counting extensions of a given type with a fixed set of roots. 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 [15]. (The upper tail parts of these theorems by Vu follow from our Corollaries 4.1 and 4.3; similarly, the upper tail part of Theorem 6 in [15] follows from our Corollary 4.4.) 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.

Corollaries 4.10–4.12 are inspired by and strongly related to results in Vu [13]; we have not been able to derive Theorem 1.3 in [13] 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 [13].

6. Applications to random graphs

We give in this section an application of the general results developed in this paper to subgraph counts of random graphs.

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 Sections 2 and 3. We have (for $n\geq k$)

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

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, i.e. it has the same meaning as the $\Theta()$ notation, but reflects better its symmetric nature.

In 1990, Janson [1] proved that the lower tail of the distribution of X_G decays exponentially in the expectation of the least expected subgraph of G. Namely, let $\Psi_H := n^{v_H} p^{e_H}$, which is roughly the expected number of copies of H in $G(n,p)$. Then, for every $\varepsilon>0$,

$$\mathbb{P}(X_G \leq (1-\varepsilon)\,\mathbb{E}\, X_G) \leq \exp\Big(-\Theta_\varepsilon\big(\min_{H\subseteq G} \Psi_H\big)\Big).$$

This is best possible, as by the FKG inequality, $\log \mathbb{P}(X_G=0)$ is of the same order.

Assume in the sequel, for simplicity, that $p \geq n^{-1/m(G)}$, the threshold for $X_G > 0$, where $m(G) = \max_{H\subseteq G} e(H)/v(H)$. Recently, it was proved in [5] that, for every $\varepsilon > 0$,

$$p^{\Theta_\varepsilon(M_G^*)} \leq \mathbb{P}(X_G \geq (1+\varepsilon)\,\mathbb{E}\,X_G) \leq \exp\left\{-\Theta_\varepsilon(M_G^*)\right\}, \tag{6.1}$$

where,

$$M_G^* = \begin{cases} \Theta(\min_{H\subseteq G} \Psi_H^{1/\alpha_H^*}) & \text{if } p \leq n^{-1/\Delta_G}, \\ \Theta(n^2 p^{\Delta_G}) & \text{if } p \geq n^{-1/\Delta_G}, \end{cases}$$

α_H^* is the *fractional independence number* of H, that is the largest value of $\sum_v \alpha_v$ over all assignments of nonnegative weights $\alpha_v \in [0,1]$ satisfying the condition $\alpha_v + \alpha_u \leq 1$ for all edges uv of H, while Δ_G is the maximum degree.

Note that the upper tail decays to zero much slower than the lower tail. Also, unlike the lower tail, the exponents of the above estimates for the upper tail of X_G are a logarithmic factor apart. Hence the following problem remains (narrowly) open.

Problem 6.1. What are the asymptotics of $-\ln \mathbb{P}(X_G \geq (1+\varepsilon)\,\mathbb{E}\,X_G)$?

We will now see that our deletion method yields in general estimates on the upper tail of X_G which are weaker than those established in [5]. However, in two particular cases ($G = K_4$ and $G = C_4$), with an additional argument, we are able to improve the above upper bound on the upper tail of X_G.

For simplicity we take $\varepsilon = 1$. Any constant $\varepsilon \leq 1$ would give the same results with, at most, an extra factor ε^2 in the exponent. Provided that for each $H \subseteq G$ with $v(H) \geq 2$, the expectation $\mathbb{E}\,X_H$ is at least some large constant times $\ln^{v(H)-1} n$, Corollary 4.3 yields that

$$\mathbb{P}(X_G \geq 2\,\mathbb{E}\,X_G) \leq \exp\Big(-\Theta\Big(\min_{H\subseteq G: v(H)\geq 2} (\mathbb{E}\,X_H)^{1/(v(H)-1)}\Big)\Big). \tag{6.2}$$

Recall that the graph G is said to be *balanced* if $e(H)/v(H) \leq e(G)/v(G)$ for every $H \subseteq G$, see [3]. For balanced graphs, estimate (6.2) assumes quite a simple form: if $\mathbb{E}\,X_G \geq C \ln^{v(G)-1} n$, then

$$\mathbb{P}(X_G \geq 2\,\mathbb{E}\,X_G) \leq \exp(-\Theta((\mathbb{E}\,X_G)^{1/(v(G)-1)})). \tag{6.3}$$

which is essentially the same bound as in Vu [15, Theorem 3].

When p is large, the bound in (6.2) is surpassed by a simple application of Corollary 2.10, which immediately yields that

$$\mathbb{P}(X_G \geq 2\,\mathbb{E}\,X_G) \leq \exp(-\Theta(n^2 p^{e(G)})). \tag{6.4}$$

This, however, except for stars, is still weaker than the bound in (6.1).

Before turning to our next example, let us notice that for k-regular graphs G we have $M_G^* = \Theta(n^2 p^k)$ (see [5]).

Example 6.2. Let us now consider the case when $G = K_4$. This graph is balanced, and assuming $\mu \asymp n^4 p^6 \geq C \ln^3 n$, (6.3) yields

$$\mathbb{P}(X_{K_4} \geq 2\,\mathbb{E}\,X_{K_4}) \leq \exp(-c\mu^{1/3}) \leq \exp(-c' n^{4/3} p^2),$$

while (6.4) yields

$$\mathbb{P}(X_{K_4} \geq 2\,\mathbb{E}\,X_{K_4}) \leq \exp(-cn^2 p^6),$$

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

For some p, we can do substantially better by using Theorem 2.6 directly and the following argument to estimate the term $\mathbb{P}(X_{\{i,j\}} > t/12r)$, where recall $X_{\{i,j\}}$ is, in this case, the number of copies of K_4 containing the edge ij.

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. Each such subgraph 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 $X_{\{i,j\}} = 0$ otherwise).

Expose first all edges in $\mathbb{G}(n,p)$ adjacent to i or j. Let $Z \sim \mathrm{Bi}(n-2, p^2)$ be the number of common neighbors of i and j. Then expose the remaining edges. Conditioned on $Z = z$, there are $\binom{z}{2}$ possible edges that would complete a subgraph counted by W, so $W \sim \mathrm{Bi}(\binom{z}{2}, p)$. Hence, for any $t, r, a > 0$ we have

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

Assume that $p \leq n^{-1/2-\gamma}$, for some $\gamma > 0$, and $\mu \geq C \ln n$, for some large $C > 0$, and choose $t = \mu$, $a = n^2 p^3$ and $r = n^2 p^3 \ln^{1/2} n$. We will apply the Chernoff bound in the form

$$\mathbb{P}(\mathrm{Bi}(n,p) \geq x) \leq \exp\Big(-x \ln \frac{x}{enp}\Big), \tag{6.5}$$

see [3, Corollary 2.4], to both Z and W.

Since $a/np^2 = np > n^{1/3}$ and $t/12ra^2 p > c_2 n^\gamma$, this yields, by Theorem 2.6,

$$\mathbb{P}(X_{K_4} \geq 2\,\mathbb{E}\,X_{K_4}) \leq e^{-r/3} + n^2 (e^{-c_3 a \ln n} + e^{-c_4 \gamma (\ln n) t/12r}) \tag{6.6}$$

and thus

$$\mathbb{P}(X_{K_4} \geq 2\,\mathbb{E}\,X_{K_4}) \leq e^{-c(\gamma)n^2p^3\ln^{1/2}n}, \tag{6.7}$$

getting away from the upper bound in (6.1).

Example 6.3. The case $G=C_4$ was treated in detail in [7] where for $p\leq n^{-2/3}$ an upper bound equivalent to that in (6.1) was established. Here we show how to improve it similarly to Example 6.2. In this case we have $\mu\asymp n^4p^4$ and $M_G^*\asymp n^2p^2$. Let us assume that $C\ln^{1/4}n/n\leq p\leq n^{-2/3-\gamma}$ for some $\gamma>0$. For fixed i and j, $X_{\{i,j\}}$ counts copies of C_4 containing the edge ij. Let W be the number of paths of length 3 connecting i and j, Z – the number of neighbors of i or j different from i and j (set N), and U – the number of edges with both endpoints in N. Then

$$X_{\{i,j\}} \leq W \leq 2U,$$

where $Z\sim\mathrm{Bi}(n-2,2p-p^2)$ and, given $Z=z$, $U\sim\mathrm{Bi}\Big(\binom{\lfloor z\rfloor}{2},p\Big)$. Hence, for any $t,r,a>0$ we have

$$\mathbb{P}(X_{\{i,j\}} > t/12r) \leq \mathbb{P}(\mathrm{Bi}(n-2,2p) > a) + \mathbb{P}\Big(\mathrm{Bi}\Big(\binom{\lfloor a\rfloor}{2},p\Big) > t/24r\Big).$$

Setting $t=\mu$ and $a=r=n^2p^2\ln^{1/2}n$, and applying Theorem 2.6 and twice the Chernoff bound (6.5) as in Example 6.2, we arrive at

$$\mathbb{P}(X_{C_4} \geq 2\,\mathbb{E}\,X_{C_4}) \leq e^{-c(\gamma)n^2p^2\ln^{1/2}n}. \tag{6.8}$$

(Here we need the extra factor $\ln^{1/2}n$ in a, since for p close to $1/n$ the term $\ln(a/\mathbb{E}\,Z)$ is of order smaller than $\ln^{1/2}n$.)

Remark 6.4. At the moment we are unable to obtain such a good upper bound as (6.7) in other ranges of p and for other graphs G. The complete graph K_4 and the 4-cycle C_4 both seem to be exceptionally suited for our method because their vertex set can be broken into two pairs, allowing us to use the natural independence of the edges of $\mathbb{G}(n,p)$ twice, together with the independence of the common neighbors. Moreover, luckily, for these two graphs $\mathbb{E}\,X_G$ is roughly the square of M_G^*. As for the range, we need p small in order to gain the $\ln n$ term in the last exponent of (6.6) and, respectively, in its analog for C_4.

Remark 6.5. 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.

Remark 6.6. For detailed proofs of the estimates (6.2) and (6.3), as well as for more examples of the application of the deletion method to other small subgraphs, see the preprint version of the present paper [6].

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(2)** (1990), 221–230.

[2] S. Janson: New versions of Suen's correlation inequality, *Random Struct. Alg.* **13(3–4)** (1998), 467–483.

[3] S. Janson, T. Łuczak and A. Ruciński: *Random Graphs*, Wiley, New York, 2000.

[4] S. Janson and 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, K. Oleszkiewicz and A. Ruciński: Upper tails for subgraph counts in random graphs, to appear in *Israel J. Math.* Available from http://www.math.uu.se/~svante/.

[6] S. Janson and A. Ruciński: The deletion method for upper tail estimates, Preprint, 2000. Available from http://www.math.uu.se/~svante/.

[7] S. Janson and A. Ruciński: The infamous upper tail, *Random Struct. Alg.* **20(3)** (2002), 317–342.

[8] J. H. Kim and V. H. Vu: Concentration of multivariate polynomials and applications *Combinatorica* **20(3)** (2000), 417–434.

[9] J. H. Kim and V. H. Vu: Divide and conquer martingales and the number of triangles in a random graph, *Random Struct. Alg.* **24(2)** (2004), 166–174.

[10] V. Rödl and A. Ruciński: Random graphs with monochromatic triangles in every edge coloring, *Random Struct. Alg.* **5(2)** (1994), 253–270.

[11] V. Rödl and A. Ruciński: Threshold functions for Ramsey properties, *J. Amer. Math. Soc.* **8(4)** (1995), 917–942.

[12] W. C. S. Suen: A correlation inequality and a Poisson limit theorem for nonoverlapping balanced subgraphs of a random graph, *Random Struct. Alg.* **1(2)** (1990), 231–242.

[13] V. H. Vu: On the concentration of multivariate polynomials with small expectation, *Random Struct. Alg.* **16(4)** (2000), 344–363.

[14] V. H. Vu: New bounds on nearly perfect matchings in hypergraphs: higher codegrees do help; *Random Struct. Alg.* **17(1)** (2000), 29–63.

[15] V. H. Vu: A large deviation result on the number of small subgraphs of a random graph, *Combin. Probab. Comput.* **10(1)** (2001), 79–94.

[16] V. H. Vu: Concentration of non-Lipschitz functions and applications, *Random Struct. Alg.* **20(3)** (2002), 262–316.

Svante Janson

Department of Mathematics
Uppsala University
PO Box 480
S-751 06 Uppsala
Sweden
svante.janson@math.uu.se
http://www.math.uu.se/~svante/

Andrzej Ruciński

Department of Discrete Mathematics
Adam Mickiewicz Universiy
Poznań
Poland
rucinski@amu.edu.pl
http://main.amu.edu.pl/~rucinski/