The Infamous Upper Tail *

Svante Janson
Uppsala University
Sweden

Andrzej Ruciński†
A. Mickiewicz University
Poznań, Poland

May 28, 2001; revised November 19, 2001

Abstract

Let Γ be a finite index set and $k \geq 1$ a given integer. Let further $\mathcal{S} \subseteq [\Gamma]^{\leq k}$ be an arbitrary family of k element subsets of Γ. Consider a (binomial) random subset $\Gamma_{\mathbf{p}}$ of Γ, where $\mathbf{p} = (p_i : i \in \Gamma)$ and a random variable X counting the elements of $\mathcal{S}$ that are contained in this random subset.

In this paper we survey techniques of obtaining upper bounds on the upper tail probabilities $\mathbb{P}(X \geq \lambda + t)$ for $t > 0$. Seven techniques, ranging from Azuma's inequality to the purely combinatorial deletion method, are described, illustrated and compared against each other for a couple of typical applications.

As one application, we obtain essentially optimal bounds for the upper tails for the numbers of subgraphs isomorphic to K_4 or C_4 in a random graph $G(n,p)$, for certain ranges of p.

1 Introduction

Let Γ be a finite index set and $k \geq 1$ a given number, and let $[\Gamma]^{\leq k}$ be the family of all subsets $A \subseteq \Gamma$ with $|A| \leq k$. Suppose that I_A, $A \in [\Gamma]^{\leq k}$, is a family of non-negative random variables such that each I_A is independent of $\{I_B : B \cap A = \emptyset\}$. Let $X := \sum_A I_A$ and $\lambda := \mathbb{E}\, X = \sum_A \mathbb{E}\, I_A$.

In this paper we survey techniques of obtaining upper bounds on the upper tail probabilities $\mathbb{P}(X \geq \lambda + t)$ for $t > 0$. In many instances such inequalities come together with their lower tail counterparts as a two-sided concentration result. This is the case of celebrated Azuma's and Talagrand's inequalities (see next section). The simplest situation takes place when $k = 1$. Then X is a sum of independent summands, and if these happen to be 0–1 random variables, Chernoff's

†Supported by KBN grant 2 P03A 032 16

bounds apply. With $\phi(x) = (1+x)\log(1+x) - x$, $x \geq -1$ (and $\phi(x) = \infty$ for $x < -1$), we have, see e.g. [1], [7] and [9, Theorem 2.8]

$$\mathbb{P}(X \geq \lambda + t) \leq \exp\Bigl(-\lambda\phi\Bigl(\frac{t}{\lambda}\Bigr)\Bigr) \leq \exp\Bigl(-\frac{t^2}{2(\lambda + t/3)}\Bigr), \qquad t \geq 0; \tag{1}$$

$$\mathbb{P}(X \leq \lambda - t) \leq \exp\Bigl(-\lambda\phi\Bigl(\frac{-t}{\lambda}\Bigr)\Bigr) \leq \exp\Bigl(-\frac{t^2}{2\lambda}\Bigr), \qquad t \geq 0. \tag{2}$$

There is a little asymmetry between the lower and upper tail, but for $t \leq \lambda$ the order of magnitude of the exponents is the same.

For arbitrary k, a special case of our general framework can be described as follows. Suppose that ξ_i, $i \in \Gamma$, is a family of independent 0–1 random variables, and that $I_A = \prod_{i \in A} \xi_i$ when $A \in \mathcal{S}$ for a given family $\mathcal{S} \subseteq [\Gamma]^{\leq k}$, while $I_A = 0$ when $A \notin \mathcal{S}$. In other words, the indicator random variables ξ_i describe a (binomial) random subset $\Gamma_{\mathbf{p}}$ of Γ, where $\mathbf{p} = (p_i : i \in \Gamma)$, $p_i = \mathbb{P}(\xi_i = 1)$ [we write Γ_p if $p_i = p$ for all i], and X is the number of elements of $\mathcal{S}$ that are contained in this random subset. For the lower tail of the distribution of X, the following analogue of the Chernoff bound holds [8], [9, Theorem 2.14].

Theorem 0. Let $X = \sum_{A \in \mathcal{S}} I_A$ as above, and let $\lambda = \mathbb{E}\, X = \sum_A \mathbb{E}\, I_A$ and $\bar{\Delta} = \sum\sum_{A \cap B \neq \emptyset} \mathbb{E}(I_A I_B)$. Then, with $\phi(x) = (1+x)\log(1+x) - x$, for $0 \leq t \leq \lambda$,

$$\mathbb{P}(X \leq \lambda - t) \leq \exp\left(-\frac{\phi(-t/\lambda)\lambda^2}{\bar{\Delta}}\right) \leq \exp\left(-\frac{t^2}{2\bar{\Delta}}\right).$$

It follows from the FKG inequality (see for example [9]) that the above bound is tight: if $t = \lambda$, $\bar{\Delta} - \lambda = o(\lambda)$ and $\max p_i = o(1)$, then

$$\mathbb{P}(X = 0) = \exp\{-\lambda(1 + o(1))\}.$$

However, an upper tail analogue cannot be true in general. One slightly artificial counterexample is presented in [9, Remark 2.17]. A more natural one is the following from [20] (here slightly adapted).

(Counter)Example. Let $\Gamma = [n]^2$, $k = 3$, and $\mathcal{S}$ be the family of the edge sets of all triangles in K_n, the complete graph on $[n]$. Furthermore, let $p_i = p = p(n)$ for all $i \in [n]^2$. Then X is simply the number of triangles in the random graph $G(n,p)$. Assuming $\frac{\log n}{n} \ll p = o(1)$, fix three disjoint subsets of vertices, each of order $v = (2\lambda)^{1/3}$, where recall $\lambda = \mathbb{E}\, X$. Then with probability p^{3v^2} there is a complete tripartite subgraph on the three sets, yielding $v^3 = 2\lambda$ triangles in $G(n,p)$, and thus

$$-\log \mathbb{P}(X \geq 2\lambda) \leq 3v^2 \log \frac{1}{p} \leq 3(2\lambda)^{2/3} \log n \ll \lambda$$

so $\mathbb{P}(X \geq 2\lambda) \gg e^{-c\lambda}$ for every $c > 0$.

In the next section we present several techniques of obtaining exponential bounds on the upper tail of X. Then, in the last section we illustrate them by a few examples and based on these examples, we compare them against each other. For X being the number of copies of K_4 in a random graph $G(n,p)$, one of our methods yields in a range of p an optimal bound, that means a bound which up to a logarithmic factor in the exponent matches the lower bound obtained by Vu's (counter)example. For the number of copies of C_4, two of the methods yield an optimal bound for certain ranges of p.

We use c_k to denote various positive constants depending on k only.

This paper is meant to be an extension of Section 2.6 of [9]. We hope that the examples treated here can serve as inspiration and suggestions for future applications of the methods.

2 Methods

2.1 Inequalities based on Lipschitz condition

The first method is a version of Azuma's inequality [7] tailored for combinatorial applications, see e.g. [12], [13] and [9, Remark 2.28].

Theorem 1. *Let $Z_1, \ldots, Z_M$ be independent random variables, with Z_j taking values in a set Λ_j. Assume that a function $f : \Lambda_1 \times \Lambda_2 \times \cdots \times \Lambda_M \to \mathbb{R}$ satisfies, for some constants b_j, $j = 1, \ldots, M$, the following Lipschitz condition:*

(L) *If two vectors $\mathbf{z}, \mathbf{z}' \in \Lambda_1 \times \Lambda_2 \times \cdots \times \Lambda_M$ differ only in the jth coordinate, then $|f(\mathbf{z}) - f(\mathbf{z}')| \leq b_j$.*

Then, the random variable $X = f(Z_1, \ldots, Z_M)$ satisfies, for any $t \geq 0$,

$$\mathbb{P}(X \geq \lambda + t) \leq \exp\Big\{-2t^2 \Big/ \sum_1^M b_j^2\Big\}, \tag{3}$$

$$\mathbb{P}(X \leq \lambda - t) \leq \exp\Big\{-2t^2 \Big/ \sum_1^M b_j^2\Big\}. \tag{4}$$

Returning to the random set Γ_p, one typically defines the random variables Z_j via the random indicators ξ_i, $i \in \Gamma$. Given a partition $A_1, \ldots, A_M$ of Γ, each Z_j is then taken as the random vector $(\xi_i : i \in A_j) \in \{0,1\}^{A_j}$, and for a given function $f : 2^\Gamma \to \mathbb{R}$, the Lipschitz condition (L) in Theorem 1 is equivalent to saying that for any two subsets $A, B \subseteq \Gamma$, $|f(A) - f(B)| \leq b_j$ whenever the symmetric difference of the sets A and B is contained in A_j.

When $\Gamma = [n]^2$ and so $\Gamma_p = G(n,p)$, there are two common choices of the partition $[n]^2 = A_1 \cup \cdots \cup A_M$. The vertex exposure martingale corresponds to the choice $M = n$ and $A_j = [j]^2 \setminus [j-1]^2$. The edge exposure martingale is one in which $M = \binom{n}{2}$ and $|A_j| = 1$ for each j. E.g., with $b_j = 1$, edge exposure is applicable provided the random variable X changes by at most 1 if a single edge is added or deleted, while vertex exposure is applicable provided X changes by at most 1 if any number of edges incident to a single vertex are added and/or deleted.

Talagrand [17] found another method that yields similar results. A combinatorial version of the Talagrand inequality requires, besides the Lipschitz condition, one more, quite technical condition, but in return it yields very often stronger bounds than Azuma's inequality. (For proof, see e.g. [17] or [9, Theorem 2.29].)

Theorem 2 (Talagrand). *Let $Z_1, \ldots, Z_M$ be independent random variables, with Z_j taking values in a set Λ_j. Assume that a function $f : \Lambda_1 \times \Lambda_2 \times \cdots \times \Lambda_M \to \mathbb{R}$ satisfies, for some constants b_j, $j = 1, \ldots, M$, and some function ψ, the following two conditions:*

(L) *If two vectors $\mathbf{z}, \mathbf{z}' \in \Lambda_1, \ldots, \Lambda_M$ differ only in the jth coordinate, then $|f(\mathbf{z}) - f(\mathbf{z}')| \leq b_j$.*

(C) *If $\mathbf{z} \in \Lambda$ and $r \in \mathbb{R}$ with $f(\mathbf{z}) \geq r$, then there exists a set $J \subseteq \{1, \ldots, M\}$ with $\sum_{j \in J} b_j^2 \leq \psi(r)$, such that for all $\mathbf{y} \in \Lambda$ with $y_j = z_j$ when $j \in J$, we have $f(\mathbf{y}) \geq r$.*

Then, the random variable $X = f(Z_1, \ldots, Z_M)$ satisfies, for any $r \in \mathbb{R}$ and $t \geq 0$,

$$\mathbb{P}(X \leq r - t)\, \mathbb{P}(X \geq r) \leq e^{-t^2/4\psi(r)}. \tag{5}$$

In particular, if m is a median of X, then for every $t \geq 0$,

$$\mathbb{P}(X \leq m - t) \leq 2e^{-t^2/4\psi(m)} \tag{6}$$

and

$$\mathbb{P}(X \geq m + t) \leq 2e^{-t^2/4\psi(m+t)}. \tag{7}$$

Remark 1. A recent inequality of Boucheron, Lugosi and Massart [2] is sometimes an interesting alternative to Talagrand's inequality; in several applications it yields essentially the same result (with better constants). We do not, however, see any way to use their inequality in the set-up treated here.

2.2 Kim–Vu concentration via average smoothness

Inequalities from the previous section become weaker when the Lipschitz coefficient are large. Kim and Vu in [11] developed a method yielding concentration bounds which depend only on the "average" Lipschitz coefficients, typically much smaller than the "worst-case" ones. Very recently Vu wrote an excellent expository paper on that method [21].

Their setup is less general than that of Azuma's and Talagrand [though, still more general than that of Theorem 0]. Let $X = X(\xi_i : i \in \Gamma = [N])$ be a polynomial of degree k, where again ξ_i are independent random 0–1 variables. For a nonempty set $A \in \Gamma^{[k]}$ let $\partial_A X$ be the partial derivative of X with respect to the variables in A and define, for $j = 0, \ldots, k$, $\mathbb{E}_j(X) = \max_{|A| \geq j} \mathbb{E}(\partial_A X)$. (Thus $\mathbb{E}_0(X) = \mathbb{E}(X) = \lambda$.)

We begin with the Main Theorem of [11], the first main theorem proved by this method.

Theorem 3A (Kim and Vu). *For any $\ell \geq 1$, if $\lambda \geq \mathbb{E}_1(X)$, then*

$$\mathbb{P}\left(|X - \lambda| \geq c_k \ell^k \sqrt{\lambda\, \mathbb{E}_1(X)}\right) \leq \exp\{-\ell + (k-1)\log N\}. \tag{8}$$

The above theorem is derived from a more general, but more technical, concentration result, also proved in [11]. Let, for each $i = 1, \ldots, M$,

$$E_i = \Big| \mathbb{E}(X|\xi_1, \ldots, \xi_{i-1}, \xi_i = 1) - \mathbb{E}(X|\xi_1, \ldots, \xi_{i-1}, \xi_i = 0) \Big|.$$

[Note that, as a conditional expectation, E_i is a random variable which is a function of $(\xi_1, \ldots, \xi_{i-1})$.] Let further $M = \max_i E_i$, $W = \sum_i p_i E_i$ and $V = \sum_i p_i q_i E_i^2$.

Theorem 3B (Kim and Vu). *Let a, v, ℓ be positive numbers such that $0 < \ell < v/a^2$. Then*

$$\begin{aligned}\mathbb{P}\left(|X - \lambda| > \sqrt{\ell v}\right) &< 2\exp\{-\ell/4\} + \mathbb{P}(M > a \text{ or } V > v) \\ &\leq 2\exp\{-\ell/4\} + \sum_i \mathbb{P}(E_i > a) + \mathbb{P}(W > v/a). \end{aligned} \tag{9}$$

These results have been developed further and successfully applied to a variety of problems by Kim and Vu, see for example [11, 18, 19, 20] and the survey [21], where also further references are given. In particular, Vu [21, Corollary 3.4] has proved the following general and widely applicable result, using Theorem 3B and induction. For further similar results see [18, 21]. It is easily seen that Theorem 3C always yields at least as strong bounds as Theorem 3A; this is illustrated by the examples in Section 3.

Theorem 3C (Vu). *Let $\mathcal{E}_0 > \mathcal{E}_1 > \cdots > \mathcal{E}_k$ and ℓ be positive numbers such that $\mathcal{E}_j \geq \mathbb{E}_j(X)$, $0 \leq j \leq k$, and $\mathcal{E}_j/\mathcal{E}_{j+1} \geq \ell + j \log n$, $0 \leq j \leq k-1$. Then*

$$\mathbb{P}\left(|X - \lambda| \geq \sqrt{\ell \mathcal{E}_0 \mathcal{E}_1}\right) \leq C_k \exp\{-c_k \ell\}. \tag{10}$$

2.3 Combinatorial techniques

In this subsection we collect techniques that require only an elementary, combinatorial argument. We will be assuming throughout that all elements of $\mathcal{S}$ have the same size k, and that for all i, $p_i = p$ for some $0 < p \leq 1$. Thus $\lambda = \mathbb{E}\, X = |\mathcal{S}|p^k$. For convenience, the deviation parameter t will be expressed here in the form $t = \rho\, \mathbb{E}\, X = \rho\lambda$, though ρ, in general, is not necessarily a constant, and we will for simplicity only consider the case $0 < \rho \leq 1$. Finally, set $|\Gamma| = N$.

2.3.1 Converting to the lower tail

Here we aim to convert an upper tail probability into a lower tail and then to apply Theorem 0. This approach works mainly for dense families $\mathcal{S}$.

Theorem 4. *Let $k \geq 1$ be an integer, and assume that $\mathcal{S} \subseteq [\Gamma]^k$, $|\mathcal{S}| = \eta\binom{N}{k}$, $0 < \eta \leq 1$, $0 < p < 1$ and $0 < \rho \leq 1$. Then there exists a constant $c_k > 0$, depending on k only, such that*

$$\mathbb{P}\big(X \geq (1+\rho)\lambda\big) \leq 2e^{-c_k \rho^2 \eta^2 N p}.$$

Proof. Set $Z = |\Gamma_p|$, $\bar{\mathcal{S}} = [\Gamma]^k \setminus \mathcal{S}$, and

$$\bar{X} = \sum_{A \in \bar{\mathcal{S}}} I_A = \binom{Z}{k} - X,$$

and choose $\delta = \rho\eta/3k$. Then, by Chernoff's bound (1),

$$\mathbb{P}\big(X \geq (1+\rho)\lambda\big) = \mathbb{P}\left(\bar{X} \leq \binom{Z}{k} - (1+\rho)\lambda\right)$$
$$\leq \mathbb{P}\left(\bar{X} \leq \binom{(1+\delta)Np}{k} - (1+\rho)\lambda\right) + e^{-\delta^2 Np/3},$$

Now, $k\delta \leq 1/3$ and thus

$$(1+\delta)^k < e^{k\delta} < 1 + e^{1/3}k\delta < 1 + \tfrac{1}{2}\rho\eta. \tag{11}$$

Moreover, we may assume that $(1+\rho)\lambda \leq |\mathcal{S}|$, since otherwise the probability in question is 0. Hence $(1+\rho)p^k \leq 1$, which in combination with (11) yields $(1+\delta)^k p^k < 1$, or $(1+\delta)p < 1$, and thus $\binom{(1+\delta)Np}{k} < (1+\delta)^k \binom{N}{k} p^k$. Consequently, using (11) again,

$$\mathbb{P}\left(\bar{X} \leq \binom{(1+\delta)Np}{k} - (1+\rho)\lambda\right) \leq \mathbb{P}\left(\bar{X} \leq \Big[(1+\delta)^k - \eta(1+\rho)\Big]\binom{N}{k}p^k\right)$$
$$\leq \mathbb{P}\left(\bar{X} \leq (1 - \eta - \tfrac{1}{2}\rho\eta)\binom{N}{k}p^k\right) = \mathbb{P}\left(\bar{X} \leq \mathbb{E}\,\bar{X} - \tfrac{1}{2}\rho\eta\binom{N}{k}p^k\right).$$

In order to apply Theorem 0 to $\bar{X}$, assume that $Np \geq 1$ (otherwise the inequality is trivial if $c_k \leq \log 2$) and note that then the quantity $\bar{\Delta}$ corresponding to the family $\bar{\mathcal{S}}$ is smaller than $kN^{2k-1}p^{2k-1}$. Hence, by Theorem 0,

$$\mathbb{P}\left(\bar{X} \leq \mathbb{E}\,\bar{X} - \tfrac{1}{2}\rho\eta\binom{N}{k}p^k\right) \leq \exp\left(-\frac{\left(\rho\eta\binom{N}{k}p^k\right)^2}{8\bar{\Delta}}\right) \leq e^{-c'_k\rho^2\eta^2 Np},$$

for some c'_k. The theorem follows with $c_k = \min(1/27k^2, c'_k)$. □

2.3.2 Breaking into disjoint matchings

The underlying idea is to break the family $\mathcal{S}$ into disjoint subfamilies of disjoint sets, and apply Chernoff's bound to one subfamily. Set $L = L(\mathcal{S})$ for the standard dependency graph of the family of indicators $\{I_A : A \in \mathcal{S}\}$, where an edge joins A and B if and only if $A \cap B \neq \emptyset$. Note that the maximum degree $\Delta(L)$ can be as large as $|\mathcal{S}| - 1$. The following result has appeared in a slightly more complicated form in [14].

Theorem 5. *Let $t = \lambda\rho$, where $\rho > 0$. Then*

$$\mathbb{P}(X \geq \lambda + t) \leq (\Delta(L)+1)\exp\left(-\frac{\rho^2\lambda}{4(\Delta(L)+1)(1+\rho/3)}\right).$$

Proof. A matching in $\mathcal{S}$ is a subfamily $\mathcal{M} \subseteq \mathcal{S}$ consisting of pairwise disjoint sets.

By the well known Hajnal–Szemerédi Theorem [6], the vertex set of the graph L can be partitioned into $\Delta(L)+1$ independent sets, each of size equal to either $\lceil |\mathcal{S}|/(\Delta(L)+1)\rceil$ or $\lfloor |\mathcal{S}|/(\Delta(L)+1)\rfloor$. These sets correspond to matchings $\mathcal{M}_i$, $i=1,\ldots,\Delta(L)+1$, in $\mathcal{S}$. Note that for each i, $|\mathcal{M}_i| > |\mathcal{S}|/2(\Delta(L)+1)$. If $X \geq \lambda + t$, then, by simple averaging, there exists a matching $\mathcal{M}_i$ such that $|[\Gamma_p]^k \cap \mathcal{M}_i| \geq p^k|\mathcal{M}_i| + t|\mathcal{M}_i|/|\mathcal{S}|$. Since $|[\Gamma_p]^k \cap \mathcal{M}_i|$ is a random variable with the binomial distribution $\mathrm{Bi}(|\mathcal{M}_i|, p^k)$, we conclude by (1) that

$$\begin{aligned}\mathbb{P}(X \geq \lambda + t) &\leq \sum_{i=1}^{\Delta(L)+1} \exp\left(-\frac{t^2|\mathcal{M}_i|}{2|\mathcal{S}|(\lambda+t/3)}\right)\\ &\leq (\Delta(L)+1)\exp\left(-\frac{t^2}{4(\Delta(L)+1)(\lambda+t/3)}\right). \qquad \square\end{aligned}$$

As our later examples show, Theorem 5 can be applied to quite sparse families $\mathcal{S}$.

2.3.3 The deletion method

The next idea for establishing a bound on the upper tail of X resembles the tactic of sweeping under the rug. We delete some elements of Γ_p and claim the concentration of X in the remainder. Originally, this approach was used in the context of partition properties of random graphs [15]. Recently, it was developed further in [10], where it is shown to often yield essentially the same results as the method of Section 2.2.

Lemma 1. *Let $\mathcal{S} \subseteq [\Gamma]^k$ and $0 < p < 1$. Then, for every pair of positive real numbers r and t, with probability at least $1 - \exp\left(-\frac{rt}{k(\lambda+t)}\right)$, there exists a set $E_0 \subset \Gamma_p$ of size r such that $\Gamma_p \setminus E_0$ contains fewer than $\lambda + t$ sets from $\mathcal{S}$.*

Proof. Given r and t, let $\mathcal{A}$ be the event that for each set $E \subset \Gamma_p$ of size r, $\Gamma_p \setminus E$ contains at least $\lambda+t$ sets from $\mathcal{S}$. Let Z be the number of $\kappa = \lceil r/k \rceil$-element sequences of disjoint sets from $\mathcal{S}$ in Γ_p. If the event $\mathcal{A}$ holds, then $Z \geq (\lambda+t)^\kappa$, since we may choose κ elements sequentially with at least $\lambda + t$ choices each time. On the other hand, $\mathbb{E}(Z) \leq |\mathcal{S}|^\kappa p^{k\kappa} = \lambda^\kappa$ and thus, by Markov's inequality,

$$\mathbb{P}(\mathcal{A}) \leq \mathbb{P}\big(Z \geq (\lambda+t)^\kappa\big) \leq \frac{\mathbb{E}(Z)}{(\lambda+t)^\kappa} \leq \left(\frac{\lambda}{\lambda+t}\right)^{r/k} \leq \exp\Big(-\frac{rt}{k(\lambda+t)}\Big). \qquad \square$$

In subsection 3.3 we show a direct application of Lemma 1. Now we derive from it a genuine bound on the upper tail of X in terms of the maximum degree in the subhypergraph $\mathcal{S}_p = \mathcal{S} \cap [\Gamma_p]^k$. Let, for $I \subset \Gamma$, $Y_I := \sum_{J \supseteq I} X_J$ and

$$Y_1^* := \max_{i \in \Gamma} Y_{\{i\}}.$$

Clearly, $Y_1^* = \Delta(\mathcal{S}_p)$ and there are at most rY_1^* elements of $\mathcal{S}_p$ which could have been destroyed by removing r elements from Γ_p.

Theorem 6A. *Let $t = \rho\lambda$, $0 < \rho \le 1$. Then, for every real $r > 0$,*

$$\mathbb{P}(X \ge \lambda + t) \le \exp\{-\rho r/3k\} + \mathbb{P}(Y_1^* > t/2r).$$

Proof. If $X \ge \lambda + t$ and $Y_1^* \le t/2r$ then for every set $E \subset \Gamma_p$ with $|E| \le r$, $\Gamma_p \setminus E$ contains at least

$$X - rY_1^* \ge \lambda + t/2$$

sets from $\mathcal{S}$. By Lemma 1 the probability of the latter event is at most

$$\exp\left(-\frac{rt/2}{k(\lambda + t/2)}\right) \le \exp\{-\rho r/3k\}. \qquad \square$$

The probability $\mathbb{P}(Y_1^* > t/2r)$ can be annihilated by choosing $r = \frac{t}{2\Delta(\mathcal{S})}$. This gives as a corollary a bound in terms of maximum degree in $\mathcal{S}$.

Theorem 6B. *Let $t = \rho\lambda$, $0 < \rho \le 1$. Then*

$$\mathbb{P}(X \ge \lambda + t) \le \exp\{-\rho^2\lambda/(6k\Delta(\mathcal{S}))\}.$$

Since $\Delta(\mathcal{S}) - 1 \le \Delta(L(\mathcal{S})) \le k\Delta(\mathcal{S})$, this theorem yields almost the same bound as Theorem 5.

Alternatively, one can bound $\mathbb{P}(Y_1^* > t/2r) \le \sum \mathbb{P}(Y_{\{i\}} > t/2r)$ in Theorem 6A and hope for the best. For $k = 2$, $Y_{\{i\}}$ is the number of surviving neighbors of vertex i, which is a sum of independent random variables, so Chernoff's inequality can be applied. For $k > 2$ each $Y_{\{i\}}$ is a sum of dependent random variables but of the same type as X, which gives room for induction. One way of doing it leads to the following result, the proof of which is presented in [10] (cf. Corollary 4.1 there). Let $\Delta_j = \Delta_j(\mathcal{S})$ be the maximum number of elements of $\mathcal{S}$ containing a given j-element set, $j = 0, \ldots, k$.

Theorem 6C. *Let $\lambda_j^* = \Delta_j p^{k-j}$ [thus $\lambda_0^* = \lambda$]. Then, with $c_k = 1/12k$, and for all $0 < \rho \le 1$,*

$$\mathbb{P}(X \ge (1+\rho)\lambda) \le 2N^{k-1} \exp\left\{-c_k \min_{1 \le j \le k} \left(\frac{\rho^2\lambda}{\lambda_j^*}\right)^{1/j}\right\}.$$

2.3.4 Approximating by a disjoint subfamily

Our last method originates in Spencer [16] (cf. Example 1, below). Let X_0 be the largest number of disjoint sets of $\mathcal{S}$ which are present in Γ_p. Clearly, $X \geq X_0$, but sometimes X is not much larger than X_0. Intuitively, X_0 should have a distribution similar to that of a sum of independent random variables, and thus a Chernoff-type bound could be true. Our next lemma makes it precise.

Lemma 2. *If $t \geq 0$, then, with $\phi(x) = (1+x)\log(1+x) - x$ and $\lambda = \mathbb{E}\, X$,*

$$\mathbb{P}(X_0 \geq \lambda + t) \leq \exp\left(-\lambda\phi\Big(\frac{t}{\lambda}\Big)\right) \leq \exp\left(-\frac{t^2}{2(\lambda + t/3)}\right). \qquad \square$$

For the proof of Lemma 2, which is similar to the proof of Lemma 1, see [9, Lemma 2.46].

In order to relate X to X_0 we invoke a simple graph theoretic fact. For an arbitrary graph G set $\alpha(G)$ for its independence number, $\beta(G)$ for the size of the largest induced matching in G, and $\Delta_e(G)$ for the maximum of $|N_G(v) \cup N_G(v')|$ over all pairs (v, v') of adjacent vertices of G. Note that $\Delta_e(G) \leq 2\Delta(G)$.

Lemma 3. *For every graph G, we have $|V(G)| \leq \alpha(G) + \beta(G)\Delta_e(G)$.*

Proof. Let M be the set of edges in a maximal induced matching of G. The vertices outside M can be divided into two groups: those adjacent to M [at most $\beta(G)(\Delta_e(G) - 2)$ of them] and those which are not. The latter group of vertices form an independent set in G and the lemma follows. $\square$

Consider the intersection graph $L_p = L(\mathcal{S}_p)$ of $\mathcal{S}_p = \mathcal{S} \cap [\Gamma_p]^k$, in which every vertex represents one set of $\mathcal{S}_p$, and the edges join pairs of vertices representing pairs of intersecting sets. (Note that L_p is an induced subgraph of L defined above.) Then X is the number of vertices and X_0 is the independence number of L_p. Thus Lemma 3 implies that

$$X \leq X_0 + \beta(L_p)\Delta_e(L_p) \leq X_0 + 2\beta(L_p)\Delta(L_p). \tag{12}$$

We have already estimated the probability that X_0 is large; hence estimates of $\beta(L_p)$ and $\Delta(L_p)$ yield bounds on the upper tail of X. This leads to the following result. Recall that $\Delta_j = \Delta_j(\mathcal{S})$ is the maximum number of elements of $\mathcal{S}$ containing a given j-element set, $j = 0, \ldots, k$.

Theorem 7. *Define* $R = \lambda \max_{1\le j\le k-1} \binom{k}{j} \Delta'_j p^{k-j}$, *where* $\Delta'_j = \Delta_j - 1$. *Then, for every* $t > 0$ *and* $r \ge R$,

$$\mathbb{P}(X \ge \lambda + t) \le \exp\left(-\frac{t^2}{8\lambda + 2t}\right) + (k-1)\exp\{-r\} + \mathbb{P}\left(Y_1^* > \frac{t}{12k(k-1)r}\right). \tag{13}$$

Remark 2. Very often the third term can be bounded (e.g. by a Chernoff inequality, see the examples in Section 3) by $\exp(-\Theta(t/r))$. Then, if, in addition, $t = \Theta(\lambda) \to \infty$, the first term is smaller than the sum of the other two (either r or t/r is not greater than $\sqrt{t}$), and therefore can be ignored. Also, note that for fixed ρ, if we ignore the first term, we are up to constants in the exponents left with the same bound as Theorem 6A, although now only for $r \ge R$. Hence the above bound then is never better (up to constants in the exponent) than that from Theorem 6A, and it yields essentially the same result when the optimal r for Theorem 6A satisfies $r \ge R$.

Proof. By (12), for any $\ell > 0$,

$$\mathbb{P}(X \ge \lambda + t) \le \mathbb{P}(X_0 \ge \lambda + t/2) + \mathbb{P}(\beta(L_p) > \ell) + \mathbb{P}(\Delta(L_p) > t/(4\ell)).$$

By Lemma 2, the first term is smaller than $\exp\left(-\frac{t^2/4}{2(\lambda+t/6)}\right) \le \exp\left(-\frac{t^2}{8\lambda+2t}\right)$. For the third term observe that $\Delta(L_p) \le k\Delta(\mathcal{S}_p) = kY_1^*$, so $\mathbb{P}(\Delta(L_p) > t/4\ell) \le \mathbb{P}(Y_1^* > t/(4k\ell))$.

It remains to estimate $\mathbb{P}(\beta(L_p) > \ell)$. First note that an induced matching in L_p with m edges corresponds to $2m$ distinct sets $A_i, B_i \in \mathcal{S}_p$, $1 \le i \le m$, such that $A_i \cap B_i \ne \emptyset$, but $A_i \cap A_j = A_i \cap B_j = B_i \cap B_j = \emptyset$, $i \ne j$. Consequently, if $\tilde{\mathcal{S}} = \{A \cup B : A, B \in \mathcal{S}, A \ne B, A \cap B \ne \emptyset\}$, then $\beta(L_p)$ is the maximum size of a disjoint subfamily of $\tilde{\mathcal{S}}_p$, i.e. $\beta(L_p) = \alpha(L(\tilde{\mathcal{S}}_p))$. Thus, $\beta(L_p)$ is a random variable of the same type as X_0, except that $\tilde{\mathcal{S}}$ is not a uniform hypergraph. To overcome this mild difficulty, we write $\tilde{\mathcal{S}} = \bigcup_{j=1}^{k-1} \tilde{\mathcal{S}}^j$, where $\tilde{\mathcal{S}}^j = \{A \cup B : A, B \in \mathcal{S}, |A \cap B| = j\}$, and thus have, with $Y_0^j = \alpha(L(\tilde{\mathcal{S}}_p^j))$,

$$\beta(L_p) = \alpha\left(L\left(\bigcup_{j=1}^{k-1} \tilde{\mathcal{S}}_p^j\right)\right) \le \sum_{j=1}^{k-1} Y_0^j.$$

Hence,

$$\mathbb{P}(\beta(L_p) > \ell) \le \sum_{j=1}^{k-1} \mathbb{P}\big(Y_0^j > \ell/(k-1)\big).$$

In order to apply Lemma 2 to Y_0^j, we need to estimate the expected size of $\tilde{\mathcal{S}}_p^j$. For each $A \in \mathcal{S}$ there are at most $\binom{k}{j}\Delta'_j$ choices of $B \in \mathcal{S}$ with $|A \cap B| = j$. Hence, setting $Y^j = |\tilde{\mathcal{S}}_p^j|$,

$$\mathbb{E}\, Y^j \le \binom{k}{j} \Delta'_j |\mathcal{S}| p^{2k-j} = \binom{k}{j} \Delta'_j p^{k-j} \lambda \le R \le r$$

and, by Lemma 2, with $\ell = 3(k-1)r$,

$$\mathbb{P}\big(Y_0^j > \ell/(k-1)\big) = \mathbb{P}(Y_0^j > 3r) \le \mathbb{P}(Y_0^j > \mathbb{E}\,Y^j + 2r) \le e^{-r}. \qquad \square$$

Sometimes, $\beta(L_p)$ and $\Delta(L_p)$ (or $\Delta_e(L_p)$) can be estimated with more ease. This is the case of the next example from [16] where this technique originated.

Example 1. Let $X^{(v)}$ be the number of triangles of $G(n,p)$ containing a given vertex v, where $p = (\omega \log n)^{1/3} n^{-2/3}$ and $\omega \to \infty$ with $\omega \le \log n$, say. The problem is to estimate $\min_v X^{(v)}$ and $\max_v X^{(v)}$. Set $X = X^{(1)}$. A standard application of Chebyshev's inequality yields that $X/\lambda \xrightarrow{\mathrm{p}} 1$, and for example $0.9\lambda < X < 1.1\lambda$ with probability $1 - O(1/\lambda)$, where $\lambda = \Theta(\omega \log n)$. This, however, is not enough to claim that such a concentration of the number of triangles holds for every vertex of $G(n,p)$. For this to be true, we need to decrease the error probability down to $o(1/n)$. Of course, there is no problem with the lower tail. By Theorem 0, $\mathbb{P}\big(X \le 0.9\lambda\big) \le \exp\{-\Omega(\lambda)\} = o(1/n)$.

It turns out that with probability $1 - o(1/n)$, we have $X \le X_0 + 12$. This follows quite easily by the first inequality in (12). Indeed, with that probability $\beta(L_p) \le 3$ and $\Delta_e(L_p) \le 4$. [The right hand side of (12) gives only $X \le X_0 + 18$, as, with probability $1 - o(1/n)$, we may only bound $\Delta(L_p) \le 3$.] Hence, for large enough n, by Lemma 2,

$$\mathbb{P}(X \ge 1.1\lambda) \le \mathbb{P}(X_0 \ge 1.1\lambda - 12) + o(1/n) \le \mathbb{P}(X_0 \ge 1.05\lambda) + o(1/n) \le \exp\{-c\lambda\} + o(1/n) = o(1/n).$$

3 Applications

In this section we illustrate all methods introduced earlier by a couple of pivotal applications and compare the results. We will refer to the methods via the following numbers and/or nicknames: 1) Azuma (Theorem 1), 2) Talagrand (Theorem 2), 3A,B,C) Kim–Vu (Theorems 3A, 3B, 3C), 4) Complement (Theorem 4), 5) Break-up (Theorem 5), 6A,B,C) Deletion (Theorems 6A, 6B, 6C), 7) Approximation (Theorem 7).

We always express $t = \rho\lambda$, $0 < \rho \le 1$, and always compare bounds on $\mathbb{P}(X > \lambda + t)$ (except for Method 2 where we have the median m in place of λ), ignoring insignificant constants. We denote insignificant constants (that may be made explicit) by c; the value of c may depend on k and other parameters, and may change from one occurence to the next.

3.1 Random induced subhypergraphs

The following problem about a random subgraph obtained by a random deletion of vertices of a given graph was studied in [14]. Here we state it for k-uniform hypergraphs.

Let $H = (V, E)$ be a k-uniform hypergraph with $|V| = N$ and $|E| = \eta\binom{N}{k}$, $0 < \eta \leq 1$, and let V_p be a binomial random subset of the vertex set V, $0 < p < 1$. Here every vertex i is associated with a 0–1 random variable ξ_i, where $\mathbb{P}(\xi_i = 1) = p$. Thus, $Y = |V_p|$ has the binomial distribution $\mathrm{Bi}(N, p)$, and $\mathcal{S}$ is the family of all hyperedges of H. We let Δ_H denote the maximum degree of H, i.e. the maximum number of hyperedges containing a given vertex. Note that $|E| \leq N\Delta_H/k$. Moreover, $\Delta_j = \Delta_j(H)$ is the maximum number of hyperedges of H containing a given j-element set, $j = 0, \ldots, k$; thus $\Delta_1 = \Delta_H$.

We want to show that with probability very close to 1, the random variable $X = |[V_p]^k \cap E|$ is, say, not larger than $2\eta\binom{Np}{k}$, i.e., that the density of the random subgraph is not much larger than the density of the initial hypergraph. Note that $\lambda = \mathbb{E}\,X = \eta\binom{N}{k}p^k \sim \eta\binom{Np}{k}$, and so this problem falls into our general framework of estimating upper tails. In fact, this is precisely the setup of our combinatorial methods 4–7 described in general terms in Section 2. Below, we study asymptotics as $N \to \infty$, where η, p and ρ may depend on N, with $\rho \leq 1$, while $k \geq 2$ is fixed. Throughout we assume that $Np \to \infty$, and in particular $Np \geq 1$.

We will try all seven methods in two basic cases. In the general case, where no assumption is made on H, we will use the trivial bounds $\Delta_j(H) \leq N^{k-j}$, $j = 1, \ldots, k$. In the *highly regular* case (and presumably sparse, meaning $\eta \to 0$) we will be assuming that $\Delta_j = \Theta(\eta N^{k-j})$, $j = 1, \ldots, k-1$. (The unspecified constants c below may depend on the constants implicit in this Θ.) Note that $\Delta_0 = |E| = \Theta(\eta N^k)$ by definition, while $\Delta_k = 1$ unless H is empty.

As a warm-up consider first the special case in which H is the complete k-uniform hypergraph K_N^k, i.e. $\eta = 1$. Then $X = \binom{Y}{k} < Y^k/k!$, and $\lambda = (Np)^k\big(1/k! + o(1/Np)\big)$. Hence, assuming ρNp is large enough,

$$\mathbb{P}\left(X > (1+\rho)\lambda\right) \leq \mathbb{P}\left(Y^k > (1+\rho/2)(Np)^k\right) = \mathbb{P}\left(Y > (1+\rho/2)^{1/k}Np\right) \leq \mathbb{P}\left(Y > \left(1 + \frac{\rho}{3k}\right)Np\right),$$

and by Chernoff's bound (1), we obtain

$$\mathbb{P}\left(X > (1+\rho)\lambda\right) \leq \exp\{-c\rho^2 Np\}. \tag{14}$$

This bound is essentially sharp and we will use it to measure the accuracy of all our methods applied to the general case.

1) Azuma. Adding/deleting a vertex may change the value of X by at most the degree of that vertex. Thus, applying Theorem 1 with $b_i = \deg_H(i)$ we obtain the bound

$$\exp\left\{-c\frac{\rho^2\eta^2 N^{2k-1}p^{2k}}{\Delta_H^2}\right\}.$$

2) Talagrand. Here again the Lipschitz condition is satisfied with $b_i = \deg_H(i)$. However, the obvious choice $\psi(r) = N\Delta_H^2$ yields only the same bound as Azuma, with an inferior constant. To get a better estimate we have to modify X by a sort of truncation. Let

$$X' = \max\{|[V']^k \cap E| : V' \subseteq V_p, |V'| \leq 2Np\}.$$

Clearly, X' satisfies condition (L) with the same constants b_i as X does, and condition (C) with $\psi(r) = 2Np\Delta_H^2$. By Chernoff's bound (1),

$$\mathbb{P}(X \neq X') \leq \mathbb{P}(|V_p| > 2Np) \leq e^{-cNp}.$$

This and (7) yield the bound

$$\exp\left\{-c\frac{\rho^2\eta^2 N^{2k-1}p^{2k-1}}{\Delta_H^2}\right\} + e^{-cNp} < 2\exp\left\{-c\frac{\rho^2\eta^2 N^{2k-1}p^{2k-1}}{\Delta_H^2}\right\}.$$

3A) Kim–Vu A. We can write

$$X = \sum_{A\in\mathcal{S}}\prod_{i\in A}\xi_i.$$

Thus,

$$\partial_A X = \sum_{A\subseteq B\in\mathcal{S}}\prod_{i\in B\setminus A}\xi_i,$$

and,

$$\mathbb{E}_1(X) = \max_{1\leq j\leq k}\Delta_j(H)p^{k-j} \leq \max_{1\leq j\leq k}(Np)^{k-j} = (Np)^{k-1}. \tag{15}$$

Hence, (8) with $l = c_k^{-1/k}\rho^{1/k}\lambda^{1/2k}(\mathbb{E}_1(X))^{-1/2k}$implies (the case $\mathbb{E}_1(X) > \lambda$ is trivial here, since then ℓ is bounded)

$$\mathbb{P}\left(|X-\lambda| \geq \rho\lambda\right) \leq N^{k-1}\exp\{-c\rho^{1/k}(\eta Np)^{1/2k}\}. \tag{16}$$

In the highly regular case, $\mathbb{E}_1(X) = \Theta\big(\max[1, \eta(Np)^{k-1}]\big)$ and the exponent in (16) is improved to $-c\rho^{1/k}(Np)^{1/2k}$ if $\eta(Np)^{k-1} \geq 1$ and $-c\rho^{1/k}\eta^{1/2k}(Np)^{1/2}$ otherwise.

3B) Kim–Vu B. Observe that

$$E_i = \sum_{j=0}^{k-1} \sum_{i\in A\in E,\, |A\setminus[i]|=j} p^j \prod_{l\in A\cap[i-1]} \xi_l.$$

This is a polynomial of degree $k-1$, so in principle one could apply induction here. However, as this approach seems to be quite complicated, we refer instead to Theorem 3C, see below, where the work with the induction already has been done, in a general setting, by Vu [21].

We here thus focus on the case $k=2$, i.e. the case of graphs, because only then E_i becomes a sum of independent variables ($j=0$) and a constant term ($j=1$). Indeed, for $k=2$,

$$E_i = \sum_{\substack{j<i\\ \{i,j\}\in E}} \xi_j + \sum_{\substack{j>i\\ \{i,j\}\in E}} p \le Z + p\Delta_H,$$

where $Z \in \mathrm{Bi}(\Delta_H, p)$. Also,

$$W = p\sum_i E_i \le p\Delta_H Y + p^2|E|.$$

By Chernoff's bound (1), $\mathbb{P}(Y \ge 3Np) < e^{-Np}$, and so, with probability at least $1-e^{-Np}$, we have $W < 4\Delta_H Np^2$. Now, choose $a = bp\Delta_H$, where $b \ge 9$, and $v = 4a\Delta_H Np^2$. By (1), with $\mathbb{E}\, Z = \Delta_H p$,

$$\mathbb{P}(E_i > a) \le \mathbb{P}\big(Z \ge (b-1)\,\mathbb{E}\, Z\big) = \mathbb{P}\big(Z - \mathbb{E}\, Z \ge (b-2)\,\mathbb{E}\, Z\big) \le \exp\big(-\phi(b-2)\,\mathbb{E}\, Z\big),$$

where $\phi(b-2) = (b-1)\log(b-1) - b + 2 \ge b$, because $\log(b-1) > 2$. Hence, by Theorem 3B, setting $\ell = \rho^2\lambda^2/v$ and checking that $\ell < v/a^2$ (this follows from $\lambda < N\Delta_H p^2$), we obtain the estimate

$$\mathbb{P}\big(X \ge (1+\rho)\lambda\big) \le 2e^{-\ell/4} + Ne^{-bp\Delta_H} + e^{-Np}. \tag{17}$$

We have to choose b optimally. We would like $\ell/4 = bp\Delta_H$, i.e. $b = \frac{1}{8}\rho\eta(N-1)N^{1/2}\Delta_H^{-3/2}$, but also $b \ge 9$. Thus, choosing $b = \max\big(\frac{1}{8}\rho\eta(N/\Delta_H)^{3/2}, 9\big)$ in (17) we find, since then $\ell/4 < v/(4a^2) = Np/b < Np$,

$$\mathbb{P}\big(X \ge (1+\rho)\lambda\big) \le (N+3)e^{-\ell/4} \le \begin{cases} 2N\exp\left\{-c\frac{\rho\eta N^{3/2}p}{\Delta_H^{1/2}}\right\} & \text{if } \rho\eta(N/\Delta_H)^{3/2} \ge 72, \\ 2N\exp\left\{-c\frac{\rho^2\eta^2N^3p}{\Delta_H^2}\right\} & \text{otherwise.} \end{cases}$$

3C) (Kim–)Vu C. We have, generalizing (15),

$$\mathbb{E}_j(X) = \max_{i\ge j} \Delta_i(X)p^{k-i}.$$

We thus need to find ℓ and $\mathcal{E}_0,\dots,\mathcal{E}_k$ such that $\sqrt{\ell\mathcal{E}_0\mathcal{E}_1} \le \rho\lambda$, $\mathcal{E}_j \ge \Delta_j(H)p^{k-j}$, $0 \le j \le k$, and $\mathcal{E}_j/\mathcal{E}_{j+1} \ge \ell + j\log N$, $0 \le j \le k-1$.

It is easily checked that if $Np \ge 2k\log N$, then these conditions are satisfied with $\mathcal{E}_0 = \lambda = \Theta(\eta(Np)^k)$, $\mathcal{E}_j = (Np)^{k-j}$, $1 \le j \le k$, and $\ell = \rho^2\lambda^2/\mathcal{E}_0\mathcal{E}_1 = \rho^2\lambda/\mathcal{E}_1 = \Theta(\rho^2\eta Np)$. (It is also easily checked that these choices are essentially optimal without further assumptions.) Hence,

$$\mathbb{P}\big(X \ge (1+\rho)\lambda\big) \le C\exp(-c\ell) \le C\exp\left(-c\rho^2\eta Np\right), \qquad Np \ge 2k\log N.$$

As a trivial consequence, for any N and p,

$$\mathbb{P}\big(X \ge (1+\rho)\lambda\big) \le N\exp\left(-c\rho^2\eta Np\right).$$

In the highly regular case, we instead choose $\mathcal{E}_j = C\max\big((2\ell)^{k-j},\, \eta(Np)^{k-j}\big)$, $0 \le j \le k$, and $\ell = c\min(\rho^{2/k}\eta^{1/k}Np,\, \rho^2 Np)$; it is easily checked that the conditions then are satisfied, provided C is large, c is small and $\ell \ge k\log N$. Consequently, in the highly regular case,

$$\mathbb{P}\big(X \ge (1+\rho)\lambda\big) \le N\exp\big(-c\min(\rho^{2/k}\eta^{1/k}Np,\, \rho^2 Np)\big).$$

4) Complement. By Theorem 4,

$$\mathbb{P}\big(X \ge (1+\rho)\lambda\big) \le 2e^{-c\rho^2\eta^2 Np}.$$

5) Break-up. By Theorem 5, with $\Delta(L) < k\Delta_H$,

$$\mathbb{P}\big(X \ge (1+\rho)\lambda\big) \le k\Delta_H \exp\left\{-\frac{\rho^2\lambda}{4k\Delta_H(1+\rho/3)}\right\}.$$

6A) Deletion A. As for Theorem 3B, we consider only the case $k=2$; for larger k one can use induction, but we refer instead to Theorem 6C, where the induction already is done. If $k=2$, Theorem 6A yields

$$\mathbb{P}\big(X \ge (1+\rho)\lambda\big) \le e^{-\rho r/6} + \sum_i \mathbb{P}(Y_{\{i\}} > t/2r),$$

where $Y_{\{i\}} = \xi_i Z_i$ with $Z_i \in \mathrm{Bi}(\deg_H(i), p)$. Hence, if $t/2r \ge 7$, (1) implies $\mathbb{P}(Y_{\{i\}} > t/2r) \le \exp(-t/2r)$, see [9, Corollary 2.4], and thus

$$\mathbb{P}\big(X \ge (1+\rho)\lambda\big) \le e^{-\rho r/6} + Ne^{-\rho\lambda/2r}.$$

Choosing the optimal (ignoring polynomial factors) $r = \min\bigl(\sqrt{3\lambda},\ t/14\Delta_H p\bigr)$, we find

$$\mathbb{P}\bigl(X \geq (1+\rho)\lambda\bigr) \leq (N+1)e^{-\rho r/6} \leq \exp\Bigl(-c\min\Bigl(\rho\sqrt{\lambda},\ \frac{\rho^2\lambda}{\Delta_H p}\Bigr)\Bigr).$$

6B) Deletion B. By Theorem 6B,

$$\mathbb{P}\bigl(X \geq (1+\rho)\lambda\bigr) \leq \exp\left\{-\frac{\rho^2\lambda}{6k\Delta_H}\right\}.$$

This is essentially the same bound as the one given by Method 5.

6C) Deletion C. It follows from Theorem 6C, using $\Delta_j \leq N^{k-j}$, that

$$\mathbb{P}\bigl(X \geq (1+\rho)\lambda\bigr) \leq 2N^{k-1}\exp\{-c\eta\rho^2 Np\}.$$

In the highly regular case we obtain the (sometimes) better

$$\mathbb{P}\bigl(X \geq (1+\rho)\lambda\bigr) \leq 2N^{k-1}\exp\{-c\min(\rho^2, \rho^{2/k}\eta^{1/k})Np\}.$$

7) Approximation. This method is useless for dense hypergraphs H. Indeed, for it to be successful one needs that the ratio t/r should be at least of the order of $\mathbb{E}\, Y_1^* = \Theta((Np)^{k-1})$. However, if η is constant, we have $r \geq R = \Theta((Np)^{2k-1})$ while only $t = \Theta((Np)^k)$. Hence, the estimate (13) does not tend to 0.

Yet, the method can still yield good bounds for sparse and highly regular H. To avoid dependence in Y_1^*, we consider only the special case of graphs, i.e. $k=2$, and assume for simplicity that ρ is constant. Let us assume that $\Delta_1 = \Delta_H = \Theta(\eta N)$. Then $R = \Theta(\eta^2(Np)^3)$, $\mathbb{E}\, Y_1^* = \Theta(\eta Np)$, and via the Chernoff bound (1), the three terms on the right hand side of (13) are $\exp\bigl(-\Theta(\eta(Np)^2)\bigr)$, $(k-1)\exp(-r)$ and $\exp\bigl(-\Theta(t/r)\bigr)$, where we need $r \geq R$ and $t/r \geq C\eta Np$. A perfect choice would be $r = \sqrt{t} = \Theta(\sqrt{\eta}Np)$ which satisfies the above constraints provided $\eta \leq c(Np)^{-4/3}$. Then

$$\mathbb{P}\bigl(X \geq (1+\rho)\lambda\bigr) \leq 3\exp\{-c\sqrt{\eta}Np\}.$$

For larger η, the best choice is $r = R = \Theta(\eta^2(Np)^3)$, which yields the bound $e^{-c/(\eta Np)}$ which is meaningful if $\eta Np < 1$. In conclusion we find, for fixed ρ and with $c = c(\rho)$,

$$\mathbb{P}\bigl(X \geq (1+\rho)\lambda\bigr) \leq 3\exp\{-c\min(\eta^{1/2}Np, (\eta Np)^{-1})\}.$$

Summary of subsection 3.1.

	fixed η	$\eta \to 0$	$\eta \to 0$, H highly regular
1) Azuma	$\rho^2 N p^{2k}$	$\rho^2\eta^2 N p^{2k}$	$\rho^2 N p^{2k}$
2) Talagrand	$\rho^2 N p^{2k-1}$	$\rho^2\eta^2 N p^{2k-1}$	$\rho^2 N p^{2k-1}$
3A) Kim–Vu A	$\rho^{1/k}(Np)^{1/2k}$	$\rho^{1/k}\eta^{1/2k}(Np)^{1/2k}$	$\rho^{1/k}\min\bigl((Np)^{1/2k},\, \eta^{1/2k}(Np)^{1/2}\bigr)$
3B) Kim–Vu B, $k=2$	$\rho^2 Np$	$\rho^2\eta^2 Np$	$\min(\rho^2,\, \rho\sqrt{\eta})Np$
3C) (Kim–)Vu C	$\rho^2 Np$	$\rho^2\eta Np$	$\min(\rho^2,\, \rho^{2/k}\eta^{1/k})Np$
4) Complement	$\rho^2 Np$	$\rho^2\eta^2 Np$	$\rho^2\eta^2 Np$
5) Break-up	$\rho^2 N p^k$	$\rho^2\eta N p^k$	$\rho^2 N p^k$
6A) Deletion A, $k=2$	$\rho^2 Np$	$\rho^2\eta Np$	$\min(\rho^2,\, \rho\sqrt{\eta})Np$
6B) Deletion B	$\rho^2 N p^k$	$\rho^2\eta N p^k$	$\rho^2 N p^k$
6C) Deletion C	$\rho^2 Np$	$\rho^2\eta Np$	$\min(\rho^2,\, \rho^{2/k}\eta^{1/k})Np$
7) Approximation, $k=2$	—	—	$c(\rho)\min(\sqrt{\eta}Np,\, (\eta Np)^{-1})$

Table 1: Exponents for upper tail bounds in the subhypergraph problem

We first summarize the obtained results in Table 1. An entry F means a bound $O(N^{c_1})\exp(-c_2 F)$ for some positive constants c_1 and c_2 (c_1 might be 0). Hence the larger F, the better, i.e. (asymptotically) smaller, bound.

Consider first the case in which both ρ and η are constants. In this case $\Delta_H = \Theta(N^{k-1})$, and we see that Methods 3C (Vu), 4 (Complement), and 6C (Deletion) all yield the optimal value of Np. For $k=2$, Methods 3B and 6A too yield this value; as indicated above, with induction and more effort, they yield estimates of the same order in the exponent for higher k too. Note that the truncated Talagrand (Np^{2k-1}) is better than Azuma (Np^{2k}), and, most interestingly, Break-up (Np^k) is better than both of them.

Now allow $\eta \to 0$ with ρ constant. Then, in the worst case $\Delta_H = \Theta(N^{k-1})$, Methods 3B and 4 give only an exponent of order $\eta^2 Np$, while Methods 3C and 6C (and 6A for $k=2$) are now the best with ηNp.

Finally, assume that H is highly regular, still keeping ρ constant. If $\eta \gg p^{k(k-1)}$, then Methods 3C and 6C, and for $k=2$ at least, 3B, 6A and 7 too, tie up with the exponent $\Theta(\lambda^{1/k}) = \Theta(\eta^{1/k}Np)$ (Method 7 only in a restricted range of η). Method 4 is unaffected by the improvement in Δ_H. On the other hand, if $\eta \ll p^{k(k-1)}$, then Methods 5 (Break-up) and 6B (Deletion) are the best with Np^k.

If also $\rho \to 0$, we see that most methods give a factor ρ^2 in the exponent, except for Methods 3A (with only $\rho^{1/k}$, but here the main term is very weak and the exponent is never better than the one given by 3C), 7 (not investigated), and, in the highly regular case, 3B, 3C, 6A and 6C (with ρ for $k = 2$, as long as $\rho \geq \eta^{1/2}$). A detailed analysis of this case, with η fixed or decaying to 0, is left to the eager reader.

3.2 Small subgraphs of random graphs

Let $\mathbb{G}(n,p)$ be a binomial random graph. It can be viewed as a random subset of the set $[n]^2$ of all pairs formed by the set $\{1,\ldots,n\}$. Thus, $N = \binom{n}{2}$. 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 $G(n,p)$. Let $k = v(G)$ and assume $n \geq k$. Then

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

where $\mathrm{Aut}(G)$ is the set of automorphisms of G and $\asymp$ means that the quotient of the two sides is bounded from above and below by positive constants. The subgraph counts X_G have received a lot of attention from the pioneering paper by Erdős and Rényi [3] to the present day. Bounds for the upper tail (in the more general context of extensions) were considered by Spencer [16], but the break-through with general exponential bounds came with Vu [20].

In this subsection we try our techniques on estimates of the upper tail of X_G for three small subgraphs: $G = K_3$, $G = K_4$ and $G = C_4$. Throughout we for simplicity assume that $\rho = 1$, i.e. $t = \lambda$, and leave the case $\rho \to 0$ to the reader. As a test of how good these results are, we compare them with the exponential lower bound on $\mathbb{P}(X_G \geq (1+\rho)\lambda)$ obtained by Vu [20], see Section 1 for the case K_3 (see also [10, (6.2)]). For our three cases, $G = K_3$, $G = K_4$ and $G = C_4$, Vu's lower bound has the exponent $-cn^2p^2\log n$, $-cn^2p^3\log n$ and $-cn^2p^2\log n$, resp.

1) Azuma. Having a choice between vertex- and edge-exposure, we realize that the latter is better. Indeed, although the martingale is longer, the Lipschitz constants tend to be smaller, as fixing one edge leaves less freedom for creating a copy of a given graph than when fixing just a single vertex. Ignoring constants, for $G = K_3$, $G = K_4$, and $G = C_4$, we have $b_i = n$, $b_i = n^2$ and again $b_i = n^2$, resp. Thus, the exponents in the estimates are of the order n^2p^6, n^2p^{12}, and n^2p^8, resp.

2) Talagrand. A truncation similar to that in 3.1.2, restricting the number of edges to, say, n^2p, plus a Chernoff inequality, allows us to apply (7) with b_i as for Azuma and $\psi(r) = n^2p(\max b_i)^2$. This provides the exponents n^2p^5, n^2p^{11} and n^2p^7, resp., thus, by one power of p better than Method 1.

In the case $G = K_3$, for $p \gg n^{-1/3}$, a direct application of the Talagrand inequality, with the certificate being the set of all edges of $\mathbb{G}(n,p)$ (they all a.a.s. belong to the copies of K_3), yields only the exponent n^2p^6. For K_4 and C_4 too, a similar direct approach yields the same exponents as Method 1.

3A) Kim–Vu A. For $G = K_3$, we have $k = 3$ and $\mathbb{E}_1(X) = \max\{np^2, 1\}$. Hence, by Theorem 3A

$$\mathbb{P}(X_{K_3} \geq 2\lambda) \leq n^4 \begin{cases} \exp\left\{-cn^{1/3}p^{1/6}\right\} & \text{if } p \geq n^{-1/2} \\ \exp\left\{-cn^{1/2}p^{1/2}\right\} & \text{otherwise.} \end{cases}$$

Similarly, with $k = 6$ and $\mathbb{E}_1(X) = \max\{n^2p^5, 1\}$,

$$\mathbb{P}(X_{K_4} \geq 2\lambda) \leq n^{10} \begin{cases} \exp\left\{-cn^{1/6}p^{1/12}\right\} & \text{if } p \geq n^{-2/5} \\ \exp\left\{-cn^{1/3}p^{1/2}\right\} & \text{otherwise.} \end{cases}$$

and, with $k = 4$ and $\mathbb{E}_1(X) = \max\{n^2p^3, 1\}$,

$$\mathbb{P}(X_{C_4} \geq 2\lambda) \leq n^6 \begin{cases} \exp\left\{-cn^{1/4}p^{1/8}\right\} & \text{if } p \geq n^{-2/3} \\ \exp\left\{-cn^{1/2}p^{1/2}\right\} & \text{otherwise.} \end{cases}$$

3B, C) Kim–Vu. To apply Theorem 3B even to X_{K_3} we have to use induction. As we already know, one induction scheme leads to Theorem 3C. It is easily verified that for our three examples, and more generally for any balanced graph G, Theorem 3C applies with $k = e(G)$, $\ell = \lambda^{1/k}$ and $\mathcal{E}_j = (C\lambda)^{1-j/k}$, provided $\lambda^{1/k} \geq \log n$, which leads to

$$\mathbb{P}(X_G \geq 2\lambda) \leq n \exp\left\{-c\lambda^{1/e(G)}\right\} = n \exp\left\{-cn^{v(G)/e(G)}p\right\}.$$

For K_3, K_4 and C_4 this yields the exponents np, $n^{2/3}p$ and np, resp.

The induction used to prove Theorem 3C means in this case adding one edge at a time to G. However, there are better induction schemes. Vu [20, Theorem 3] has shown, using Theorem 3B and an induction adding vertices one by one, the better estimate

$$\mathbb{P}(X_G \geq 2\lambda) \leq n \exp\left\{-c\lambda^{1/(v(G)-1)}\right\}.$$

for every balanced G. (The induction hypothesis is actually stated more generally for extension counts, see [20] for details.) In our three cases, this yields the exponents $n^{3/2}p^{3/2}$, $n^{4/3}p^2$ and $n^{4/3}p^{4/3}$, resp.

4) Complement. This method works well for dense families $\mathcal{S}$, which is not the case here. There are $\Theta(n^6)$ 3-element subsets of $[n]^2$ and only $\Theta(n^3)$ triangles. Thus, $\eta = \Theta(n^{-3})$ and so, the exponent in Theorem 4 is $\eta^2 Np = O(p/n) = o(1)$. Things get only worse for $G = K_4$ and $G = C_4$.

5) Break-up. By Theorem 5, with $\Delta(L) \leq e(G)\Delta_H$ and Δ_H equal to the Lipschitz constants from 3.2.1 (Azuma), we obtain exponents n^2p^3, n^2p^6 and n^2p^4, resp.

6) Deletion. First, note that (ignoring constants in the exponent) Theorem 6B gives the same bounds as Theorem 5, and that Theorem 6C gives the same bounds as Theorem 3C. For K_3, K_4 and C_4 this yields the exponents n^2p^3, n^2p^6 and n^2p^4, resp., for 6B and np, $n^{2/3}p$ and np, resp., for 6C.

Recall that Theorem 6C is based on Theorem 6A and induction; in this case the induction is over the number of edges in the graph G. Just as for the Kim–Vu method, there are better induction schemes, and an induction over the number of vertices in G yields the better estimate

$$\mathbb{P}(X_G \geq 2\lambda) \leq n \exp\left\{-c\lambda^{1/(v(G)-1)}\right\}$$

for every balanced G, exactly as for Method 3. (Again the induction hypothesis is more general. See [10] for details and a general theorem.) In our three cases, this yields the exponents $n^{3/2}p^{3/2}$, $n^{4/3}p^2$ and $n^{4/3}p^{4/3}$, resp.

There are even more efficient ways to use Theorem 6A, although they so far require more effort and *ad hoc* arguments. First, Theorem 6A can be used directly for K_3. In this case $Y_{\{i\}} = \xi_i Z_i$, where $Z_i \in \mathrm{Bi}(n-1, p^2)$ is the number of paths of length 2 between the endpoints of edge i. The Chernoff bound (1) thus yields, see [9, Corollary 2.4], that if $\lambda/2r \geq 7np^2$, then

$$\mathbb{P}(Y_{\{i\}} > \lambda/2r) \leq \mathbb{P}(Z_i > \lambda/2r) \leq \exp\{-\lambda/2r\}.$$

Consequently Theorem 6A yields, for any r such that $\lambda/2r \geq 7np^2$,

$$\mathbb{P}(X_{K_3} \geq 2\lambda) \leq \exp\{-r/9\} + N\exp\{-\lambda/2r\}.$$

Choosing $r = c\sqrt{\lambda}$, we obtain

$$\mathbb{P}(X_{K_3} \geq 2\lambda) \leq n^2 \exp\left\{-cn^{3/2}p^{3/2}\right\},$$

as obtained above from [10]. This can be slightly improved by choosing instead $r = c\sqrt{\lambda \log n}$, and using (1) more carefully, leading to

$$\mathbb{P}(X_{K_3} \geq 2\lambda) \leq n^2 \exp\left\{-cn^{3/2}p^{3/2}\sqrt{\log n}\right\};$$

for details see [10]. Note also that choosing $r = \lambda/2n$ in Theorem 6A leads to Theorem 6B, and thus the exponent n^2p^3, which is better when $p \gg n^{-1/3}$.

For $G = K_4$, as shown in detail in [10], the term $\mathbb{P}(Y_{\{i\}} > \lambda/2r)$ can be estimated using a two-fold application of the Chernoff bound (1), which results in the estimates

$$\mathbb{P}(X_{K_4} \geq 2\lambda) \leq \begin{cases} n^2 \exp\left\{-cn^2p^3\right\} & \text{if } np^2 \leq 1, \\ n^2 \exp\left\{-cn^{4/3}p^{5/3}\right\} & \text{otherwise.} \end{cases}$$

(Actually, for $p \leq n^{-1/2-\varepsilon}$, the exponent can be improved to $cn^2p^3\sqrt{\log n}$ [10].) For $p \leq n^{-1/2}$, the exponent here differs by only a logarithmic factor from the exponent $n^2p^3 \log n$ in Vu's lower bound, so this estimate is essentially optimal. Here the advantage of the direct Chernoff application over induction is striking.

The case $G = C_4$ was not treated in [10], but a two-fold application of the Chernoff bound (1) as for K_4 yields an essentially optimal bound here too for a certain range of p. In this case $Y_{\{i\}} = \xi_i Z_i$, where Z_i is the number of paths of length 3 between the endpoints of edge i. Note that each such path is determined by the middle edge and its orientation in the path. Thus, if U is the set of vertices adjacent to at least one of the endpoints of i, and W is the number of edges with both endpoints in U, then $Y_{\{i\}} \leq 2W$, and thus, for any $a > 0$,

$$\mathbb{P}\Big(Y_{\{i\}} > \frac{\lambda}{2r}\Big) \leq \mathbb{P}\Big(W > \frac{\lambda}{4r}\Big) \leq \mathbb{P}(|U| > a) + \mathbb{P}\bigg(\mathrm{Bi}\bigg(\binom{\lceil a\rceil}{2}, p\bigg) > \frac{\lambda}{4r}\bigg),$$

where $|U| \in \mathrm{Bi}(n-2,\, 2p - p^2)$.

If we choose r and a such that $a \geq 14np > 7\,\mathbb{E}\,|U|$ and $\lambda/4r > 7\frac{a^2}{2}p$, then by (1), see [9, Corollary 2.4], this yields

$$\mathbb{P}(Y_{\{i\}} > \lambda/2r) \leq \exp(-a) + \exp(-\lambda/4r)$$

and hence Theorem 6A yields

$$\mathbb{P}(X_{C_4} \geq 2\lambda) \leq e^{-cr} + \binom{n}{2}\Big(e^{-a} + e^{-\lambda/4r}\Big).$$

If $p \geq n^{-2/3}$, we choose (for a small c) $a = r = cn^{4/3}p$ and obtain

$$\mathbb{P}(X_{C_4} \geq 2\lambda) \leq n^2 \exp\left\{-cn^{4/3}p\right\}, \qquad p \geq n^{-2/3}.$$

For $p \le n^{-2/3}$, we choose $a = cnp^{1/2}$ and $r = cn^2p^2$, and obtain

$$\mathbb{P}(X_{C_4} \ge 2\lambda) \le n^2 \exp\left\{-cn^2p^2\right\}, \qquad p \le n^{-2/3}.$$

Just as for K_4, for small p (here $p \le n^{-2/3}$), the obtained exponent matches Vu's lower bound up to a logarithmic factor.

7) Approximation. As we already know, Theorem 7 is never better than Theorem 6A (for ρ constant as here, at least), see Remark 2. Here they actually sometimes yield the same result. Indeed, for K_3 we have $R = \lambda \cdot 3\Delta_1' p^2 = \Theta(n^4p^5)$, since $\Delta_1 = n-2$ and $\Delta_2 = 1$, and thus $\Delta_2' = 0$. We apply the Chernoff bound (1) to Y_1^* as for Method 6A. Since we now need $r \ge R$, we see that the essentially optimal choice $r = c\sqrt{\lambda} = \Theta(n^{3/2}p^{3/2})$ used above for Method 6A now is allowed if $p \le cn^{-5/7}$; otherwise, we must be content with the choice $r = R$. This yields the following estimates.

$$\mathbb{P}(X_{K_3} \ge 2\lambda) \le \begin{cases} 2\exp\left\{-cn^{3/2}p^{3/2}\right\} & \text{if } p \le n^{-5/7}, \\ \exp\left\{-c/(np^2)\right\} & \text{otherwise.} \end{cases}$$

For $G = K_4$, Method 7 fares worse. Since $\Delta_3 = n-3$, we have $R \ge np\lambda \gg \lambda$, provided $\lambda \ge 1$ to avoid trivialities, and thus $r \ge R$ implies $\lambda/r \to 0$ and we do not get any meaningful bound.

For $G = C_4$ we have, assuming $\lambda \ge 1$ and thus $np \ge 1$, $R = \Theta(\lambda n^2p^3) = \Theta(n^6p^7)$. If $p \le n^{-4/5}$, this allows the essentially optimal choice $r = cn^2p^2$ used for Method 6A above, and the same double application of the Chernoff bound yields again the optimal bound

$$\mathbb{P}(X_{C_4} \ge 2\lambda) \le n^2 \exp\left\{-cn^2p^2\right\}, \qquad p \le n^{-4/5},$$

although now for a smaller range of p only. For $n^{-4/5} \le p \le cn^{-2/3}$ we have to take $r = R$, and can estimate the tail of $Y_{\{i\}}$ as for Method 6A, now using $a = c/np^2$. We obtain

$$\mathbb{P}(X_{C_4} \ge 2\lambda) \le n^2 \exp\left\{-c/n^2p^3\right\}, \qquad p \ge n^{-4/5},$$

which is meaningful for $p \ll n^{-2/3}$.

Summary of subsection 3.2. The results of this subsection are summarized in Table 2 and in Figures 1–3. A point (x, y) in the figures signifies a bound of the type $c_1 \exp(-c_2 n^y)$ when $p = n^{-x}$; in other words, the figures plot the asymptotic value of $\log|\log(\text{bound})|/\log n$ as a function of $|\log p|/\log n$. Thus, the bigger y, the better bound. Note that along the x-axis, p decreases from constant values at the left end down to $1/n$ at the right end. The dotted line shows Vu's lower bound.

	K_3	K_4	C_4
1) Azuma	n^2p^6	n^2p^{12}	n^2p^8
2) Talagrand	n^2p^5	n^2p^{11}	n^2p^7
3A) Kim–Vu A	$\min(n^{1/3}p^{1/6},\, n^{1/2}p^{1/2})$	$\min(n^{1/6}p^{1/12},\, n^{1/3}p^{1/2})$	$\min(n^{1/4}p^{1/8},\, n^{1/2}p^{1/2})$
3B) Kim–Vu B [20]	$n^{3/2}p^{3/2}$	$n^{4/3}p^2$	$n^{4/3}p^{4/3}$
3C) (Kim–)Vu C	np	$n^{2/3}p$	np
4) Complement	—	—	—
5) Break-up	n^2p^3	n^2p^6	n^2p^4
6A) Deletion A [10]	$\max(n^{3/2}p^{3/2},\, n^2p^3)$	$\min(n^2p^3,\, n^{4/3}p^{5/3})$	$\min(n^2p^2,\, n^{4/3}p)$
6B) Deletion B	n^2p^3	n^2p^6	n^2p^4
6C) Deletion C	np	$n^{2/3}p$	np
7) Approximation	$\min(n^{3/2}p^{3/2},\, 1/(np^2))$	—	$\min(n^2p^2,\, 1/n^2p^3)$
Vu's lower bound	n^2p^2	n^2p^3	n^2p^2

Table 2: Exponents for upper tails in the small subgraphs problem, ignoring logarithmic factors.

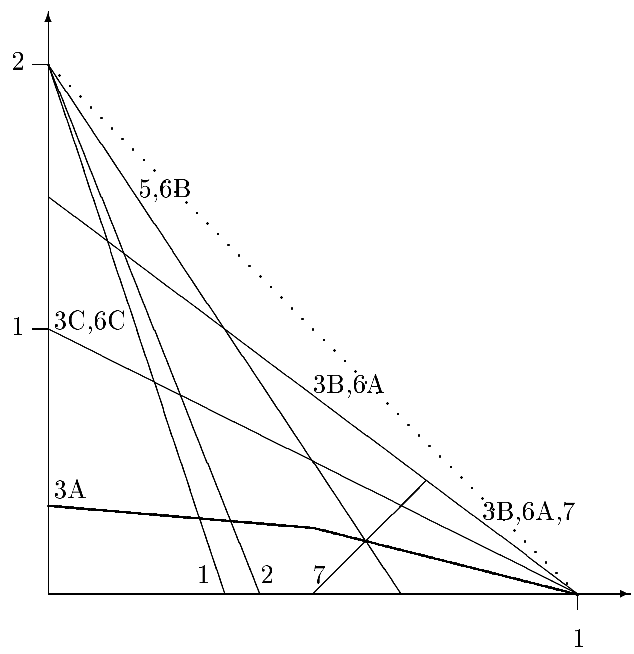

Figure 1: K_3

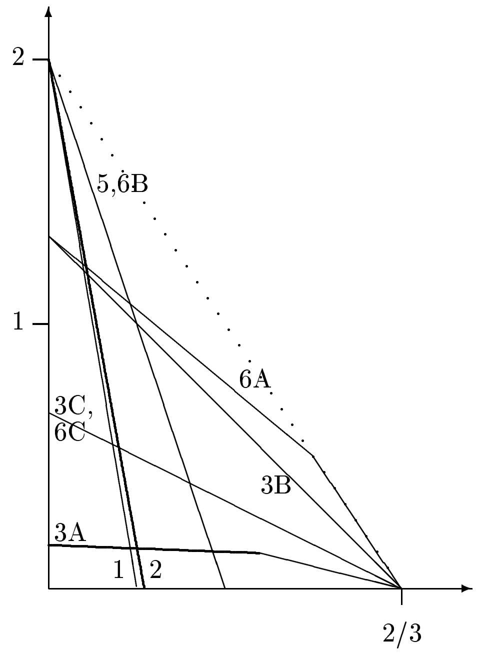

Figure 2: K_4

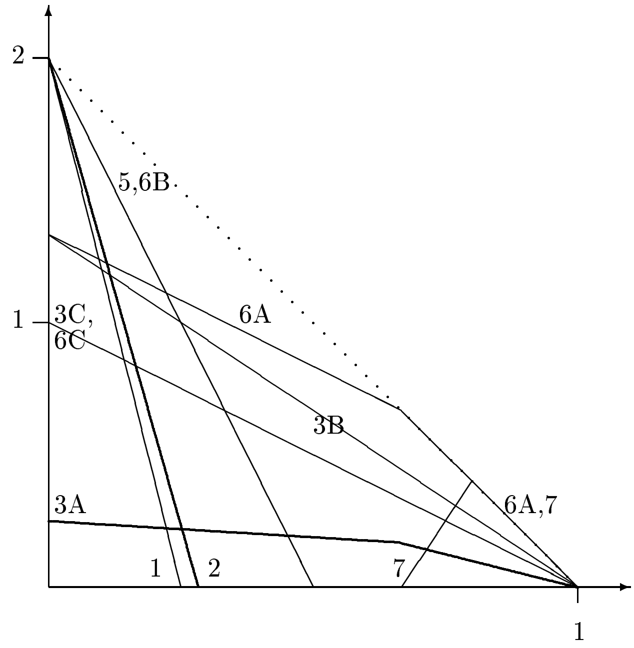

Figure 3: C_4

We leave the direct comparison of the results obtained by various methods to the anxious reader. Instead, we check how close these estimates get to the lower bound mentioned earlier. For $G = K_3$, none of them achieves it. The nearest are the bounds obtained by Methods 5 and 6 for rather dense graphs and Methods 3, 6 and 7 for sparser graphs. In the cases $G = K_4$ and $G = C_4$, quite surprisingly, Methods 6 and 7, together with the double Chernoff argument used above, both achieve the lower bound in some range of $p = p(n)$. (This can also be achieved by Method 3 [Vu, personal communication].)

Note also that the simple combinatorial Method 5 (Break-up) again beats both Azuma and Talagrand, and together with 6B is the best for rather dense graphs.

3.3 How many pairs support tepees?

Our last example is very special, but still worth mentioning. It shows the strength of Lemma 1 which gave rise to the Deletion Method 6, but so far has not been illustrated by any direct application. In the following problem we would like to obtain an upper bound on the upper tail of X_{C_4} for $p = \Theta(n^{-1/2})$, which is of the order $\exp(-\Theta(n))$, just as Vu's lower bound (ignore the logarithm, again). However, none of the Methods 1–7 gives that, and we do not know whether such a bound holds or not. Nevertheless, it turns out that if we apply Lemma 1 directly, and instead of proving a genuine bound on the tail allow ourselves to discard some edges, we will achieve our task.

The problem itself arose in the study of the width of the threshold for a Ramsey property of random graphs [4]. The solution, however, has originated in [15], where the Deletion Method traces back its roots.

Let G be a graph. The base of G, denoted by $Base(G)$, is the set of all pairs of vertices of G with a common neighbor in G. Thus, $Base(G)$ is a graph with the vertex set $V(G)$, but typically not a subgraph of G. It is a subgraph of G^2, though.

In [4] the following lemma is needed.

Lemma 4. *For every $c > 0$ and $a > 0$ there exists a' such that a.a.s. for any (spanning) subgraph F of the random graph $\mathbb{G}(n, cn^{-1/2})$ with $|E(F)| > acn^{3/2}$, the graph $Base(F)$ contains at least $a'n^3$ triangles.*

In the proof which we only outline here (for details see [5]), an application of a sparse regularity lemma (see [9, Lemma 8.19, page 216]) to F yields, for a suitably chosen $\rho > 0$, the existence of a pair of disjoint subsets of vertices, U and V with $an/2 \le |U| \le an$, $|V| = n/k$, where $k \le K = K(\rho, a)$, such that for every $W \subset V$, $|W| = \rho|V|$, all but at most $\rho|U|$ vertices of U have each at least

$$(1-\rho')ap|W|/2 = (1-\rho')ac\rho\sqrt{n}/2k \tag{18}$$

neighbors in W, where $\rho' = \rho'(\rho, a) > \rho$.

The next and last step is to show that a.a.s. every such W induces a subgraph B_W of $Base(F)[V]$ with at least $a''\binom{|W|}{2}$ edges, where $a'' = \min\{a^3c^2/20, a^4/400\}$ does not depend on ρ. This will easily imply (see [15, Lemma 2]) that $Base(F)[V]$ itself contains at least $a'''|V|^3$ triangles, which proves Lemma 4 with $a' = a'''/K^3$. The same Lemma 2 of [15] determines $\rho = \rho(a'')$.

We will underestimate the edges of B_W by counting only pairs of vertices of W with a common neighbor in U. So, let $H = F[U, W]$, i.e. H is the bipartite subgraph of F with the bipartition sets U and W and all the edges of F with one endpoint in U and the other in W. Then $B_W \supseteq Base(H)[W]$.

How to count the edges of $Base(H)[W]$? Let us number the pairs of vertices in W by $1, \dots, \binom{|W|}{2}$ and denote by x_i the number of paths of length two in H connecting the vertices of the i-th pair (such paths are called in [4] tepees). Set L for the number of those $i \in \{1, \dots, \binom{|W|}{2}\}$ for which $x_i > 0$, i.e. for the sought number of pairs of vertices of W with a common neighbor in U. Assume that $x_1, \dots, x_L > 0$.

Observe that, denoting by d_u the degree of vertex $u \in U$ in H,

$$\sum_{i=1}^{L} x_i = \sum_{u \in U} \binom{d_u}{2} \quad \text{and} \quad \sum_{i=1}^{L} \binom{x_i}{2} = X_{C_4}(H),$$

where $X_{C_4}(H)$ is the number of copies of C_4 in H.

If $L \ge \sum_{i=1}^{L} x_i/2$, then by (18) (for ρ small)

$$L \ge \frac{1}{2}\sum_{u \in U} \binom{d_u}{2} \ge \frac{1}{33}(1-\rho')^3 a^3 c^2 \rho^2 (n/k)^2 \ge \frac{1}{20} a^3 c^2 \binom{|W|}{2}.$$

Otherwise, by the Cauchy–Schwarz inequality,

$$\sum_i \binom{x_i}{2} \ge L\binom{\sum_i x_i/L}{2} \ge \frac{(\sum_i x_i)^2}{4L},$$

and a decent upper bound on $X_{C_4}(H)$ will give a lower bound on L and thus on the number of edges of $Base(H)[W]$.

To get back to the genuine random graph not affected by a malicious choice of the subgraph H, we may bound $X_{C_4}(H)$ by $X_{C_4}(U,W)$ – the number of copies of C_4 contained in the bipartite subgraph $\mathbb{G}(n,p)[U,W]$ of $\mathbb{G}(n,p)$ spanned by U and W. Obviously, the expectation of $X_{C_4}(U,W)$ is equal to $\binom{|U|}{2}\binom{|W|}{2}p^4 = \Theta(n^4p^4) = \Theta(n^2)$.

To accommodate the number of choices of the sets U and W, which is bounded by 4^n, we would need an upper tail estimate on $X_{C_4}(U,W)$ which is close to the lower bound established by Vu. None of our methods provides such a bound. Fortunately, we may use Lemma 1.

Define a random variable $Y_{-k}(U,W) = \min_{E_0\subset E(\mathbb{G}(n,p)),\,|E_0|\le k} X_{-E_0}(U,W)$, where $X_{-E_0}(U,W)$ is the number of copies of C_4 in the subgraph $G(n,p)[U,W]-E_0$. It follows quickly from Lemma 1 that a.a.s. for all U and W we have

$$Y_{-n\log n}(U,W) < 2\,\mathbb{E}\,X_{C_4}(U,W) = 2\binom{|U|}{2}\binom{|W|}{2}p^4 < \frac{a^2\rho^2c^4n^2}{2k^2}.$$

The bottom line of this argument is that even after deleting the edges of $E_0 = E_0(U,W)$, still all but, say, at most $2\rho|U|$ vertices of U have each at least $(1-2\rho')ap|W|/2 = (1-2\rho')ac\rho\sqrt{n}/2k$ neighbors in W. Indeed, as $|E_0| = n\log n$, at most, say, $n^{2/3}$ vertices of U may have each more than $n^{1/3}\log n$ edges of E_0 incident to them.

Let d_u^0 be the degree of u in $H-E_0$, $u\in U$, and similarly, let x_i^0 and L^0 be modifications of the previously introduced quantities. Note that $|E(Base(H)[W])| \ge L^0$ and consider two cases. If $L^0 \ge \sum_{i=1}^{L^0} x_i^0/2$, then, as before, (for ρ small)

$$L^0 \ge \frac{1}{2}\sum_{u\in U}\binom{d_u^0}{2} \ge \frac{1}{20}a^3c^2\binom{|W|}{2}.$$

If, on the other hand, $L^0 \le \sum_{i=1}^{L^0} x_i^0/2$, then

$$\sum_i \binom{x_i^0}{2} \ge L^0\binom{\sum_i x_i^0/L^0}{2} \ge \frac{(\sum_i x_i^0)^2}{4L^0}$$

and, so

$$L^0 \ge \frac{\frac{1}{400}a^6c^4\rho^4(n/k)^4}{2a^2c^4\rho^2(n/k)^2} \ge \frac{a^4}{400}\binom{|W|}{2},$$

as required.

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, and Edyta Szymańska for providing her desk and her whiteboard. We further thank Van Vu for sending us unpublished preprints and drafts.

References

[1] G. Bennett, Probability inequalities for the sum of independent random variables. *J. Amer. Statist. Assoc.* **57** (1962), 33–45.

[2] S. Boucheron, G. Lugosi & P. Massart, A sharp concentration inequality with applications. *Random Struct. Alg.* **16** (2000), 277–292.

[3] P. Erdős & A. Rényi, On the evolution of random graphs. *Publ. Math. Inst. Hungar. Acad. Sci.* **5** (1960), 17–61.

[4] E. Friedgut, V. Rödl, A. Ruciński & P. Tetali, Sharp thresholds for Ramsey properties. In preparation.

[5] E. Friedgut, Y. Kohayakawa, V. Rödl, A. Ruciński & P. Tetali, Ramsey games against one-armed bandit. In preparation.

[6] A. Hajnal & E. Szemerédi, Proof of a conjecture of Erdős. *Combinatorial Theory and its Applications, Vol. II*, Editors P. Erdős, A. Rényi & V.T. Sós, Colloq. Math. Soc. János Bolyai 4, North-Holland, Amsterdam, (1970), 601–623.

[7] W. Hoeffding, Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* **58** (1963), 13–30.

[8] S. Janson, Poisson approximation for large deviations. *Random Struct. Alg.* **1** (1990), 221–230.

[9] S. Janson, T. Łuczak & A. Ruciński, *Random Graphs.* Wiley, New York, 2000.

[10] S. Janson & A. Ruciński, The deletion method for upper tail estimates, Technical report U.U.D.M. 2000:28, Uppsala.

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

[12] C. McDiarmid, On the method of bounded differences. *Surveys in Combinatorics (Proceedings, Norwich 1989)*, London Math. Soc. Lecture Note Ser. 141, Cambridge Univ. Press, Cambridge 1989, pp. 148–188.

[13] C. McDiarmid, Concentration. *Probabilistic Methods for Algorithmic Discrete Mathematics*, eds. M. Habib, C. McDiarmid, J. Ramirez and B. Reed, Springer-Verlag, Berlin, 1998, pp. 195–248.

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

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

[16] J. Spencer, Counting Extensions. *J. Combin. Th. A* **55** (1990), 247–255.

[17] M. Talagrand, Concentration of measure and isoperimetric inequalities in product spaces. *Inst. Hautes Études Sci. Publ. Math.* **81** (1995), 73–205.

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

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

[20] 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.*

[21] V.H. Vu, Concentration of non-Lipschitz functions and applications. To appear in *Random Struct. Alg.*