# UPPER TAILS FOR COUNTING OBJECTS IN RANDOMLY INDUCED SUBHYPERGRAPHS AND ROOTED RANDOM GRAPHS

SVANTE JANSON AND ANDRZEJ RUCIŃSKI

Abstract. General upper tail estimates are given for counting edges in a random induced subhypergraph of a fixed hypergraph $\mathcal{H}$, with an easy proof by estimating the moments. As an application we consider the numbers of arithmetic progressions and Schur triples in random subsets of integers. In the second part of the paper we return to the subgraph counts in random graphs and provide upper tail estimates in the rooted case.

## 1. Introduction

Consider a finite sum of dependent random variables of the following form. Let $\Gamma$ be a finite ground set and let $\mathcal{S}$ be a family of its subsets. Let $\Gamma_p$ be a random, binomial subset of $\Gamma$ which independently includes each element of $\Gamma$ with probability $p$. Finally, for each $S \in \mathcal{S}$, let $I_S$ be the indicator random variable of the event $\{S \subseteq \Gamma_p\}$. Then $X = X(\Gamma, \mathcal{S}, p) = \sum_{S\in\mathcal{S}} I_S$ counts the number of members of the family $\mathcal{S}$ contained in a random subset $\Gamma_p$. A lot of research has been devoted to the study of the asymptotic distribution of $X$ when the order $N = |\Gamma|$ grows to $\infty$ and $p = p(N)$, both in a general setting and for particular instances, most notably for random graphs, see [7].

One feature which received a lot of attention is the rate of decay of the tails of $X$, the lower tail $\mathbb{P}(X \le t\,\mathbb{E}\,X)$ for $0 < t < 1$, and the upper tail $\mathbb{P}(X \ge t\,\mathbb{E}\,X)$ for $t > 1$. Good estimates for the lower tail follow from the FKG inequality (lower bound) and Janson's inequality (upper bound), see [7], Section 2.2. Often, these two bounds asymptotically match under some restrictions on the dependencies among the summands $I_S$. This is, in particular, the case of subgraph counts in random graphs, see [7], Section 3.1.

The upper tails tend to be harder to analyze. Some ad hoc results can be found in [7], [10], [5], [6], among others. For the subgraph count problem a quite satisfactory and complete result has been obtained in [4], where

*Date*: May 7, 2009.

2000 *Mathematics Subject Classification.* 60C05; 05C80, 05C65.

Second author supported by Polish grant N201036 32/2546. Research was performed while the authors visited Institut Mittag-Leffler in Djursholm, Sweden, during the program 'Discrete Probability', 2009.

the logarithms of the upper and lower bound on $\mathbb{P}(X \geq t\,\mathbb{E}\,X)$ are of the same order of magnitude except for a logarithmic term. A generalization to random hypergraphs can be found in [1].

This paper can be viewed as a follow-up paper to [4]. Using the proof techniques developed therein, those results are extended in two directions. First, we return to the more general model of set systems (or hypergraphs) and obtain some straightforward estimates for the upper tail of $X$, covering, in particular, the number of arithmetic progressions of given length in a random subset of integers. Then, we return to the subgraph counts to study the rooted version of the problem, only to discover some unexpected features there.

## 2. Counting edges of randomly induced subhypergraphs

Let $\mathcal{H}$ be a $k$-uniform hypergraph on a vertex set $\Gamma$ with $|\Gamma| = N$ and with $|\mathcal{H}| = aN^q$ edges, where $a = a(N) > 0$ and $0 < q \leq k$. Consider a random, binomial subset $\Gamma_p$ of $\Gamma$, where $0 < p = p(N) < 1$, and the random variable $X = |\mathcal{H}[\Gamma_p]|$ counting the edges of $\mathcal{H}$ that are entirely present in $\Gamma_p$. Note that

$$\mu := \mathbb{E}\,X = |\mathcal{H}|p^k = aN^q p^k.$$

For $j = 0, 1, \dots, k$, let

$$\Delta_j = \max_{S \in \binom{\Gamma}{j}} |\{T \in \mathcal{H} : T \supseteq S\}|,$$

i.e., the maximum number of edges that contain $j$ given vertices.

**Theorem 2.1.** *Let $q$ be an integer, $1 \leq q \leq k$, and let $a_0 > 0$ and $t > 1$ be real numbers. There exists a constant $c = c(q, a_0, t)$ such that if $\mathcal{H}$ satisfies the following four conditions:*

(i) $a(N) = |\mathcal{H}|/N^q \geq a_0$,
(ii) *for all $j \leq q$ we have $\Delta_j = O(N^{q-j})$,*
(iii) *for all $j > q$ we have $\Delta_j = O(1)$,*
(iv) *there exists $C > 0$ and $\Gamma_0 \subseteq \Gamma$ such that $|\Gamma_0| \leq C\mu^{1/q}$ and $|\mathcal{H}[\Gamma_0]| \geq t\mu$,*

*then, with $X = |\mathcal{H}[\Gamma_p]|$,*

$$p^{C\mu^{1/q}} = \exp\{-C\mu^{1/q}\log(1/p)\} \leq \mathbb{P}(X \geq t\mu) \leq \exp\{-c\mu^{1/q}\}.$$

Before giving the proof, we make some comments.

- The two exponents are of the same order of magnitude except for the logarithmic term $\log(1/p)$; this inaccuracy disappears obviously for $p$ constant.
- Note that $\mathbb{P}(X \geq t\mu) > 0 \iff t\mu \leq |\mathcal{H}| \iff tp^k \leq 1$, so the theorem is interesting for $t \leq p^{-k}$ only. (For larger $t$, $\mathbb{P}(X \geq t\mu) = 0$ so the lower bound fails, while the upper bound is trivial; further, (iv) fails.)

- Condition (iii) is redundant, since it follows from (ii) with $j = q$, but we prefer to include it explicitly for emphasis, and for comparison with Theorem 2.2 which allows non-integer values of $q$ (note that for non-integer $q$, (iii) does not follow from (ii)).
- As we will see in the proof, the upper bound follows only from conditions (i)–(iii), while the lower bound is a consequence of condition (iv) alone.

*Proof.* Take $C$ and $\Gamma_0$ as in assumption (iv). We have

$$\mathbb{P}(X \geq t\mu) \geq \mathbb{P}(\Gamma_p \supseteq \Gamma_0) = p^{|\Gamma_0|},$$

which proves the lower bound.

For the upper bound, we use the same approach as in [4]. By Markov's inequality, for every $m$ we have

$$\mathbb{P}(X \geq t\mu) \leq \frac{\mathbb{E}\, X^m}{t^m \mu^m}.$$

It remains to show that for a sufficiently small $c_1 = c_1(q, a_0, t)$ and $m = \lceil c_1 \mu^{1/q} \rceil$ we have, say, $\mathbb{E}\, X^m \leq t^{m/2}\mu^m$.

Having chosen $m-1$ (not necessarily distinct) edges $E_1, \dots, E_{m-1}$ of $\mathcal{H}$, let $N_j$ be the number of edges $E_m$ such that $\left|E_m \cap \bigcup_{i=1}^{m-1} E_i\right| = j$, and let $N_{\geq j} = \sum_{k \geq j} N_k$. We estimate these numbers as follows: For $j = 0$,

$$N_0 \leq N_{\geq 0} = |\mathcal{H}|. \tag{2.1}$$

For $1 \leq j \leq q$, by (ii),

$$N_j \leq N_{\geq j} = O(m^j \Delta_j) = O(m^j N^{q-j}), \tag{2.2}$$

since if $\left|E_m \cap \bigcup_{i=1}^{m-1} E_i\right| \geq j$, then there exists a set $A \subseteq \bigcup_{i=1}^{m-1} E_i$ with $|A| = j$ and $E_m \supseteq A$, and there are $O(m^j)$ such sets $A$, and at most $\Delta_j$ edges $E_m$ for each $A$. For $j > q$ we obtain

$$N_j \leq N_{\geq q} = O(m^q) \tag{2.3}$$

from (2.2) (with $j = q$).

Arguing as in [4] we have from (2.1)–(2.3), by induction on $m$,

$$\begin{aligned}\mathbb{E}\, X^m &\leq \mu \left( |\mathcal{H}|p^k + \sum_{j=1}^{q} O(m^j N^{q-j}) p^{k-j} + \sum_{j=q+1}^{k} O(m^q) p^{k-j} \right)^{m-1} \\ &= \mu^m \left( 1 + O(a^{-1}) \sum_{j=1}^{q} \left( \frac{m}{Np} \right)^j + O(1)\frac{m^q}{\mu} \right)^{m-1}\end{aligned}$$

for every $m \geq 1$. Now choose $m = \lceil c_1 \mu^{1/q} \rceil \geq 1$, as said above. If $m \geq 2$, then $m/(Np) \leq 2c_1 \mu^{1/q}/(Np) = 2c_1 a^{1/q} p^{k/q-1} \leq 2c_1 a^{1/q}$, and thus, using (i), the term in parenthesis in the last line can be made arbitrarily close to 1 for all $m \geq 2$ by choosing $c_1 > 0$ small enough; in particular, it can be made less than $t^{1/2}$. Hence, for the chosen $m$, $\mathbb{E}\, X^m \le t^{m/2}\mu^m$ if $m \ge 2$, and trivially if $m = 1$ too. This completes the proof. $\square$

In the case of non-integer $q$, the upper bound gets further away from the lower bound. Indeed, we then have the following result.

**Theorem 2.2.** *Let $q$, $a_0$ and $t$ be real numbers, with $0 < q \le k$, $a_0 > 0$ and $t > 1$. There exists a constant $c = c(q, a_0, t)$ such that under the same assumptions* (i)–(iv) *as in Theorem 2.1,*

$$\mathbb{P}(X \ge t\mu) \le \exp\left\{-c \max\left(\mu^{1/q} p^{k(1/\lfloor q\rfloor - 1/q)}, \mu^{1/\lceil q\rceil}\right)\right\}$$

*and*

$$\mathbb{P}(X \ge t\mu) \ge p^{C\mu^{1/q}} = \exp\{-C\mu^{1/q}\log(1/p)\}.$$

*Proof.* The only difference in the proof is when we bound $N_j$ to estimate $\mathbb{E}\, X^m$. Namely, for $j \ge \lceil q\rceil$, we either use $N_j \le N_{\ge \lfloor q\rfloor} = O(m^{\lfloor q\rfloor} N^{q-\lfloor q\rfloor})$, or $N_j \le N_{\ge\lceil q\rceil} = O(m^{\lceil q\rceil})$. We then choose

$$m = \lceil c_1 \max\bigl(\mu^{1/q} p^{k(1/\lfloor q\rfloor - 1/q)}, \mu^{1/\lceil q\rceil}\bigr)\rceil$$

for a small constant $c_1$. (We may assume $\mu \ge 1$, since otherwise $m = 1$ and, recalling that $t > 1$, the estimate $\mathbb{E}\, X \le t^{1/2}\mu$ is trivial.) $\square$

2.1. **Integer solutions of linear homogeneous systems.** For an $l \times k$ integer matrix $A$, where $l < k$, assume that every $l \times l$ submatrix $B$ of $A$ has full rank $r(B) = l = r(A)$. Consider the system of homogeneous linear equations $Ax = 0$, where $x = (x_1, \dots, x_k)$ is a column vector and 0 is a column vector of dimension $l$. We assume also that there exists a distinct-valued positive integer solution of $Ax = 0$. These assumptions seem to be quite restrictive, but, in fact, we cover at least one important case: the arithmetic progressions of length $k$ which can be viewed as distinct-valued solutions to a system of $l = k - 2$ equations.

Let $\Gamma = [N] := \{1, 2, \dots, N\}$ and $0 < p = p(N) < 1$. Then $\Gamma_p$ is a random subset of the first $N$ integers with density $p$. Define a $k$-uniform hypergraph $\mathcal{H}_A = \mathcal{H}_A(N)$ as the family of all solution sets $\{x_1, \dots, x_k\}$ of the system $Ax = 0$ with $x_i$ distinct and in $[N]$. Let us check that for some $a_0, q$, and $C$ the assumptions (i)–(iv) of Theorem 2.1 hold, at least in the interesting case $\mu = |\mathcal{H}_A| p^k \ge 1$ and $t\mu \le |\mathcal{H}_A|$, which can be equivalently restated as

$$\mu \ge 1 \quad \text{and} \quad t \le p^{-k}. \tag{2.4}$$

Set $q = k - l$.

(i), (iv): We will show that there exists $a_0 > 0$ such that for sufficiently large $m \le N$ we have

$$|\mathcal{H}_A(m)| \ge a_0 m^q. \tag{2.5}$$

Taking $m = N$ in (2.5) we obtain $|\mathcal{H}_A| \ge a_0 N^q$, which is (i). Taking $m = \min\bigl(\lceil (t a_0^{-1}\mu)^{1/q}\rceil, N\bigr)$ in (2.5) and $\Gamma_0 = [m]$ we obtain (iv) with $C = 2(t a_0^{-1})^{1/q}$, using the assumptions in (2.4).

Let $\mathbf{x}_0 \in Z^k$ be a positive integer solution of $Ax = 0$. Let $M_0$ by the largest of its coefficients $x_{01}, \dots, x_{0k}$. Let $\mathbf{x}_1, \dots, \mathbf{x}_q$ be $q$ linearly independent integer solutions of $Ax = 0$. (There exist $q$ linearly independent rational solutions, and we may multiply these by their common denominators and thus assume that they are integer solutions.) Let $M$ be the maximum of the absolute values of the coefficients in $\mathbf{x}_1, \dots, \mathbf{x}_q$.

Given $m$, let $d := \lfloor m/(M_0 + 1) \rfloor$. For any integers $a_1, \dots, a_q$, the sum $d\mathbf{x}_0 + \sum_{i=1}^{q} a_i \mathbf{x}_i$ yields an integer solution of $Ax = 0$, and these solutions are all distinct. If further $|a_i| < d/(2qM)$ for all $i$, this solution has all coefficients positive, less than $m$, and distinct. The number of these solutions is $\Theta(d^q) = \Theta(m^q)$. Hence, (2.5) holds.

(ii), (iii): By elementary algebraic properties of systems of linear equations, every system $By = c$, where $B$ is an integer $l \times h$ matrix, has no more than $N^{h-r(B)}$ solutions in $[N]$. Thus, $\Delta_0 = |\mathcal{H}_A| \le N^{k-l} = N^q$. For every subset $J$ of the columns of $A$, define $A_J$ as the submatrix obtained from $A$ by removing the columns in $J$. This means that when we fix values of some $j$ variables, then the obtained system of equations is of the form $By = c$, where $y$ consists of the remaining unknowns, $B = A_J$, and $J$ is the set of columns of $A$ corresponding to the fixed variables. Hence, the number of solutions with $j$ given elements corresponding to the given columns $J$ is at most $N^{k-j-r(A_J)}$. Now, for all $j \le q = k - l$, if $|J| = j$ then, by our assumption on $A$, $r(A_J) = l$, so (summing over $J$) $\Delta_j = O(N^{k-j-l}) = O(N^{q-l})$. On the other hand, if $j > k - l$ then $r(A_J) = k - j$, so $\Delta_j = O(N^0) = O(1)$.

Hence, given (2.4), Theorem 2.1 applies for such $\mathcal{H}_A$ with $q = k - l$ and $\mu = \Theta(N^{k-l} p^k)$.

**Example 2.3.** In particular, we obtain quite sharp estimates for the tails of the numbers of arithmetic progressions of length $k$ in $[N]_p$. Indeed, they are given by the system $x_i - 2x_{i+1} + x_{i+2} = 0$, $i = 1, \dots, k-2$. It is easy to check that for $l = k - 2$ every $l \times l$ submatrix has full rank, and we have the following result.

**Corollary 2.4.** *Let $X$ be the number of arithmetic progressions of length $k$ in $[N]_p$, $k \ge 3$, and let $\mu$, $t > 1$, and $p$ satisfy* (2.4)*. Then there exist $c, C > 0$ such that*

$$p^{CNp^{k/2}} = \exp\{-CNp^{k/2} \log(1/p)\} \le \mathbb{P}(X \ge t\mu) \le \exp\{-cNp^{k/2}\}. \qquad \square$$

**Example 2.5.** A Schur triple is a triple $\{x, y, z\}$ of positive integers such that $x + y = z$, $x \ne y$. In this case we have $k = 3$, $l = 1$ and so, $q = 2$.

**Corollary 2.6.** *Let $X$ be the number of Schur triples in $[N]_p$, and let $\mu$, $t > 1$, and $p$ satisfy* (2.4) *with $k = 3$. Then there exist $c, C > 0$ such that*

$$p^{CNp^{3/2}} = \exp\{-CNp^{3/2} \log(1/p)\} \le \mathbb{P}(X \ge t\mu) \le \exp\{-cNp^{3/2}\}. \qquad \square$$

**Remark 2.7.** Arithmetic progressions are partition regular, a name introduced by Rado for all linear systems the solutions of which satisfy theorems similar to the van der Waerden theorem. But, in addition, they are also density regular, which means that every subset of integers of positive density contains them (Szemerédi's theorem). Partition properties of random subsets of integers with respect to density regular systems were studied in [9]. Schur triples form an example of partition regular but not density regular linear system. Partition properties of random subsets of integers with respect to Schur triples were studied in [3].

**Remark 2.8.** We have here treated the set of solutions $x$ to $Ax = 0$ as a hypergraph, i.e., we have treated the solutions $x$ as $k$-sets rather than $k$-vectors. This is fine for the examples of arithmetic progressions and Schur triples treated above, but in general it may be more natural to regard the solutions $x$ as vectors (or, equivalently, sequences) in $[N]^k$, rather than as sets. We then define $\mathcal{H}_A$ as the subset $\{x : Ax = 0\}$ of $[N]^k$. In this way, we distinguish between solutions that are permutations of each other (for example, $(x, y, z)$ and $(y, x, z)$ in the Schur triple case), and we allow repeated values.

It is possible to prove a version of Theorem 2.1 for this case, using essentially the same proof, but the possibility of repeated elements of $\Gamma = [N]$ complicates the conditions; we now need bounds on the number of vectors in $\mathcal{H}_A$ that have $j$ coordinates fixed, and at most $\ell$ distinct values of the other coordinates. We omit the details.

### 2.2. **Further examples and remarks.**

**Example 2.9.** In the dense case, that is, when $q = k$, assumption (iv) holds trivially by averaging over all subsets $\Gamma_0$ of a suitable size, provided the necessary condition $t \leq p^{-k}$ is satisfied, but this result has been known already (cf. [5] and [6]). In particular, this case covers the number of matchings of size $k$ in a random $r$-uniform hypergraph $G^{(r)}(n, p)$, by considering a $k$-uniform hypergraph $\mathcal{H}$ where the vertices are the edges of the complete $r$-uniform hypergraph $K_n^{(r)}$ and the edges are the matchings of size $k$ in $K_n^{(r)}$. Then the assumptions of Theorem 2.1 hold with $q = k$.

**Remark 2.10.** It can be very hard to improve upon Theorem 2.2, because it contains the triangle count problem from [4]. Indeed, with $\Gamma = \binom{[n]}{2}$ and $\mathcal{H}$ being the family of the edge sets of all triangles in $K_n$, we have $N = \binom{n}{2}$ and $|\mathcal{H}| = \Theta(n^3) = \Theta(N^{3/2})$, so $q = 3/2$. To get the result from [4], we would need to improve the upper bound, but this seems to be impossible without "seeing" the vertices of the random graph.

## 3. Rooted subgraphs of random graphs

A *rooted graph* $(R, G)$ is a graph $G$ with a fixed independent set $R$; we also say that the graph is *rooted at* $R$. (For simplicity, we sometimes use $G$ to denote the rooted graph $(R, G)$ when $R$ is clear from the context.) Counting rooted subgraphs of a random graph $G(n, p)$ with a fixed set $R$ of roots plays an important role in studying the so called extension statements and 0–1 laws in random graphs, see, e.g., [7, Sections 3.4 and 10.2]. Another application can be found in [8], where a sharp concentration of the number of paths of given length connecting two given vertices is utilized. Here we give a quite accurate estimate of the upper tail of the number of rooted copies of a given rooted graph in $G(n,p)$; the result is similar to our main result in [4] for unrooted graphs, but somewhat simpler, except for a new complication for constant $p$.

A rooted graph $(R',H)$ is a *rooted subgraph* of $(R,G)$ if $H$ is a subgraph of $G$ and $R' = V(H) \cap R$. We let $N^R(G,H)$ denote the number of rooted copies of $H$ in $G$.

Given a rooted graph $(R,G)$ and a graph $F$ on the vertex set $V(F) = [n] = \{1,2,\dots,n\}$, let $r = |R|$ and regard $F$ as rooted on $[r] = \{1,\dots,r\}$; we say that a rooted subgraph of $([r],F)$ isomorphic to $(R,G)$ is an *$R$-rooted copy of $G$ in $F$*. Thus $N^R(F,G)$ is the number of $R$-rooted copies of $G$ in $F$. In particular, when $F$ is a random graph $G(n,p)$, we let the random variable $X = X_G^R = X_G^R(n,p)$ be the number $N^R(G(n,p),G)$ of $R$-rooted copies of $G$ in $G(n,p)$. We further define

$$\mu = \mu_R(G,n,p) := \mathbb{E}\, X_G^R = N^R(K_n,G)p^{e(G)}. \tag{3.1}$$

For a subgraph $H$ of $G$ let $H - R$ be the graph obtained from $H$ by deleting all vertices of $R$ (together with incident edges), and define

$$\Psi_H^R = \Psi_H^R(n,p) := n^{v(H-R)}p^{e(H)}. \tag{3.2}$$

Note that $\Psi_H^R = \Theta(\mathbb{E}\, X_H^{R'})$, with $R' = R \cap V(H)$, but as defined, it does not depend on the actual set $R'$ of roots of $H$.

Recall that, for a graph $H$, the fractional independence number $\alpha^*(H)$ is defined as the maximum value of $\sum_i x_i$ over all assignments $(x_i)_{i\in V(H)}$ such that $0 \le x_i \le 1$ for all vertices $i \in V(H)$ and $x_i + x_j \le 1$ for every edge $ij \in H$. We let

$$M_{R,G} = M_{R,G}(n,p) = \min_{H\subseteq G,\, e(H)>0} \left(\Psi_H^R\right)^{1/\alpha^*(H-R)}. \tag{3.3}$$

We further let

$$m_R(G) := \max_{H\subseteq G,\, e(H)>0} \frac{e(H)}{v(H-R)} > 0, \tag{3.4}$$

and note that (3.3), (3.2) and (3.4) imply that

$$M_{R,G} < 1 \iff np^{m_R(G)} < 1. \tag{3.5}$$

By the same argument as for the unrooted case in [7, Section 3.1], it is easy to show that $p = n^{-1/m_R(G)}$ is the threshold for the appearance of an $R$-rooted copy of $G$ in $G(n,p)$.

Let $e_R(G) = e(G) - e(G-R)$ be the number of edges in $G$ incident with the root set $R$. We assume below that $e_R(G) > 0$; the case $e_R(G) = 0$ is uninteresting since then $X_G^R$ equals the number of copies of the unrooted graph $G-R$ in $G(n,p)-[r]$, which we identify with $G(n-r,p)$, so $X_G^R(n,p) = X_{G-R}(n-r,p)$ and we may apply the results of [4].

**Theorem 3.1.** *For every rooted graph $(R,G)$ with $e_R(G)>0$ and for every $t>1$ there exist constants $c=c(t,G)$ and $C=C(t,G)$ such that for all $n \ge v(G)$, with $p_1 := t^{-1/e_R(G)}$ and $p_2 := t^{-1/e(G)}$:*

(a) *If $p \le n^{-1/m_R(G)}$, then*

$$p^C = \exp\{-C\log(1/p)\} \le \mathbb{P}(X_G^R \ge t\mu) \le \exp\{-c\}.$$

(b) *If $n^{-1/m_R(G)} \le p \le p_1$, then*

$$p^{CM_{R,G}} = \exp\{-CM_{R,G}\log(1/p)\} \le \mathbb{P}(X_G^R \ge t\mu) \le \exp\{-cM_{R,G}\}.$$

(c) *If $p_1 \le p \le p_2$, then*

$$\exp\{-C(n+(p-p_1)^2n^2)\} \le \mathbb{P}(X_G^R \ge t\mu) \le \exp\{-c(n+(p-p_1)^2n^2)\}.$$

(d) *If $p_2 < p \le 1$, then*

$$\mathbb{P}(X_G^R \ge t\mu) = 0.$$

Note that $0 < p_1 \le p_2 < 1$, and that $p_1$ and $p_2$ do not depend on $n$. Before giving the proof, we make some comments.

(i) Case (d) is trivial, because $p > p_2 \iff tp^{e(G)} > 1 \iff t\mu > N^R(K_n,G)$, see (3.1), so it is impossible to get at least $t\mu$ rooted copies of $G$ on $n$ vertices.

(ii) Case (a) is uninteresting and included only to show that the estimates in (b) extend in a continuous way to smaller $p$. (Note that $M_{R,G}=1$ at the threshold $p=n^{-1/m_R(G)}$, cf. (3.5).) Indeed, in case (a) we are below the threshold, so typically $X_G^R=0$.

(iii) If $e_R(G)=e(G)$, or equivalently $e(G-R)=0$, i.e., all edges in $G$ have a root as one endpoint, then $p_1=p_2$ and case (c) disappears, so that (b) is valid until the cutoff at $p_2$. For all other $G$, $p_1<p_2$ and case (c) appears, so there is a phase transition at $p_1$.

(iv) In the unrooted case in [4] there is also a phase transition at $p = n^{-1/\Delta_G}$. This has no counterpart in the rooted case.

(v) Since $e_R(G)>0$, $G$ has a rooted subgraph $H_0$ which is just a single edge with one endpoint in $R$; we have $\Psi_{H_0}^R = np$ and $\alpha^*(H_0-R) = \alpha^*(K_1)=1$, so

$$M_{R,G} \le \left(\Psi_{H_0}^R\right)^{1/\alpha^*(H_0-R)} = np \le n. \tag{3.6}$$

Hence, the upper bound in (b) is never stronger than $\exp\{-\Theta(n)\}$.

(vi) In (b) the exponents in the lower and upper bound are of the same order of magnitude except for the logarithmic term $\log(1/p)$; this inaccuracy disappears obviously for $p$ constant.

(vii) For any fixed $p>0$ (or $p=p(n)\in[p_0,1]$ for some constant $p_0>0$), $\Psi_H^R = \Theta(n^{v(H-R)})$. Since $\alpha^*(H-R) \le v(H-R)$ for all $H \subseteq G$, with equality for at least one $H$ with $e(H)>0$, viz. a single rooted edge, (3.3) shows that then $M_{R,G} = \Theta(n)$. Consequently, the result in (b) can be written for constant $p \leq p_1$ as $\mathbb{P}(X_G^R \geq t\mu) = \exp\{-\Theta(n)\}$. This shows that the bounds in (b) and (c) agree at $p = p_1$. Moreover, we obtain the following corollary.

**Corollary 3.2.** *With assumptions and notations as in Theorem 3.1, assume further that $p$ is fixed.*

(a) *If $0 < p \leq p_1$, then*
$$\mathbb{P}(X_G^R \geq t\mu) = \exp\{-\Theta(n)\}.$$
(b) *If $p_1 < p \leq p_2$, then*
$$\mathbb{P}(X_G^R \geq t\mu) = \exp\{-\Theta(n^2)\}.$$
(c) *If $p_2 < p \leq 1$, then*
$$\mathbb{P}(X_G^R \geq t\mu) = 0.$$

The sudden jump in the exponent from $n$ to $n^2$ at $p = p_1$ (for $G$ with $e(G - R) > 0$, so $p_1 < p_2$) may be surprising, and has no counterpart in the unrooted case in [4]. It may roughly be explained as follows (see the proof): If $p < p_1$, then it suffices (typically) to have all $\Theta(n)$ edges from the roots present in $G(n,p)$ in order to have more than $t\mu$ rooted copies of $G$. However, if $p > p_1$, this is not enough, and we need also (typically) a larger proportion than $p$ of the $\binom{n-r}{2}$ other possible edges, which by the usual Chernoff bound has probability only $\exp\{-\Theta(n^2)\}$.

*Proof of Theorem 3.1.* We mostly follow closely the proof for the unrooted case from [4], and therefore omit some details. As remarked above, (d) is trivial. Part (a) can be proved by a modification of the argument below, replacing $M_{R,G}$ by 1; we omit the details and refer to the corresponding argument in [4]. Hence we consider only (b) and (c). We let $C_1, C_2 \dots$ and $c_1, c_2, \dots$ denote constants that may depend on $G$ and $t$, but not on $n$ or $p$.

*Upper bounds*: If $(R, H)$ is a rooted graph, let $N^R(n, m, H)$ be the maximum of $N^R(F, H)$ over all rooted graphs $F$ with $v(F) \leq n$ and $e(F) \leq m$ and with a set of roots of size $|R|$. In other words, $N^R(n, m, H)$ is the maximum number of copies of $(R, H)$ that can be packed in $n$ vertices and $m$ edges with a given set of $|R|$ roots.

Let us start with the observation that if the minimum degree $\delta(H) > 0$ then
$$N^R(n, m, H) \leq N^R(2m, m, H) = O(N(2m, m, H - R)). \tag{3.7}$$
Indeed, for any $F$ with $v(F) \leq n$, $e(F) \leq m$, and $\delta(F) > 0$, we have $v(F) \leq 2m$, so the left hand side inequality follows. To prove the right hand side inequality, assume that $F$ and $H$ have the same set of roots $R$. Then
$$N^R(F, H) \leq N(F - R, H - R) \times 2^{|R|(v(H)-|R|)} = O(N(2m, m, H - R)).$$

Now, to prove the upper bound on $\mathbb{P}(X_G^R \geq t\mu)$, as before, we want to show that, say, $\mathbb{E}\, X^m \leq t^{m/2}\mu^m$, where $X = X_G^R$, $\mu = \mathbb{E}\, X$, and $m$ is suitably large. Similarly as in [4] and, as a matter of fact, similarly to the proof of Theorem 2.1 here, an inductive argument yields, for all $m \geq 1$,

$$\mathbb{E}\, X^m \leq \mu^m \biggl(1 + C_1 \sum_{H \subseteq G} \frac{N^{R'}\bigl(n,(m-1)e(G),H\bigr)}{\Psi_H^R}\biggr)^{m-1}, \tag{3.8}$$

where the sum extends over all rooted subgraphs $(R', H)$ of $(R, G)$ with $\delta(H) > 0$. ($H$ corresponds to the subgraph spanned by the edges in the intersection of the $m$th copy of $G$ and the union of the $m-1$ previous copies, and as such has $\delta(H) > 0$.)

We take $m := \lceil c_1 M_{R,G} \rceil$ for a suitable small constant $c_1 \in (0,1)$ to be fixed later. By (3.7), [4, Theorem 1.3] and (3.3), for every $H \subseteq G$ with $\delta(H) > 0$, assuming $m \geq 2$,

$$\begin{aligned} N^{R'}\bigl(n,(m-1)e(G),H\bigr) &\leq C_2 N\bigl(2(m-1)e(G),(m-1)e(G),H-R\bigr) \\ &= \Theta(m^{\alpha^*(H-R)}) = \Theta\bigl((c_1 M_{R,G})^{\alpha^*(H-R)}\bigr) \\ &\leq C_3 c_1 \Psi_H^R. \end{aligned}$$

Hence, (3.8) yields (the case $m = 1$ being trivial), $\mathbb{E}\, X^m \leq \mu^m (1+C_4 c_1)^{m-1}$. We choose $c_1$ so small that $1 + C_4 c_1 \leq t^{1/2}$, and then Markov's inequality yields

$$\mathbb{P}(X \geq t\mu) \leq \frac{\mathbb{E}\, X^m}{t^m \mu^m} \leq t^{-m/2} \leq \exp\{-c_2 M_{R,G}\}. \tag{3.9}$$

In particular, this yields the upper bound in (b).

For the upper bound in (c), we note that each rooted copy of $G$ in $K_n$ yields a copy of $G - R$ in $K_n - R = K_{n-r}$; conversely each copy of $G - R$ in $K_n - R$ can be extended to exactly $g$ rooted copies of $G$ in $K_n$, for some integer $g \geq 1$ depending on $G$. Hence, $X_G^R(n,p) \leq g X_{G-R}(n-r,p)$. Further, $N^R(K_n, G) = g N(K_{n-r}, G-R)$ so

$$\begin{aligned} \mu &= N^R(K_n, G) p^{e(G)} = g N(K_{n-r}, G-R) p^{e(G-R)+e_R(G)} \\ &= g\mu(G-R, n-r, p) p^{e_R(G)}. \end{aligned} \tag{3.10}$$

Consequently,

$$\begin{aligned} \mathbb{P}(X_G^R \geq t\mu) &\leq \mathbb{P}\bigl(g X_{G-R}(n-r,p) \geq t g \mu(G-R,n-r,p) p^{e_R(G)}\bigr) \\ &= \mathbb{P}\bigl(X_{G-R}(n-r,p) \geq t p^{e_R(G)} \mu(G-R,n-r,p)\bigr). \end{aligned} \tag{3.11}$$

Let $\tilde{t} := t p^{e_R(G)}$, and note that, for (c), $1 \leq \tilde{t} \leq t$. By [4, Theorems 1.2 and 1.5, and Remark 8.2], recalling that $t$ is fixed and $p \geq p_1$,

$$\mathbb{P}\bigl(X_{G-R}(n-r,p) \geq \tilde{t}\mu(G-R,n-r,p)\bigr) \leq \exp\{-c_3(\tilde{t}-1)^2 n^2\}. \tag{3.12}$$

Further,

$$\tilde{t} - 1 = t p^{e_R(G)} - 1 = (p/p_1)^{e_R(G)} - 1 \geq p/p_1 - 1 \geq p - p_1,$$

so (3.11)–(3.12) yield

$$\mathbb{P}(X_G^R \geq t\mu) \leq \exp\{-c_3(p-p_1)^2 n^2\}. \tag{3.13}$$

The upper bound in (c) now follows by taking the geometric mean of (3.9) and (3.13), noting that in this range of $p$, $M_{R,G} = \Theta(n)$ as remarked in (vii) above.

*Lower bounds*: Let $H$ be a subgraph of $G$ such that $e(H) > 0$ and

$$M := M_{R,G} = \left(\Psi_H^R\right)^{1/\alpha^*(H-R)}.$$

Since we consider parts (b) and (c) only, $M \geq 1$ by (3.5).

Set $p_0 = (3v_G t)^{-1}$ and assume first that $p \leq p_0$. (Note that $p_0 < t^{-1} \leq p_1$.) We construct, as in [4], a graph $F$ with

$$v(F) \leq 3(v_G - r)tM, \quad e(F) = O(M), \quad \text{and} \quad N(F, H-R) \geq 2t\Psi_H^R. \tag{3.14}$$

This is done as follows. Let $(x_i)_{i \in V(H-R)}$ be an optimal assignment for the fractional independence problem, that is, $0 \leq x_i \leq 1$, $x_i + x_j \leq 1$ for every edge $ij \in H - R$, and $\sum_i x_i = \alpha^*(H-R)$. Construct $F$ by blowing up each vertex of $H - R$ to a set of $\lceil 2tM^{x_i} \rceil$ vertices and replacing each edge of $H - R$ by the complete bipartite graph. This yields (3.14), where we have put 3 rather than 2 because of the ceiling. Now, by (3.6),

$$v(F) \leq 3(v_G - r)tM \leq 3(v_G - r)tnp \leq (1 - r/v_G)n \leq n - r.$$

We may thus fix a copy $F_1$ of $F$ with $V(F_1) \subseteq [n] \setminus [r]$; we further let $F_2$ be $F_1$ enlarged by adding all roots $1, \dots, r$ together with all $rv(F_1) = O(M)$ edges between the roots and $V(F_1)$. Now, exactly as in [4], it follows from [4, Lemma 3.3] that

$$\mathbb{P}(X_G^R \geq t\mu) \geq \frac{1}{4} p^{e(G)}\, \mathbb{P}(G(n,p) \supseteq F_2) = \frac{1}{4} p^{e(G)+e(F_2)} = p^{\Theta(M)}.$$

This proves the lower bound in (b) when $p \leq p_0$.

Assume now that $p_0 \leq p \leq p_2$ and note that the lower bound we want to prove can be written as $\exp\{-\Theta(n)\}$, see (vii) above or Corollary 3.2.

Consider first the case $e(G - R) = 0$ and observe that then the maximum number of copies of $G$ are obtained as soon as all edges from the roots appear, so, denoting this event by $\mathcal{E}^R$,

$$\mathbb{P}(X_G^R \geq t\mu) \geq \mathbb{P}\big(\mathcal{E}^R\big) = p^{r(n-r)} \geq e^{-C_5 n},$$

which proves the lower bound in (b) in this case. (Since $e(G-R) = 0$ implies $p_1 = p_2$, (c) is trivial.)

Thus, it remains to consider the case when $p_0 \leq p \leq p_2$ and $e(G-R) > 0$. We note first the trivial bound

$$\mathbb{P}(X_G^R \geq t\mu) \geq \mathbb{P}\big(G(n,p) = K_n\big) = p^{\binom{n}{2}} \geq e^{-C_6 n^2}. \tag{3.15}$$

Let $Z$ be the number of edges in $G(n-r, p)$. Since $Z$ has binomial distribution $\mathrm{Bin}\big(\binom{n-r}{2}, p\big)$ with mean $p\binom{n-r}{2}$, it is easily seen that if $(1 + 3\delta)p \leq 1$, and $\mathcal{E}_\delta$ is the event $\{Z \geq (1 + 3\delta)p\binom{n-r}{2}\}$, then

$$\mathbb{P}(\mathcal{E}_\delta) \geq c_4 \exp\{-C_7 n^2 \delta^2\}. \tag{3.16}$$

(The Chernoff bounds are essentially sharp, as is easily seen using Stirling's formula.) The number $X_{G-R}(n-r,p)$ of copies of $G-R$ in $G(n-r,p)$ is a sum of $N(K_{n-r},G-R)$ indicator variables $I_\alpha$. Conditioned on $Z=z$, each of them has the expectation

$$\mathbb{P}(I_\alpha = 1 \mid Z = z) = \frac{(z)_{e(G-R)}}{\left(\binom{n-r}{2}\right)_{e(G-R)}} = \left(\frac{z}{\binom{n-r}{2}}\right)^{e(G-R)}\Big(1 - O(z^{-1})\Big). \tag{3.17}$$

Let $N_\delta := (1+3\delta)p\binom{n-r}{2}$. If $\delta \geq n^{-1}$, $z \geq N_\delta$ and $n$ is large enough, then (3.17) yields

$$\begin{aligned}\mathbb{P}(I_\alpha = 1 \mid Z = z) &\geq (1+3\delta)^{e(G-R)}p^{e(G-R)}\Big(1-O(n^{-2})\Big)\\ &\geq (1+2\delta)p^{e(G-R)}.\end{aligned}$$

Consequently, if $\delta \geq n^{-1}$ and $n$ is large enough, then $\mathbb{P}(I_\alpha = 1 \mid \mathcal{E}_\delta) \geq (1+2\delta)p^{e(G-R)}$, and summing over $\alpha$ we find

$$\mathbb{E}(X_{G-R}(n-r,p) \mid \mathcal{E}_\delta) \geq (1+2\delta)\,\mathbb{E}\,X_{G-R}(n-r,p) = (1+2\delta)\mu(G-R).$$

Hence, by Lemma 3.2 of [4], as in the proof of Lemma 3.3 therein, with 1/2 replaced by $\frac{1+\delta}{1+2\delta}$, we obtain

$$\mathbb{P}\big(X_{G-R} \geq (1+\delta)\mu(G-R) \mid \mathcal{E}_\delta\big) \geq \Big(\frac{\delta}{1+2\delta}\Big)^2 \frac{\mu(G-R)}{N(K_{n-r},G-R)} \geq c_5\delta^2.$$

Assuming also the presence of all edges from the roots, i.e., the event $\mathcal{E}^R$, we have $X_G^R = gX_{G-R}$ (where $g$ is as in the proof of the upper bound); further, by (3.10), $\mu = g\mu(G-R)p^{e_R(G)}$; hence the inequality $X_{G-R} \geq (1+\delta)\mu(G-R)$ is equivalent to

$$X_G^R \geq (1+\delta)p^{-e_R(G)}\mu. \tag{3.18}$$

Consequently,

$$\mathbb{P}\big(X_G^R \geq (1+\delta)p^{-e_R(G)}\mu \mid \mathcal{E}^R, \mathcal{E}_\delta\big) \geq \mathbb{P}\big(X_{G-R} \geq (1+\delta)\mu(G-R) \mid \mathcal{E}_\delta\big) \geq c_5\delta^2$$

and thus, by (3.16),

$$\begin{aligned}\mathbb{P}\big(X_G^R \geq (1+\delta)p^{-e_R(G)}\mu\big) &\geq c_5\delta^2\,\mathbb{P}(\mathcal{E}_\delta \text{ and } \mathcal{E}^R) = c_5\delta^2\,\mathbb{P}(\mathcal{E}_\delta)\,\mathbb{P}(\mathcal{E}^R)\\ &\geq c_6 n^{-2}\exp\{-C_7\delta^2n^2\}p^{rn} = \exp\{-\Theta(\delta^2n^2+n)\},\end{aligned}$$

provided $1/n \leq \delta \leq \frac{1}{3}(p^{-1}-1)$ and $n$ is large enough.

For $p_0 \leq p \leq p_1$, we choose $\delta = n^{-1}$; then the right hand side of (3.18) is greater than $p_1^{-e_R(G)}\mu = t\mu$, so we obtain

$$\mathbb{P}(X_G^R \geq t\mu) \geq \exp\{-\Theta(n)\},$$

which as remarked above is equivalent to the lower bound in (b) for this range of $p$.

Finally, if $p_1 \le p \le p_2$, we take

$$\delta := \max\left\{tp^{e_R(G)} - 1, 1/n\right\} = \max\left\{(p/p_1)^{e_R(G)} - 1, 1/n\right\} = \Theta(p-p_1+1/n),$$

so that the right hand side of (3.18) is again at least $t\mu$. This yields the lower bound in (c) when $n$ is large enough and $p \ge p_1$ is small enough to guarantee that $\delta \le \frac{1}{3}(p^{-1}-1)$. For larger $p$, as well as for small $n$, we simply use (3.15). This completes the proof of the lower bound in (c). $\square$

3.1. **Examples and remarks.** It is easy to see that the minimum defining $M = M_{R,G}$ in (3.3) is achieved by a subgraph $H$ of $G$ such that $H - R$ is connected and, for every vertex $v \in H$, $H$ contains all edges leading from $v$ to $R$. These observations simplify computations of the bounds in Theorem 3.1.

**Example 3.3. Cliques rooted at a vertex.** Let $G = K_k$, $k \ge 2$, and $r = |R| = 1$. Then $m_R(G) = k/2$ and $e_R(G) = k-1$. To find $M$, consider first the candidates $H = K_2$ (with the root contained in $H$) and $H = G = K_k$. For $H = K_2$, we have, as shown in general in comment (v) above, $\left(\Psi_H^R\right)^{1/\alpha^*(H-R)} = np$. For $H = K_k$ we have $\Psi_{K_k} = n^{k-1}p^{\binom{k}{2}}$ and $\alpha^*(K_k - R) = \alpha^*(K_{k-1}) = (k-1)/2$, and thus $\left(\Psi_{K_k}^R\right)^{1/\alpha^*(K_k-R)} = n^2p^k$. Hence,

$$M \le \min\left\{np, n^2p^k\right\}; \tag{3.19}$$

we will show that equality holds.

To this end, consider a general $H \subseteq G$ with $e(H-R) > 0$ and let $F := H - R$. Then $e(H) \le e(F) + v(F)$ and so, see (3.2),

$$\begin{aligned}\frac{\Psi_H^R}{(np)^{\alpha^*(H-R)}} &\ge \frac{n^{v(F)}p^{e(F)+v(F)}}{(np)^{\alpha^*(F)}} \\ &= \left(np^{k-1}\right)^{v(F)-\alpha^*(F)} p^{e(F)-(k-2)(v(F)-\alpha^*(F))}\end{aligned} \tag{3.20}$$

and, dividing (3.20) by $\left(np^{k-1}\right)^{\alpha^*(H-R)}$,

$$\frac{\Psi_H^R}{(n^2p^k)^{\alpha^*(H-R)}} \ge \left(np^{k-1}\right)^{v(F)-2\alpha^*(F)} p^{e(F)-(k-2)(v(F)-\alpha^*(F))}. \tag{3.21}$$

Since $\frac{1}{2}v(F) \le \alpha^*(F) \le v(F)$, we have $v(F) - \alpha^*(F) \ge 0$ while $v(F) - 2\alpha^*(F) \le 0$, so $\left(np^{k-1}\right)^{v(F)-\alpha^*(F)} \ge 1$ if $np^{k-1} \ge 1$ and $\left(np^{k-1}\right)^{v(F)-2\alpha^*(F)} \ge 1$ if $np^{k-1} \le 1$. Further, by [4, Lemma 6.1], since $F \subseteq G - R = K_{k-1}$, we have $e(F) \le (k-2)(v(F) - \alpha^*(F))$, and thus $p^{e(F)-(k-2)(v(F)-\alpha^*(F))} \ge 1$ for all $p \in (0,1]$. Consequently, at least one of the right hand sides of (3.20) and (3.21) is $\ge 1$, so

$$\Psi_H^R \ge \min\left\{(np)^{\alpha^*(H-R)}, (n^2p^k)^{\alpha^*(H-R)}\right\},$$

or $(\Psi_H^R)^{1/\alpha^*(H-R)} \geq \min\{np, n^2p^k\}$. Finally, by (3.3) and (3.19),

$$M = \min\{np, n^2p^k\} = \begin{cases} n^2p^k, & p \leq n^{-1/(k-1)}, \\ np, & p \geq n^{-1/(k-1)}. \end{cases}$$

**Example 3.4. Bipartite graphs rooted at one whole side.** These are exactly the graphs with $e(G-R) = 0$, and so $p_1 = p_2$ (see comment (iii) after Theorem 3.1). The two classes of the bipartition are $R$ and $S = V(G) \setminus R$. Since the only connected subgraph of $G-R$ is $K_1$, and $\alpha^*(K_1) = 1$, we have from (3.3) and the comments above that $M = np^{\Delta_S(G)}$, where $\Delta_S(G) := \max_{v \in S} d_G(v)$ is the maximum degree in $G$ among all the vertices of $S$. Consequently, the upper bound in part (b) of Theorem 3.1 is of the form

$$\mathbb{P}(X_G^R \geq t\mu) \leq \exp\{-\Theta(np^{\Delta_S(G)})\}.$$

It follows from the above example that the bounds on $\mathbb{P}(X_G^R \geq t\mu)$ for $K_{s,2}$ with $r = 2$ and for even cycles $C_{2s}$ with $r = s$ are the same, since in both cases $\Delta_S(G) = 2$ . This is a special case of a more general phenomenon that the bounds depend only on the structure of $G - R$ and the degree sequence $|N_G(v) \cap R|$, $v \in V(G) \setminus R$. Our next example provides one more instance of that.

**Example 3.5. Paths rooted at the endpoints and cycles rooted at a vertex.** Let $G = P_k$ be a path with $k$ vertices, $k \geq 3$, and let $R$ be the set of its two endpoints. Then $m_R(G) = \frac{k-1}{k-2}$, and so $p \geq n^{-1/m_R(G)}$ implies that $np \to \infty$ as $n \to \infty$. The minimum in $M$ can be achieved only on a subpath $H$ on at most $k-2$ vertices containing one root, or $H = P_k$. So,

$$M = \min\left\{ \min_{1 \leq l \leq k-3} \left(n^l p^l\right)^{1/\lceil l/2 \rceil}, \left(n^{k-2}p^{k-1}\right)^{1/\lceil (k-2)/2 \rceil} \right\}.$$

The terms with even $l$ are all equal to $(np)^2$ while for odd $l$ they are equal to $(np)^{2l/(l+1)}$, which means that the smallest among them is $np$, the term corresponding to a single rooted edge. Hence, for even $k$, $M = np$ if $p \geq n^{-(k-2)/k}$, and otherwise $M = n^2p^{2(k-1)/(k-2)}$, the term corresponding to $H = G$. A similar cutoff for odd $k$ occurs at $n^{-(k-3)/(k-1)}$ with $M$ taking the values of $n^{2(k-2)/(k-1)}p^2$ and $np$, in turn.

Finally, note that if $R'$ is a single vertex in a cycle $C_{k-1}$, $k \geq 4$, then $m_{R'}(C_{k-1}) = m_R(P_k)$, $\Psi_{C_{k-1}}^{R'} = \Psi_{P_k}^R$, $\alpha^*(C_{k-1} - R') = \alpha^*(P_k - R)$, and the same is true for all other candidates for the minimum in $M$, that is, paths with a root at one end. Thus, $M_{R',C_{k-1}} = M_{R,P_k}$ and the upper tail bounds provided by Theorem 3.1 are the same for these two rooted graphs.

**Remark 3.6.** In the unrooted case, the lower tails are typically much smaller than the upper tails (see Remark 8.3 in [4]), and at best they can be of the same order of magnitude, e.g., when $p$ is fixed. Here, we encounter an opposite situation. Namely, for every $(R, G)$ with $e_R(G) > 0$ and a fixed $p$, by the FKG inequality, we have for any $t > 1$

$$\mathbb{P}(X_G^R \leq t\mu) \geq \mathbb{P}(X_G^R = 0) \geq \mathbb{P}(e_R(G(n,p)) = 0) = \exp\{-\Theta(n)\}$$

while for $t > 1$ and $p_1 < p \le p_2$, by Corollary 3.2,

$$\mathbb{P}(X_G^R \ge t\mu) = \exp\{-\Theta(n^2)\}.$$

**Remark 3.7.** If there are no isolated vertices in $H - R$ and $n \ge v(H)$, $m \ge e(H)$, then (3.7) may be improved to

$$N^R(n, m, H) = \Theta(N(n, m, H-R)) = \Theta(N(\min(n, 2m), m, H-R)). \quad (3.22)$$

Note, however, that this fails if $H$ contains a vertex whose all neighbors are among the roots; for example if $H$ is a rooted edge and $n > m$, then $N^R(n, m, H) = m$ and $N(n, m, H - R) = n$.

For the lower bound in (3.22), take a graph $F_0$ (with $V(F_0) \cap R = \emptyset$) which achieves the maximum in $N(n - r, m/3r, H - R)$; we may assume that $F_0$ has no isolated vertices, and thus at most $2m/(3r)$ vertices. Then join all vertices of $R$ to all vertices of $F_0$, obtaining a graph $F_1$ which contains $R$, has at most $n$ vertices, at most $m/(3r) + r2m/(3r) \le m$ edges, and is such that $N^R(F_1, H) \ge N(F_0, H - R)$. Hence, $N^R(n, m, H) \ge N(n - r, m/(3r), H - R)$. Finally, provided $m \ge 3re(H - R)$, we use the fact proved in [4] that if $n' = \Theta(n)$, $m' = \Theta(m)$ and $n, n' \ge v(H)$, $m, m' \ge e(H)$, then $N(n', m', H) = \Theta(N(n, m, H))$ (this follows directly from [4, Theorem 1.3]). The case $e(H) \le m < 3re(H - R)$ is trivial, since then both sides of (3.22) are $\Theta(1)$.

## References


[1] A. Dudek, J. Polcyn, and A. Ruciński, Subhypergraph counts in extremal and random hypergraphs and the fractional $q$-independence, *J. Combin. Optim.* DOI 10.1007/s10878-008-9174-9, Published online: 17 July 2008.

[2] E. Friedgut and J. Kahn, On the number of copies of one hypergraph in another, *Israel J. Math.* **105** (1998), 251–256.

[3] R. Graham, V. Rödl, and A. Ruciński, On Schur properties of random subsets of integers, *J. Number Th.* **61**(2) (1996), 388–408.

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

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

[6] S. Janson and A. Ruciński, The deletion method for upper tail estimates, *Combinatorica* **24**(4) (2004), 615–640

[7] S. Janson, T. Łuczak, and A. Ruciński, *Random Graphs*, John Wiley and Sons, New York (2000).

[8] T. Łuczak and P. Prałat, Chasing robbers on random graphs: zigzag theorem, preprint (2008).

[9] V. Rödl and A. Ruciński, Rado partition theorem for random subsets of integers, *Proc. London Math. Soc.* **74**(3) (1997), 481–502.

[10] 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.

Department of Mathematics, Uppsala University, PO Box 480, SE-751 06 Uppsala, Sweden
*E-mail address*: svante.janson@math.uu.se
*URL*: http://www.math.uu.se/∼svante/

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