J Comb Optim
DOI 10.1007/s10878-008-9174-9

Subhypergraph counts in extremal and random hypergraphs and the fractional q-independence

Andrzej Dudek · Joanna Polcyn · Andrzej Ruciński

Abstract We study the extremal parameter $N(n, m, H)$ which is the largest number of copies of a hypergraph H that can be formed of at most n vertices and m edges. Generalizing previous work of Alon (Isr. J. Math. 38:116–130, 1981), Friedgut and Kahn (Isr. J. Math. 105:251–256, 1998) and Janson, Oleszkiewicz and the third author (Isr. J. Math. 142:61–92, 2004), we obtain an asymptotic formula for $N(n, m, H)$ which is strongly related to the solution $\alpha_q(H)$ of a linear programming problem, called here the fractional q-independence number of H. We observe that $\alpha_q(H)$ is a piecewise linear function of q and determine it explicitly for some ranges of q and some classes of H. As an application, we derive exponential bounds on the upper tail of the distribution of the number of copies of H in a random hypergraph.

Keywords Hypergraphs · Fractional independence · Random hypergraphs

1 Introduction

The subject of this paper touches upon three areas of discrete mathematics: extremal combinatorics, linear programming and probability. A *hypergraph* is a pair (V, E),

The research of A. Ruciński was supported by the Polish grant N201036 32/2546. Part of this research was done while on leave at Emory University.

A. Dudek (✉)
Department of Mathematics and Computer Science, Emory University, Atlanta, GA, USA
e-mail: adudek@emory.edu

J. Polcyn · A. Ruciński
Department of Discrete Mathematics, Adam Mickiewicz University, Poznań, Poland

J. Polcyn
e-mail: joaska@amu.edu.pl

A. Ruciński
e-mail: rucinski@amu.edu.pl

where V is a finite set of *vertices*, while $E \subseteq 2^V$ is a family of subsets of V, called *edges*. For two hypergraphs F and H, let $N(F, H)$ be the number of isomorphic copies of H in F, that is,

$$N(F, H) = |\{H' \subseteq F : H' \cong H\}|,$$

where $\cong$ represents the isomorphism equivalence relation between hypergraphs. Furthermore, let

$$N(n, m, H) = \max\{N(F, H) : v_F \leq n, e_F \leq m\},$$

where v_F and e_F represent the number of vertices and edges of F, respectively. In other words, $N(n, m, H)$ is the largest number of copies of H that can be constructed using at most n vertices and m edges. Note that $N(n, m, H) = 0$ if either $n < v_H$ or $m < e_H$. Therefore, we will be always assuming that $n \geq v_H$ and $m \geq e_H$. Throughout the paper we use notation $a_n = \Theta(b_n)$ to mean that $cb_n < a_n < Cb_n$ for some $C > c > 0$.

Originally, there was an interest in a related problem with no restriction on the number of vertices. Let $N(m, H) = \max\{N(F, H) : e_F \leq m\}$. Alon (1981) proved that in the case of graphs, that is, when each edge of H has size 2,

$$N(m, H) = \Theta\big(m^{\alpha(H)}\big), \tag{1}$$

where $\alpha(H)$ is the fractional independence number of H (see below for the definition). Typically, in the literature the independence number is denoted by α and its fractional counterpart by α^*. Here we drop the asterisk for the ease of notation. Later, Friedgut and Kahn (1998) extended Alon's result (1) to all hypergraphs H using Shearer's entropy lemma (Chung et al. 1986).

In the random context, studied in Janson et al. (2004), it was necessary to restrict also the number of vertices. Following the ideas from Friedgut and Kahn (1998), an asymptotic formula for $N(n, m, H)$ was obtained in Janson et al. (2004) in the case when H is a graph. The proof in Janson et al. (2004) extends mutatis mutandis to hypergraphs yielding an asymptotic formula for $N(n, m, H)$, for every hypergraph H. In order to state this result, we need to introduce a parameter which is central for the entire paper.

For a real number $q \geq 0$ and a hypergraph H, the *fractional q-independence number* of H, denoted by $\alpha_q(H)$, is the optimum value of the following linear program $LP = LP(q, H)$.

$$\begin{aligned}
\text{(LP)}\quad & \text{Maximize} && \sum_{v \in V(H)} x_v \\
& \text{subject to} && \sum_{v \in e} x_v \leq q, \quad \text{for every } e \in E(H), \\
& && x_v \in [0, 1], \quad \text{for every } v \in V(H).
\end{aligned}$$

Note that for graphs, $\alpha(H) = \alpha_1(H)$ is the ordinary fractional independence number. The name *fractional independence number* comes from the fact that if we restrict ourselves to $x_v \in \{0, 1\}$, then for any optimal solution the set of vertices v with $x_v = 1$

is a maximum size independent set in H. There are several ways (one for each integer q) to generalize the notion of independence number to hypergraphs. Obviously, in the fractional relaxation there is no need to restrict q to integers. Now we are ready to state our result.

Theorem 1.1 *For every hypergraph H,*

$$N(n,m,H)=\Theta(n^{\alpha_q(H)}), \tag{2}$$

where $q=\log_n m$.

We will see in Sect. 3 that (1) is a special case of Theorem 1.1.

In Janson et al. (2004), Theorem 1.1, combined with an explicit formula for $\alpha_q(H)$, was just a tool in proving almost tight bounds on the upper tail of the distribution of the number X_H of the copies of a graph H in a random graph $G(n,p)$ (cf. Theorem 4.1 in Sect. 4). In the hypergraph case, due to its generality, it is not so easy anymore to apply formula (2) in the random hypergraph context. The problem is that the parameter $\alpha_q(H)$ is not always expressible as an explicit function of q. The goal of this paper is to initiate a study of the parameter $\alpha_q(H)$ and, as application, obtain more explicit formulae for $N(n,m,H)$, and, consequently, some estimates of the upper tails of subhypergraph counts in random hypergraphs.

More hypergraph notation and terminology We set $k_H=\max_{e\in E(H)}|e|$. If $|e|=k$ for each $e\in H$, we call H *k-uniform*. For each $v\in V(H)$, let $d_H(v)$ be the degree of v in H, namely the number of edges in H that contain v. We set $\Delta_H=\max_{v\in V(H)} d_H(v)$. If $d_H(v)=d$ for each vertex $v\in V(H)$, then we say that H is *d-regular*. A hypergraph is called *regular* if it is d-regular for some d. For a hypergraph H, we will use notation $v_H=v(H)=|V(H)|$ and $e_H=e(H)=|E(H)|$.

2 The q-independence number of hypergraphs

In this section we give several facts about the parameter $\alpha_q(H)$ which will be useful in computing explicit formulae for $N(n,m,H)$. The most important feature of $\alpha_q(H)$ is that it is a piecewise linear function of q. Indeed, if $\alpha_q(H)=b(H)+c(H)q$ for some range of q, then, by Theorem 1.1 , $N(n,m,H)=\Theta(n^{b(H)}m^{c(H)})$ in that range, which is a relatively simple formula. Let i_H be the number of isolated vertices in H.

Proposition 2.1 *For every hypergraph H,*

(i) *$\alpha_q(H)$ is a nondecreasing function of q, $i_H\le\alpha_q(H)\le v_H$, $\alpha_0(H)=i_H$, and $\alpha_q(H)=v_H$ for all $q\ge k_H$,*
(ii) *$\alpha_q(H)$ is a concave function of q,*
(iii) *$\alpha_q(H)$ is a piecewise linear, continuous function of q,*
(iv) *if H' is a spanning subhypergraph of H then $\alpha_q(H)\le\alpha_q(H')$,*
(v) *if H consists of connected components $H_1,\dots,H_s$ then $\alpha_q(H)=\alpha_q(H_1)+\dots+\alpha_q(H_s)$.*

Proof Parts (i) and (v) follow directly from the definition of $\alpha_q(H)$. For part (ii), observe that the average of the optimal solutions of $LP(q_1, H)$ and $LP(q_2, H)$ is a feasible solution of $LP(\frac{1}{2}(q_1 + q_2), H)$ for all $q_1 < q_2$. To verify part (iii), we recall Theorem 10.2 from Chvátal (1983) which, in our context, implies that there exist linear functions $f_1, \dots, f_\ell$ such that $\alpha_q(H) = \min\{f_i(q) : 1 \le i \le \ell\}$ for all $q \ge 0$ (this, of course, also implies (ii)). For (iv), notice that every feasible solution of $LP(q, H)$ is a feasible solution to $LP(q, H')$. □

Note that property (v) implies, in particular, that $\alpha_q(H) = \alpha_q(H_0) + i_H$, where H_0 is H stripped from its isolated vertices. For simplicity, we will be assuming that $i_H = 0$. For further references we reformulate parts (ii) and (iii) of Proposition 2.1 in a more detailed version.

Corollary 2.2 *There exists a (unique) finite sequence of real numbers* $0 = q_0 < q_1 < \dots < q_\ell = k_H$, *and a sequence of linear functions of* q

$$f_i^H(q) = b_i + c_i q \tag{3}$$

such that

(i) $0 = b_1 < \dots < b_\ell$ *and* $c_1 > \dots c_\ell \ge 0$,
(ii) *for each* $i = 1, \dots, \ell$

$$\alpha_q(H) = f_i^H(q) \quad \text{for } q \in [q_{i-1}, q_i],$$

(iii) *for all* $i = 1, \dots, \ell$ *and all* $q > 0$,

$$\alpha_q(H) \le f_i^H(q). \tag{4}$$

We will now determine $\alpha_q(H)$ at both ends of the range $0 \le q \le k_H$.

Fact 2.3

(i) *For all hypergraphs* H, *and all* $0 \le q \le 1$

$$\alpha_q(H) = \alpha_1(H) q,$$

that is, $q_1 \ge 1$, $b_1 = 0$, *and* $c_1 = \alpha_1(H)$;
(ii) *For all* k*-uniform hypergraphs* H *and all* $k - 1 \le q \le k$

$$\alpha_q(H) = k\alpha_{k-1}(H) - (k-1)v_H + (v_H - \alpha_{k-1}(H))\, q,$$

that is, $q_{\ell-1} \le k - 1$, $b_\ell = k\alpha_{k-1}(H) - (k-1)v_H$ *and* $c_\ell = v_H - \alpha_{k-1}(H)$.

Proof (i) Let x_v, $v \in V(H)$, be an optimal solution of $LP(q, H)$. Set $y_v = x_v/q$ and note that, because no vertex is isolated, $0 \le x_v \le q$, and so, $0 \le y_v \le 1$. Moreover, $\sum_{v \in e} y_v \le 1$ for each $e \in H$. Thus, (y_v) is a feasible solution of $LP(1, H)$ with $\sum_{v \in V(H)} y_v = \alpha_q(H)/q$. This implies that $\alpha_q/q \le \alpha_1(H)$. The reverse inequality follows from the concavity of α_q.

(ii) Let x_v, $v \in V(H)$, be an optimal solution of *LP*(q, H). Since $q \geq k-1$, we have $x_v \geq q-(k-1)$ for all $v \in V(H)$, because otherwise we could increase an x_v to $q-(k-1)$ without violating any constraints. Therefore, we may write

$$x_v = q-(k-1)+(k-q)\xi_v,$$

where $0 \leq \xi_v \leq 1$. Note that for each $e \in H$, $\sum_{v\in e}\xi_v \leq k-1$, since $\sum_{v\in e} x_v \leq q$.

Thus, (ξ_v) is a feasible solution of *LP*$(k-1, H)$, and x_v's are optimal if and only if the ξ_v's are. Therefore

$$\begin{aligned}\alpha_q(H) = \sum_{v\in V(H)} x_v &= (q-(k-1))v_H + (k-q)\sum_{v\in V(H)}\xi_v \\ &= (q-(k-1))v_H + (k-q)\alpha_{k-1}(H). \qquad \square\end{aligned}$$

Our next task is to derive bounds on the difference $\alpha_{q'}(H)-\alpha_q(H)$ and the quotient $\alpha_{q'}(H)/\alpha_q(H)$. From now on we assume that H is k-uniform.

Fact 2.4 *For all* $0<q<q'\leq k$

$$\text{(i)} \quad q'-q \leq \alpha_{q'}(H)-\alpha_q(H) \leq \alpha_1(H)(q'-q), \quad \textit{and}$$

$$\text{(ii)} \quad \frac{\alpha_{q'}(H)}{\alpha_q(H)} \leq \frac{q'}{q}.$$

Proof Let $q_{i-1} \leq q \leq q_i$ and $q_{j-1} \leq q' \leq q_j$ for some, possibly equal, i and j. Then, by (4), $c_j(q'-q) \leq \alpha_{q'}(H)-\alpha_q(H) \leq c_i(q'-q)$, and, by Corollary 2.2(i) and Fact 2.3 (both parts), $c_j \geq c_\ell = v_H - \alpha_{k-1}(H) \geq 1$ while $c_i \leq c_1 = \alpha_1(H)$. To bound the quotient, note that, again by (4),

$$\frac{\alpha_{q'}(H)}{\alpha_q(H)} \leq \frac{b_i + c_i q'}{b_i + c_i q} \leq \frac{q'}{q}. \qquad \square$$

Next, we prove general bounds on $\alpha_q(H)$. First, adding in (LP) all inequalities together yields

$$\sum_{v\in V(H)} x_v \leq \sum_{e\in E(H)}\sum_{v\in e} x_v \leq e_H q, \tag{5}$$

and consequently,

$$\alpha_q(H) \leq e_H q. \tag{6}$$

Moreover, the equality in (6) holds if every edge $e \in H$ contains $t_e \geq q$ vertices of degree 1. Indeed, then

$$x_v = \begin{cases} \frac{q}{t_e} & d_H(v) = 1, v \in e, \\ 0 & d_H(v) \geq 2, \end{cases}$$

is a feasible solution of *LP*(q, H) with the value of the objective function $e_H q$.

More bounds are given in the next result.

Fact 2.5 *For all* $0 \le q \le k$

$$\frac{v_H}{k} q \le \alpha_q(H) \le k\alpha_{k-1}(H) - (k-1)v_H + (v_H - \alpha_{k-1}(H))q$$
$$\le v_H - \frac{(k-q)e_H}{\Delta_H}. \tag{7}$$

Proof The first inequality follows from Fact 2.4(ii) and Proposition 2.1(i) ($\alpha_k = v_H$). The second inequality is a consequence of Fact 2.3(ii) combined with (4). The last inequality is equivalent to

$$\alpha_{k-1}(H) \le v_H - \frac{e_H}{\Delta_H}. \tag{8}$$

To prove (8), observe that any optimal solution of $LP(k-1, H)$ satisfies $1 \le \sum_{v\in e}(1-x_v)$, for every $e \in E(H)$. Thus,

$$e_H \le \sum_{e\in E(H)} \sum_{v\in e}(1-x_v) \le \Delta_H \sum_{v\in V}(1-x_v) = \Delta_H(v_H - \alpha_{k-1}).$$ □

Note that in view of Fact 2.3(ii), the (sharper) upper bound in (7) is achieved for $k-1 \le q \le k$. Also, for regular, nonempty k-uniform hypergraphs all bounds in (7) coincide. Indeed, if H is d-regular, $d > 0$, then $ke_H = dv_H$, and

$$v_H - \frac{(k-q)e_H}{\Delta_H} = v_H - \frac{(k-q)v_H}{k} = \frac{v_H}{k} q.$$

In fact, the lower bound in (7) is attained for all q by a broader class of hypergraphs, including those with perfect matchings.

Fact 2.6 *If a k-uniform hypergraph H contains a spanning subhypergraph H' such that each of the connected components of H' is regular but not an isolated vertex, then, for all* $0 \le q \le k$,

$$\alpha_q(H) = \frac{v_H}{k} q.$$

Proof By Proposition 2.1(iv), we have $\alpha_q(H) \le \alpha_q(H')$, so, in view of the lower bound in (7), all we need is that $\alpha_q(H') = \frac{v_H}{k} q$ (note that $v_H = v_{H'}$). As we have just pointed out, this is true for each connected component of H', and thus, by Proposition 2.1(v), it is true for H' as well. □

We conclude this section with finding the values of $\alpha_q(H)$ for two particular classes of k-uniform hypergraphs. *A loose cycle* $L_t^{(k)}$, where $k-1$ divides t, is a hypergraph with vertices $v_0, \dots, v_{t-1}$ and edges

$$\{v_1, v_2, \dots, v_k\}, \{v_k, v_{k+1}, \dots, v_{2k-1}\}, \dots, \{v_{t-k+2}, v_{t-k+3}, \dots, v_0, v_1\}.$$

Fact 2.7 *Let $L_t^{(k)}$ be a loose k-uniform cycle of length t. Then*

$$\alpha_q(L_t^{(k)}) = \begin{cases} \frac{t}{k-1}q, & q \le k-2, \\ \frac{(k-2)t}{2(k-1)} + \frac{t}{2(k-1)}q, & k-2 \le q \le k. \end{cases}$$

Proof We may assume that $k \ge 3$ (for $k = 2$ the loose cycles become to 2-regular graphs and are covered by Fact 2.6). The case when $0 \le q \le k-2$ follows from (6) because each edge contains precisely $k-2 \ge q$ vertices of degree 1. Suppose that $k-2 \le q \le k$. Then, (7) implies

$$\alpha_q(L_t^{(k)}) \le t - \frac{(k-q)t}{2(k-1)} = \frac{(k-2)t}{2(k-1)} + \frac{t}{2(k-1)}q.$$

The converse inequality is derived by taking the feasible solution

$$x_v = \begin{cases} 1, & d_H(v) = 1, \\ \frac{q-(k-2)}{2}, & d_H(v) = 2. \end{cases}$$ □

Fact 2.8 *Let $K^{(k)}(t_1, \dots, t_k)$ be the k-uniform, k-partite complete hypergraph with k-partition $V = V_1 \cup \dots \cup V_k$, where $|V_i| = t_i$ and $t_1 \ge \dots \ge t_k$. Further, let $j = 1, \dots, k$ and $j-1 < q \le j$. Then*

$$\alpha_q(K^{(k)}(t_1, \dots, t_k)) = t_1 + \dots + t_{j-1} - (j-1)t_j + t_j q.$$

Proof Suppose that $0 \le j-1 < q \le j \le k$. Let $d_i = d_H(v)$ for every $v \in V_i$. Then, $d_i = \frac{e_H}{t_i}$ and $d_1 \le d_2 \le \dots \le d_k$. By (5), we have

$$d_1 \sum_{v \in V_1} x_v + d_2 \sum_{v \in V_2} x_v + \dots + d_{j-1} \sum_{v \in V_{j-1}} x_v + d_j \sum_{v \in V_j \cup \dots \cup V_k} x_v \le e_H q,$$

or, equivalently,

$$d_j \sum_{v \in V} x_v \le e_H q + (d_j - d_1) \sum_{v \in V_1} x_v + (d_j - d_2) \sum_{v \in V_2} x_v + \cdots$$
$$+ (d_j - d_{j-1}) \sum_{v \in V_{j-1}} x_v.$$

Moreover, since $x_v \le 1$ and $d_i t_i = e_H$, we have

$$d_j \sum_{v \in V} x_v \le e_H q + (d_j - d_1)t_1 + (d_j - d_2)t_2 + \dots + (d_j - d_{j-1})t_{j-1}$$
$$= e_H q + d_j(t_1 + t_2 + \dots + t_{j-1}) - (j-1)e_H,$$

which yields,

$$\alpha_q(K^{(k)}(t_1, \dots, t_k)) \le t_1 + \dots + t_{j-1} - (j-1)t_j + t_j q.$$

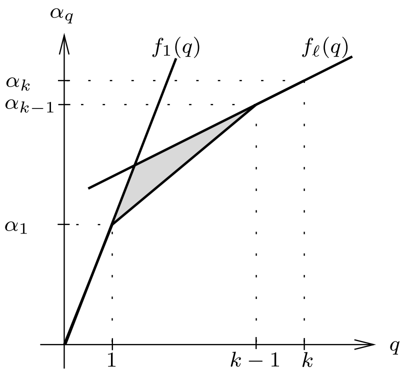

Fig. 1 The placement of (q, α_q) points, for $1 \leq q \leq k-1$

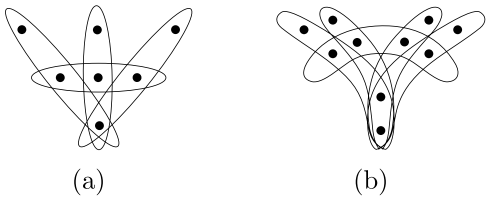

Fig. 2 Hypergraphs $H_{3,2}$ (**a**) and $H_{4,2}$ (**b**)

The converse inequality follows by taking the feasible solution

$$x_v = \begin{cases} 1, & v \in V_1 \cup \cdots \cup V_{j-1}, \\ q-j+1, & v \in V_j, \\ 0, & \text{otherwise.} \end{cases}$$ □

In summary, we have found explicit expressions for $\alpha_q(H)$ as a linear function of q for both, $0 \leq q \leq 1$ and $k-1 \leq q \leq k$. By the concavity, in the remaining range $(1 \leq q \leq k-1)$, the points $(q, \alpha_q(H))$ must lie in the triangle T_H located above the straight line segment joining the points $(1, \alpha_1(H))$ and $(k-1, \alpha_{k-1}(H))$, and, by Corollary 2.2(iii), below the lines $f_1(q)$ and $f_\ell(q)$ (see Fig. 1). This triangle may have an empty interior, as is the case of hypergraphs satisfying the assumptions of Fact 2.6 (where $\ell = 1$). Otherwise, it can be still traversed by the graph of $\alpha_q(H)$ along the upper boundary, lower boundary or through the interior. Indeed, let $H_{3,2}$ be the 3-uniform hypergraph in Fig. 2(a). Then

$$\alpha_q(H_{3,2}) = \begin{cases} 3q, & q < \frac{3}{2}, \\ \frac{5}{3}q + 2, & q \geq \frac{3}{2}, \end{cases}$$

and therefore, $\ell = 2$, $q_0 = 0$, $q_1 = \frac{3}{2}$ and $q_2 = 3$, and $\alpha_q(H_{3,2})$ traverses the upper boundary of the triangle $T_{H_{3,2}}$.

As we have seen in Fact 2.8 , the 3-partite complete hypergraph $K = K^{(3)}(3, 2, 1)$ has $\ell = 3$, $q_i = i$, $i = 0, 1, 2, 3$ and thus $\alpha_q(K)$ traverses the bottom of T_K. Finally, for the disjoint union $H_{3,2} \cup K$ we have $\ell = 4$, $q_0 = 0$, $q_1 = 1, q_2 = 3/2$,

$q_3 = 2$ and $q_4 = 3$, which means that the graph of $\alpha_q(H_{3,2} \cup K)$ traverses the interior of $T_{H_{3,2}\cup K}$.

It seems to be interesting to determine, for a fixed k, the largest value of $\ell = \ell(H)$, the number of straight line segments of the function $\alpha_q(H)$. Let

$$\tau(k) = \max_H \ell(H),$$

where the maximum is taken over all k-uniform hypergraphs H. We know only that $\tau(2) = 2$ (cf. Fact 2.3 for $k = 2$). However, for $k \geq 3$ it might be the case that $\tau(k) = \infty$. In fact we believe that this is indeed true.

Conjecture 2.9 *For $k \geq 3$ we have $\tau(k) = \infty$.*

The example of $H_{3,2} \cup K^{(3)}(3, 2, 1)$ shows that $\tau(3) \geq 4$. Moreover, it suggests the right approach to the problem. Namely, in order to obtain a hypergraph H with large $\ell(H)$, it is sufficient to construct a family of hypergraphs $H_1, H_2, \dots$ with distinct "break points" of the function $\alpha_q(H_i)$, $i = 1, 2, \dots$, and then take their vertex-disjoint union $\bigcup_i H_i$. Recently, Matoušek (2008) found two further such hypergraphs yielding $\tau(3) \geq 6$.

The construction of $H_{3,2}$ can be easily generalized for arbitrary $k \geq 3$ as follows. For a given $k \geq 3$ and $0 \leq j \leq k$ let $H_{k,j}$ be the k-uniform hypergraph with the edge set $\{e_1, \dots, e_{k+1}\}$ satisfying, for some $S \subseteq V(H_{k,j})$ of size $k - j$, the following conditions:

(i) $e_{i_1} \cap e_{i_2} = S$, for any $1 \leq i_1 < i_2 \leq k$,
(ii) $e_{k+1} \cap S = \emptyset$,
(iii) $|e_i \cap e_{k+1}| = 1$, for any $1 \leq i \leq k$.

In Fig. 2 there are two examples of graphs $H_{k,j}$. It is easy to check that

$$\alpha_q(H_{k,j}) = \begin{cases} kq, & q < j - 1 + \frac{j-1}{k-1}, \\ (k-1)(j-1) + 2q - \frac{q}{k}, & \text{otherwise.} \end{cases}$$

Let $K = K^{(k)}(k, \dots, 1)$ be the k-partite complete hypergraph with parts of size $k, \dots, 1$. Then one can check that $\ell(\bigcup_{j=2}^{k-1} H_{k,j} \cup K) = 2k - 2$, implying $\tau(k) \geq 2k - 2$.

3 Subhypergraph counts in extremal hypergraphs

As we mentioned in the introduction, the proof of Theorem 1.1 goes along the lines of the proof of its special case for graphs given in Janson et al. (2004), and therefore we omit it here. However, to give the reader some indication how the parameter $\alpha_q(H)$ becomes involved in the formula for $N(n, m, H)$, we present the proof of the lower bound, that is, we show that $N(n, m, H) \geq cn^{\alpha_q(H)}$ for some $c > 0$.

Proof of Theorem 1.1—the lower bound Let n, m and H be given, and let $q = \log_n m$. Let x_v, $v \in V$, be an optimal solution of *LP*(q, H), and $y_v = x_v \log n$. We construct a graph F by blowing up each vertex $v \in V(H)$ to a set of size $n_v = \lceil c_H \exp(y_v) \rceil$, where $c_H = v_H^{-1} e_H^{-1}$. In other words, $V(F) = \bigcup_{v \in V(H)} V_v$, where V_v's are pairwise disjoint sets, $|V_v| = n_v$, and each k-tuple of sets $(V_{v_1}, V_{v_2}, \dots, V_{v_k})$ spans the complete k-uniform, k-partite hypergraph if $\{v_1, v_2, \dots, v_k\} \in E(H)$, and it spans the empty hypergraph otherwise.

Since $y_v \le \log n$ and $\sum_{v \in e} y_v \le \log m$, for every $v \in V(H)$ and $e \in E(H)$, we have $v_F \le n$ and $e_F \le m$. Moreover,

$$N(F, H) \ge \prod_{v \in V(H)} n_v \ge c_H^{v_H} \exp\Bigl(\log n \sum_{v \in V(H)} x_v\Bigr) = c_H^{v_H} n^{\alpha_q(H)}. \qquad \square$$

Note that if H_0 denotes H stripped from its isolated vertices and $i_H = |V(H) \setminus V(H_0)|$ then

$$N(n, m, H) = \Theta\bigl(N(n, m, H_0) n^{i_H}\bigr).$$

Therefore, we will be assuming that $i_H = 0$, that is, $H_0 = H$.

Note also that it follows from Theorem 1.1 that the order of magnitude of $N(n, m, H)$ is not affected by a linear change in n, that is $N(\lfloor Cn \rfloor, m, H) = \Theta(N(n, m, H))$. With the help of Fact 2.4(i) one can show that the same is true with respect to m. This observation will be utilized in the next section.

Corollary 3.1 *For all k-uniform hypergraphs H and for every $C > 1$,*

$$N(n, \lfloor Cm \rfloor, H) = \Theta(N(n, m, H)).$$

Proof Given $C > 1$, let $q' = \log(Cm)/\log n = q + \log C/\log n$. Then, $q' > q$, and consequently, Fact 2.4 yields

$$n^{\alpha_{q'}(H)} \ge n^{\alpha_q(H)} n^{q'-q} = C n^{\alpha_q(H)}$$

and

$$n^{\alpha_{q'}(H)} \le n^{\alpha_q(H)} n^{\alpha_1(H)(q'-q)} = C^{\alpha_1(H)} n^{\alpha_q(H)}. \qquad \square$$

Our knowledge about $\alpha_q(H)$ can be used to derive explicit formulae for $N(n, m, H)$ and $N(m, H)$. First note that, by Theorem 1.1, for all hypergraphs H and all $n^{q_{i-1}} \le m \le n^{q_i}$,

$$N(n, m, H) = \Theta(n^{b_i} m^{c_i}), \tag{9}$$

where q_i, b_i and c_i were defined in (3). In particular, if $m \le n$, then, using Fact 2.3,

$$N(n, m, H) = \Theta(n^{\alpha_1(H) q}) = \Theta(m^{\alpha_1(H)})$$

and thus

$$N(m, H) = N(k_H m, m, H) = \Theta(m^{\alpha_1(H)}), \tag{10}$$

which is the result from Friedgut and Kahn (1998). If H is k-uniform, that is, $k_H = k$, then at the other end of the range, that is, for $n^{k-1} \le m \le n^k$, we have again by Fact 2.3

$$N(n,m,H) = \Theta\left(n^{k\alpha_{k-1}(H)-(k-1)v_H} m^{v_H-\alpha_{k-1}(H)}\right).$$

In the special case of graphs we thus recover a result from Janson et al. (2004).

Theorem 3.2 (Janson et al. 2004) *For every graph H*

$$N(n,m,H) = \begin{cases} \Theta(m^{\alpha(H)}), & \text{if } m \le n, \\ \Theta(n^{2\alpha(H)-v_H} m^{v_H-\alpha(H)}), & \text{if } n \le m \le \binom{n}{2}. \end{cases}$$

Turning to specific classes of graphs, by Fact 2.6 we know that if a k-uniform hypergraph H contains a spanning subhypergraph H' such that each of the connected components of H' is regular but not an isolated vertex, then for $m \le n^k$,

$$N(n,m,H) = \Theta\left(m^{v_H/k}\right). \tag{11}$$

In particular, the above formula remains true for all k-uniform hypergraphs H containing perfect matchings, as well as for all regular k-uniform hypergraphs H, including complete hypergraphs.

Moreover, as consequences of Facts 2.7 and 2.8, we know that

$$N(n,m,L_t^{(k)}) = \begin{cases} \Theta(m^{\frac{t}{k-1}}), & m \le n^{k-2}, \\ \Theta(n^{\frac{t(k-2)}{2(k-1)}} m^{\frac{t}{2(k-1)}}), & n^{k-2} \le m \le n^k, \end{cases} \tag{12}$$

and, for $n^{j-1} < m \le n^j$,

$$N\left(n,m,K^{(k)}(t_1,\dots,t_k)\right) = \Theta\left(n^{t_1+\cdots+t_{j-1}-(j-1)t_j} m^{t_j}\right).$$

4 Subhypergraph counts in random hypergraphs

Let $G^{(k)}(n,p)$, $k \ge 2$, denote the random k-uniform hypergraph with n labeled vertices and such that each of the $\binom{n}{k}$ possible edges exists with probability p, independently of all other edges. Let X_G be the number of isomorphic copies of G contained in $G^{(k)}(n,p)$ and $\mu_G = \mathbf{E}X_G$. In this section we reserve the letter H to denote subhypergraphs of G.

In the case of graphs the distribution of X_G has been studied extensively since the pioneering paper by Erdős and Rényi (1960). A general threshold for $\{X_G > 0\}$ was established by Bollobás (1981) at $p = n^{-1/m_G}$, where

$$m_G = \max_{H \subseteq G} \frac{e_H}{v_H}. \tag{13}$$

Next, it was shown that the lower tail of the distribution of X_G decays exponentially in the expectation of the least expected subgraph of G, see (Janson et al. 1990) and

(Janson 1990). Namely, let $\Psi_H = n^{v_H} p^{e_H}$, which is roughly the expected number of copies of H in $G(n, p)$. Then, for all $\varepsilon \in (0, 1]$, with $c_\varepsilon > 0$ depending on H and ε,

$$\mathbf{P}\big(X_G \le (1-\varepsilon)\mu_G\big) \le \exp\Big(-c_\varepsilon \min_{H \subseteq G,\, e_H > 0} \Psi_H\Big).$$

This is best possible, provided p stays away from 1, as by the FKG inequality, $-\log \mathbf{P}(X_G = 0) \le -\log \mathbf{P}(X_H = 0) = O(\Psi_H)$ for every $H \subseteq G$ (see Janson et al. 2000 for details). Finally, in Janson et al. (2004), almost as tight bounds were found for the upper tail of X_G, that is, for $\mathbf{P}(X_G \ge (1+\varepsilon)\mu_G)$.

All these results are easily extendable to k-uniform hypergraphs. Here we focus on a generalization of the latter one. Let G be a k-uniform hypergraph and for any integer n and $0 < p < 1$, let $\Psi_G = n^{v_G} p^{e_G}$. Let $K_r^{(k)}$ denote the complete k-uniform hypergraph on r vertices. Finally, let

$$M_G = M_G(n, p) = \max\left\{ m \le \binom{n}{k} : \forall H \subseteq G\ N(n, m, H) \le \Psi_H \right\} \tag{14}$$

for $p \ge n^{-k}$ and 1 otherwise. A reason for the special definition in the extreme case $p < n^{-k}$ (when $\Psi_{K_k^{(k)}} = n^k p < 1$), is to prevent $M_G = 0$. Since $N(n, 1, H) = 0$ unless $e_H \le 1$, it is easily checked that $1 \le M_G \le \binom{n}{k}$.

Now we can formulate a result, which is a straightforward generalization of Theorem 1.2 in Janson et al. (2004).

Theorem 4.1 *For every k-uniform hypergraph G and for every $\varepsilon > 0$ there exist constants $c(\varepsilon, G) > 0$ and $C(\varepsilon, G) > 0$ such that for all $n \ge v_G$ and $p \in (0, 1)$*

$$\mathbf{P}(X_G \ge (1+\varepsilon)\mu_G) \le \exp\{-c(\varepsilon, G) M_G(n, p)\}, \tag{15}$$

and, provided $(1+\varepsilon)\mu_G \le N(K_n^{(k)}, G)$,

$$\mathbf{P}(X_G \ge (1+\varepsilon)\mu_G) \ge \exp\{-C(\varepsilon, G) M_G(n, p) \log(1/p)\}.$$

We omit the complete proof and present only one part which points to the connection between the upper tail of X_G and the extremal parameter $N(n, m, H)$.

Sketch of proof of (15) Note that by Markov's inequality, for every positive integer m we have

$$\mathbf{P}(X_G \ge (1+\varepsilon)\mu_G) \le \frac{\mathbf{E} X_G^m}{(1+\varepsilon)^m \mu_G^m}. \tag{16}$$

In order to turn (16) into an exponential bound, it will be enough to show that

$$\mathbf{E} X_G^m \le \left(1 + \frac{\varepsilon}{2}\right)^m \mu_G^m, \tag{17}$$

for some large m. Let $G_1, G_2, \ldots, G_{N(K_n^{(k)}, G)}$ be all copies of G in the complete hypergraph $K_n^{(k)}$. Moreover, let I_{G_i} be an indicator random variable corresponding to G_i. Then,

$$
\begin{aligned}
\mathbf{E}X_G^m &= \sum_{i_1,\dots,i_m} \mathbf{E}(I_{G_{i_1}} \cdots I_{G_{i_m}}) \\
&= \sum_{i_1,\dots,i_m} p^{e(G_{i_1}\cup\dots\cup G_{i_m})} \\
&= \sum_{i_1,\dots,i_{m-1}} p^{e(G_{i_1}\cup\dots\cup G_{i_{m-1}})} \sum_{i_m} p^{e_G - e((G_{i_1}\cup\dots\cup G_{i_{m-1}})\cap G_{i_m})} \\
&= \mathbf{E}X_G^{m-1} \sum_{i_m} p^{e_G - e(F\cap G_{i_m})}, \qquad (18)
\end{aligned}
$$

where $F = G_{i_1} \cup \dots \cup G_{i_{m-1}}$. Furthermore,

$$
\begin{aligned}
\sum_i p^{e_G - e(F\cap G_i)} &= \sum_{i:F\cap G_i=\emptyset} p^{e(G)} + \sum_{i:F\cap G_i\neq\emptyset} p^{e_G - e(F\cap G_i)} \\
&\le \mu_G + \sum_{H\subseteq G, e_H>0} \sum_{i:F\cap G_i\cong H} p^{e_G - e_H} \\
&\le \mu_G + \sum_{H\subseteq G, e_H>0} N(F,H) n^{v_G - v_H} p^{e_G - e_H}. \qquad (19)
\end{aligned}
$$

The last inequality follows because there are at most $N(F, H)$ copies of H in F, and every such copy can be extended to at most $n^{v_G - v_H}$ copies of G. Since $v_F \le n$ and $e_F \le (m-1)e_G$, the definition of $N(n, m, H)$ and Corollary 3.1 imply, that

$$
N(F, H) \le N(n, (m-1)e_G, H) = \Theta(N(n, m, H)). \qquad (20)
$$

Consequently, (18), (19) and (20) yield that

$$
\begin{aligned}
\mathbf{E}X_G^m &\le \mathbf{E}X_G^{m-1}\left(\mu_G + \sum_{H\subseteq G, e_H>0} \Theta(N(n,m,H)) n^{v_G - v_H} p^{e_G - e_H}\right) \\
&= \mathbf{E}X_G^{m-1}\mu_G\left(1 + \sum_{H\subseteq G, e_H>0} \Theta\left(\frac{N(n,m,H)}{\Psi_H}\right)\right).
\end{aligned}
$$

Hence, by induction on m,

$$
\mathbf{E}X_G^m \le \mu_G^m \left(1 + \sum_{H\subseteq G, e_H>0} \Theta\left(\frac{N(n,m,H)}{\Psi_H}\right)\right)^{m-1}.
$$

Now, by the definition of M_G and in view of (the proof of) Corollary 3.1, one can choose a constant c such that

$$
\sum_{H\subseteq G, e_H>0} \Theta\left(\frac{N(n,\lfloor cM_G\rfloor,H)}{\Psi_H}\right) < \frac{\varepsilon}{2},
$$

and consequently, (17) holds with $m = \lfloor cM_G \rfloor$. □

To apply Theorem 4.1, we need to estimate M_G, and for this, in turn, it is crucial to have a fair estimate of the extremal parameter $N(n,m,H)$ for every graph $H \subseteq G$. In the case of graphs, Theorem 3.2 allows to completely determine the asymptotic order of M_G.

Theorem 4.2 (Janson et al. 2004) *For every graph G*

$$M_G(n,p) = \begin{cases} \Theta(1), & \text{if } p \le n^{-1/m_G}, \\ \Theta(\min_{H\subseteq G} \Psi_H^{1/\alpha(H)}), & \text{if } n^{-1/m_G} \le p \le n^{-1/\Delta_G}, \\ \Theta(n^2 p^{\Delta_G}), & \text{if } p \ge n^{-1/\Delta_G}. \end{cases}$$

Unfortunately, in the case of k-uniform hypergraphs, $k \ge 3$, we can do it only for those H and for those ranges of p for which the value of $\alpha_q(H)$ has been explicitly determined as a function of q, and, consequently, the asymptotic range of $N(n,m,H)$ is known (as a function of m and n).

We will first illustrate the technique of estimating M_G by two examples (regular k-uniform hypergraphs and loose k-cycles), before turning to the general case. Owing to Corollary 3.1, in determining the order of magnitude of M_G it is sufficient to find an m for which $N(n,m,H) = O(\Psi_H)$ for all $H \subseteq G$, and $N(n,m,H_0) = \Theta(\Psi_{H_0})$ for *some* $H_0 \subseteq G$.

Note that if $p < n^{-1/m_G}$ and $H \subseteq G$ satisfies $m_G = e_H/v_H$ (cf. (13)), then

$$\Psi_H = n^{v_H} p^{e_H} < n^{v_H} n^{-\frac{1}{m_G} e_H} = n^{v_H} n^{-\frac{v_H}{e_H} e_H} = 1,$$

and thus, $M_G = \Theta(1)$.

Proposition 4.3 *If G is a d-regular k-uniform hypergraph and $p \ge n^{-1/m_G} = n^{-k/d}$, then $M_G(n,p) = \Theta(n^k p^d)$.*

Proof Set $m = n^k p^d$ and recall that $q = \log_n m$. Theorem 1.1 together with the upper bound in (7) yields that, for every $H \subseteq G$,

$$\begin{aligned} N(n,m,H) &= \Theta(n^{\alpha_q(H)}) = O(n^{v_H-(k-q)e_H/\Delta_H}) = O(n^{v_H-ke_H/\Delta_H} n^{qe_H/\Delta_H}) \\ &= O(n^{v_H-ke_H/\Delta_H} m^{e_H/\Delta_H}) = O(n^{v_H-ke_H/\Delta_H} n^{ke_H/\Delta_H} p^{de_H/\Delta_H}) \\ &= O(n^{v_H} p^{e_H}) = O(\Psi_H). \end{aligned}$$

On the other hand, with $H_0 = G$, by (11),

$$N(n,m,H_0) = \Theta(m^{v_{H_0}/k}) = \Theta(n^{v_{H_0}} p^{dv_{H_0}/k}) = \Theta(n^{v_{H_0}} p^{e_{H_0}}) = \Theta(\Psi_{H_0}). \qquad \square$$

Note that for the loose k-cycle $G = L_t^{(k)}$ we have $m_G = \frac{1}{k-1}$.

Proposition 4.4 *For the loose k-cycle $G = L_t^{(k)}$ we have*

$$M_G(n,p) = \begin{cases} \Theta(1), & \text{if } p < n^{1-k}, \\ \Theta(n^{k-1}p), & \text{if } n^{1-k} \le p < n^{-1}, \\ \Theta(n^k p^2), & \text{if } p \ge n^{-1}. \end{cases}$$

Proof Let $p \le 1/n$ and set $m = n^{k-1}p$. Then $m \le n^{k-2}$ and $q = \log_n m \le k-2$. Note that for each $H \subseteq G$, by (6), $\alpha_q(H) = e_H q$. Moreover, $m_G = \frac{1}{k-1}$, that is, $e_H/v_H \le m_G = \frac{1}{k-1}$. Hence,

$$N(n,m,H) = \Theta(m^{e_H}) = \Theta(n^{(k-1)e_H} p^{e_H}) = O(n^{v_H} p^{e_H}) = O(\Psi_H),$$

while

$$N(n,m,G) = \Theta\big(m^{\frac{t}{k-1}}\big) = \Theta\big(n^t p^{\frac{t}{k-1}}\big) = \Theta(\Psi_G).$$

For $p \ge 1/n$, set $m = n^k p^2 \ge n^{k-2}$. Then, $n^{-(k-q)} = p^2$, and so, by Theorem 1.1 and the rightmost upper bound in (7),

$$N(n,m,H) = \Theta(n^{\alpha_q(H)}) = O(n^{v_H-(k-q)e_H/\Delta_H}) = O(n^{v_H} p^{2e_H/\Delta_H}) = O(\Psi_H),$$

where the last inequality follows from the fact that $\Delta_H \le 2$. On the other hand, with $H_0 = G$, by (12),

$$N(n,m,H_0) = \Theta\big(n^{\frac{(k-2)t}{2(k-1)}} m^{\frac{t}{2(k-1)}}\big) = \Theta(\Psi_{H_0}). \qquad \square$$

For a general k-uniform hypergraph G, by Corollary 2.2 applied simultaneously to all subhypergraphs $H \subseteq G$, there exists a finite sequence $q_0 < q_1 < \cdots < q_r$, such that for each $i = 1, \ldots, r$, each subhypergraph $H \subseteq G$, and for all $q_{i-1} \le q \le q_i$ we have

$$\alpha_q(H) = f_i^H(q) = b_i(H) + c_i(H)q,$$

for some $b_1(H) \le \cdots \le b_r(H)$ and $c_1(H) \ge \cdots \ge c_r(H)$. Note that $r \le \sum_{H \subseteq G} \ell(H)$, $q_{r-1} \le k-1$ and $q_r = k$. Set also $f_0^H(q) = 0$ for convenience. For $i = 0, 1, \ldots, r-1$, define

$$s_i(G) = s_i = \max_{H \subseteq G} \frac{e_H}{v_H - f_i^H(q_i)}$$

and $p_i = n^{-1/s_i}$, $i = 0, \ldots, r-1$, and $p_r = 1$.

Theorem 4.5 *For all k-uniform hypergraphs G*

$$M_G(n,p) = \Theta\Big(\min_{H \subseteq G} n^{-b_i(H)/c_i(H)} \Psi_H^{1/c_i(H)}\Big)$$

for $p_{i-1} \le p \le p_i$, $i = 1, \ldots, r$.

Proof Fix G and i, and let $m = \min_{H \subseteq G} n^{-b_i(H)/c_i(H)} \Psi_H^{1/c_i(H)}$. Further, let H_0 and H_1 be subhypergraphs of G achieving, respectively, the maximum in $s_i(G)$ and the minimum in m. We will first show that $n^{q_{i-1}} \le m \le n^{q_i}$. Indeed,

$$m \le n^{-b_i(H_0)/c_i(H_0)} \Psi_{H_0}^{1/c_i(H_0)} \le n^{(-b_i(H_0)+v_{H_0}-\frac{e_{H_0}}{s_i})/c_i(H_0)} = n^{q_i}$$

and

$$m = n^{-b_i(H_1)/c_i(H_1)} \Psi_{H_1}^{1/c_i(H_1)} \geq n^{(-b_i(H_1)+v_{H_1}-\frac{e_{H_1}}{s_{i-1}})/c_i(H_1)} \geq n^{q_{i-1}},$$

because, by continuity, $f_i^H(q_{i-1}) = f_{i-1}^H(q_{i-1})$.

Hence, by (9) and the definition of m, for all $H \subseteq G$,

$$N(n, m, H) = \Theta(n^{b_i(H)} m^{c_i(H)}) = O(\Psi_H),$$

while $N(n, m, H_1) = \Theta(\Psi_{H_1})$, what was to be proved. □

Observe that $s_0 = m_G$ and $s_1 = \max_{H \subseteq G} \frac{e_H}{v_H - \alpha_1(H)}$. Moreover, since $q_{r-1} \leq k-1$, (8) yields that $s_{r-1} \leq \Delta_G$. Thus, in particular,

$$M_G(n, p) = \begin{cases} \Theta(1), & \text{if } p \leq n^{-1/m_G}, \\ \Theta(\min_{H \subseteq G} \Psi_H^{1/\alpha_1(H)}), & \text{if } n^{-1/m_G} \leq p \leq n^{-1/s_1}, \\ \Theta(n^k p^{\Delta_G}), & \text{if } p \geq n^{-1/\Delta_G}. \end{cases}$$

It would be nice to find similar explicit formulae for $M_G(n, p)$ in the range $n^{-1/s_1} \leq p \leq n^{-1/\Delta_G}$. This boils down, however, to finding explicit formulae for $N(n, m, G)$ in the corresponding range $n \leq m \leq n^{k-1}$, which, in turn, depends on the more explicit determination of $\alpha_q(H)$ for all $1 < q < k-1$.

Acknowledgements The authors would like to thank the referee for his or her careful reading.

References

Alon N (1981) On the number of subgraphs of prescribed type of graphs with a given number of edges. Isr J Math 38(1–2):116–130

Bollobás B (1981) Threshold functions for small subgraphs. Math Proc Cambr Philos Soc 90(2):197–206

Chung F, Graham R, Frankl P, Shearer J (1986) Some intersection theorems for ordered sets and graphs. J Comb Theory Ser A 43(1):23–37

Chvátal V (1983) Linear programming. Freeman, New York

Erdős P, Rényi A (1960) On the evolution of random graphs. Publ Math Inst Hung Acad Sci 5:17–61

Friedgut E, Kahn J (1998) On the number of copies of one hypergraph in another. Isr J Math 105:251–256

Janson S (1990) Poisson approximation for large deviations. Random Struct Algorithms 1:221–230

Janson S, Łuczak T, Ruciński A (1990) An exponential bound for the probability of nonexistence of a specified subgraph in a random graph. In: Karoński M, Jaworski J, Ruciński A (eds) Random graphs '87 (Proceedings, Poznań 1987). Wiley, New York, pp 73–87

Janson S, Łuczak T, Ruciński A (2000) Random graphs. Wiley, New York

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

Matoušek J (2008) Personal communication