ELSEVIER

Discrete Mathematics 171 (1997) 153–167

DISCRETE MATHEMATICS

The number of connected sparsely edged uniform hypergraphs

Michał Karoński [*,a,1], Tomasz Łuczak [b,1,2]

[a] Department of Discrete Mathematics, Adam Mickiewicz University, Poznań, Poland

[b] Mathematical Institute of the Polish Academy of Sciences

Received 20 April 1995; revised 30 January 1996

Abstract

Certain families of d-uniform hypergraphs are counted. In particular, the number of connected d-uniform hypergraphs with r vertices and $r+k$ hyperedges, where $k = o(\log r/\log\log r)$, is found.

1. Introduction

In this paper we are concerned with counting members of families of labelled d-uniform hypergraphs with a given number of vertices and hyperedges. The description of these families, although simple, is not very short, so we postpone precise statements of our main results (Theorems 8 and 9) until the last section of the article. Here we only recall shortly some of the known facts of a similar flavour concerning 'ordinary' graphs.

Let $c_2(n,k)$ denote the number of labelled graphs with n vertices and $n+k$ edges. Thus, for instance, $c_2(n,-1)$ is the number of labelled trees on n vertices equal to n^{n-2}. Connected graphs with exactly one cycle were counted by Katz [7] and Rényi [12], while the case $k=1$ was settled by Bagaev [1]. A substantial progress in the studies of the asymptotic behaviour of $c_2(n,k)$ for large k was made by Wright. In the sequence of papers [14–16] he studied generating functions of the number of various classes of connected graphs with n vertices and $n+k$ edges, proving, among others, the following result.

* Corresponding author. Address : Department of Mathematics and Computer Science, Emory University, Atlanta, GA 30322, USA; e-mail: michal@mathcs.emory.edu.

[1] Both authors were partially supported by KBN grant 2 1087 91 01.

[2] On leave from Department of Discrete Mathematics, Adam Mickiewicz University, Poznań, Poland.

PII S0012-365X(96)00076-3

Theorem 1. *If* $1 \leqslant k = o(n^{1/3})$ *then*

$$c_2(n,k) = (1 + O(\sqrt{k^3/n}))\sqrt{2\pi}\alpha_k \frac{3^k(k-1)!}{2^{5k}\Gamma(3k/2)} n^{n+(3k-1)/2} ,$$

where

$$\alpha_1 = \alpha_2 = \tfrac{5}{36} \quad \textit{and} \quad \alpha_{k+1} = \alpha_k + \sum_{i=1}^{k-1} \frac{\alpha_i \alpha_{k-i}}{(k+1)\binom{k}{i}} \quad \textit{for } k \geqslant 2 . \tag{1}$$

The problem of computing $c_2(n,k)$ was, in full generality, solved by Bender et al. who found in [2] the asymptotic value of $c_2(n,k)$ for every $k = k(n)$ for which $0 \leqslant k \leqslant \binom{n}{2} - n$ (the statement of their result is somewhat complicated, so we omit it here). Quick methods of estimation of $c_2(n,k)$ were developed also by Bollobás [3] and Łuczak [8]. Finally, we remark that for the last few years a powerful stimulus for investigating the behaviour of $c_2(n,k)$ has come from the random graph theory, where the value of $c_2(n,k)$ plays a crucial role in the studies of the phase transition phenomenon (see [4, 5]), and in several papers on random graphs, as [9–11] and in particular in the article of Janson et al. [6], the structure of a 'typical' connected graphs with n vertices and $n+k$ edges has also been examined.

Much less is known about the number of hypergraphs of a prescribed size. Up to our knowledge, the only result in this direction was proved by Selivanov [13], who counted generalized rooted forests (see Lemma 5 in this note) and connected uniform hypergraphs with at most one cycle. Our goal is to obtain a theorem analogous to that of Wright for uniform hypergraphs. Thus, in the next section we introduce some basic definitions, which naturally generalize graph properties to hypergraphs. Then we study the kernel of a hypergraph H: a small hypergraph obtained from H which captures the main features of its structure. The next part of the paper deals with 'clean' uniform hypergraphs, which turn out to be particularly easy to count. Then we show that most of the hypergraphs which are not too dense are, in fact, clean. As a consequence of this fact we get estimates for the size of different classes of complex d-uniform hypergraphs.

Let us also mention that, unlike arguments used by Wright [14–16], and Bender et al. [2] based on delicate analysis of the behaviour of naturally defined generating functions, our approach is purely combinatorial (however, in the proof of Lemma 7, we make use of Theorem 1).

2. The structure of hypergraphs

Let us start with a few simple definitions concerning hypergraphs: as a matter of fact most of them are rather straightforward generalizations of corresponding notions for graphs.

A *hypergraph* H is a pair (V,E), where the set of hyperedges E is the family of subset of the vertex set V. A sequence of $v_0 e_1 v_1 \dots e_k v_k$, where v_i are vertices of H,

e_i are its hyperedges, and $v_{i-1}, v_i \in e_i$ for $i = 1, 2, \ldots, k$, is called a *walk* of length k between v_0 and v_k. By a *component* of $H = (V, E)$ containing $v \in V$ we mean a subhypergraph which consists of all vertices v' and edges e' which belong to some walk containing v. H is *connected* if it has only one component.

The *excess* of a hypergraph $H = (V, E)$ is defined as

$$\mathrm{ex}(H) = \sum_{e \in E} (|e| - 1) - |V| \geqslant -1 .$$

A hypergraph is called a *forest* if all its components have excess -1. On the other hand, we say that a hypergraph is *complex* if the excess of each of its components is strictly positive.

A *degree* of a vertex $v \in V$ is the number hyperedges containing v. A hyperedge e is called *pendant* if all vertices of e, except at most one, are of degree one, i.e. belong to no other hyperedges of hypergraph. A hypergraph without pendant hyperedges is *smooth*. It is not hard to observe that if at least one component of a hypergraph H has a non-negative excess than H contains a unique maximal smooth subhypergraph which can be found in the process of 'peeling off' pendant edges from H. We call this subhypergraph the *core* of H and denote it by $\mathrm{cor}(H)$. Note that since the removing of a pendant hyperedge e together with $|e| - 1$ isolated vertices produced in this way does not change the excess of a graph, for every complex hypergraph H we have $\mathrm{ex}(H) = \mathrm{ex}(\mathrm{cor}(H))$.

As the core is obtained from the hypergraph by eliminating pendant vertices, the kernel is a result of compressing paths of the core. A *proper path* in H is defined as a walk $v_0 e_1 v_1 \ldots e_k v_k$ such that:

- all vertices $v_0, \ldots, v_k$ are different and all of them (including v_0 and v_k) have degree two in the hypergraph;
- all edges $e_1, \ldots, e_k$ are different and for every $i = 1, \ldots, k$, and every $v \in e \setminus \{v_{i-1}, v_i\}$, the degree of v is one.

A proper path is maximal if it is not contained in any other proper path. Now, for a complex hypergraph H, replace in the core of H each maximal proper path joining vertices v and v' by a *chain*, i.e. a new edge $\{v, v'\}$. Note that unlike $\mathrm{cor}(H)$ the hypergraph $\ker(H)$ obtained in this procedure, called the *kernel* of H, may *not* be a subhypergraph of H ($\ker(H)$ may contain some chains even when all hyperedges of H consists of more than two elements). Nevertheless, replacing proper paths by edges does not affect the excess of a hypergraph, so $\mathrm{ex}\ker(H) = \mathrm{ex}(\mathrm{cor}(H)) = \mathrm{ex}(H)$.

3. Clean hypergraphs

As we shall soon see the size of the kernel of a hypergraph depends not on its size but only on its excess. Thus, if the excess of a hypergraph is not too large, its kernel is a small hypergraph which reflects the most characteristic structural features of H.

Hence, the analysis of the structure of the kernel of a complex hypergraph will be crucial for our argument.

For a kernel K let K^- denote the hypergraph obtained from K by removing all chains. We say that a component of K^- is a *3-star*, if it consists of three hyperedges e_1, e_2, e_3 such that

$$e_1 \cap e_2 = e_1 \cap e_3 = e_2 \cap e_3 = \{v\}$$

and each of e_1, e_2, e_3 contains precisely one vertex which becomes the end of the chain in K. The kernel K of a hypergraph is *clean* if every component of K^- is either a 3-star, or an isolated hyperedge with precisely three vertices which are ends of chains in K. Finally, we call a hypergraph *clean* if its kernel is clean.

Clean kernels play an important role in counting complex hypergraphs because they maximize the number of chains. This is true for any complex hypergraphs of a given excess, but for simplicity we prove this fact only for d*-uniform* hypergraphs i.e. hypergraphs in which each hyperedge consists of precisely $d \geqslant 2$ vertices.

Claim 2. *If K is a kernel of a complex d-uniform hypergraph with excess k, then it has at most $3k$ chains, and this maximum is attained only for clean kernels.*

Furthermore, every kernel of such a complex d-uniform hypergraph has at most $2k(3d-2) \leqslant 6kd$ vertices.

Proof. Since the case when $d = 2$ is obvious, we assume that $d \geqslant 3$. For every kernel K of a complex d-uniform hypergraph with excess k which is not clean we define another hypergraph $g(K)$ such that $g(K)$ is also the kernel of some complex d-uniform hypergraph with excess k, but $g(K)$ has more chains and more vertices than K. Furthermore, the transformation g will have the property that for every kernel K there exists $i \leqslant 3k$ such that the kernel

$$g^i(K) = \underbrace{g(g(\ldots g(}_{i}K)\ldots))$$

is clean. Since each clean kernel of a hypergraph with excess k has precisely $3k$ chains and at most $2k(3d-2) \leqslant 6kd$ vertices, the assertion will follow.

Thus, for a given K, we must construct another kernel $g(K)$. We split the definition of $g(K)$ into several cases.

Case 1. K contains a vertex v which belong to precisely two hyperedges of size d.

Let v be the lexicographically first vertex of the above type, and let $e_1 \cap e_2 = \{v\}$, where $|e_1| = |e_2| = d$. Then, to obtain $g(K)$, add to K a new vertex w, label it by the first available label, replace e_2 by $e_3 = e_2 \setminus \{v\} \cup \{w\}$ and add additional chain $\{v, w\}$.

Case 2. Two hyperedges of K of size d share at least two vertices.

Let $e_1 \cap e_2 \supseteq \{v_1, v_2\}$ be first two hyperedges and first two vertices with this property. To get $g(K)$ we add new vertices $w_0, w_1, \ldots, w_d$, replace e_2 by $e_3 = e_2 \setminus \{v_1\} \cup \{w_0\}$, and add a new hyperedge $\{w_1, \ldots, w_d\}$ and chains $\{w_0, w_1\}$ and $\{w_2, w_3\}$.

Case 3. There exist hyperedges $e_1, \dots, e_5$, all of size d, such that

$$e_1 \cap e_2 \cap e_3 = \{v_1\}, e_1 \cap e_4 \cap e_5 = \{v_2\}, \quad \text{where} \quad v_1 \neq v_2 \, .$$

In this case we enlarge K by new vertices $w_0, \dots, w_{d-1}$, put $e_6 = e_1 \setminus \{v_1\} \cup \{w_0\}$ instead of e_1, add a new hyperedge $e_7 = \{v_2, w_1, \dots w_{k-1}\}$ and a chain $\{w_0, w_1\}$.

Case 4. There exists a hyperedge e, which is not isolated in K^-, such that two vertices of e, say v_1 and v_2, belong to chains.

Choose lexicographically first e with this property and, if possible, let v_1, $v_2 \in e$ be ends of a chain (if no chain is contained in e choose as v_1, v_2 lexicographically first vertices of e). To obtain $g(K)$ add $w_1, \dots, w_d$ new vertices to K together with a hyperedge $\{w_1, \dots, w_d\}$. Now consider two subcases:

(1) if $\{v_1, v_2\}$ is a chain replace it by $\{v_1, w_1\}$ and $\{w_2, w_3\}$,

(2) if K contains chains $\{v_1, v_1'\}$ and $\{v_2, v_2'\}$ replace them by $\{v_1, w_1\}$, $\{v_1', w_2\}$, $\{v_2', w_3\}$.

Case 5. There exists $\{v\}$ which belong to at least four hyperedges of size d.

Let $e_1 \cap e_2 \cap e_3 \cap e_4 = \{v_1\}$. Note that because Cases 1–4 do not apply, for each $i = 1, 2, 3, 4$, the hyperedge e_i contains precisely one vertex of degree at least four and one vertex which is the end of a chain – all other vertices of e_i are of degree one. Let $e_1 = \{v_1, v_2, \dots, v_d\}$, where v_2 is the end of a chain, and let $\{v', v''\}$ be any chain such that $\{v', v''\} \cap e_1 = \emptyset$ but $\{v', v''\} \cup (e_2 \cup e_3 \cup e_4) \neq \emptyset$. Now we add a new vertex w to K and replace e_1 by $e_5 = e_1 \setminus \{v_1\} \cup \{w\}$, and $\{v', v''\}$ by $\{w, v'\}$ and $\{v'', v_3\}$.

It is not hard to see that the above procedure applied repeatedly to any kernel will result in a clean kernel after a finite number of steps (and so, since in each step we increase the number of chains, after no more than $3k$ steps). Furthermore, an elementary analysis of Cases 1–5 shows that if K is the kernel of a d-uniform complex hypergraph, $g(K)$ is such a kernel as well (since the verification of this fact is easy but neither very short nor especially exciting we decide to omit details). Hence the number of vertices and chains in any kernel is bounded from above by the number of vertices and chains in some clean kernel, and Claim 2 follows. □

If a kernel is clean one can simplify its structure even further. Thus, in a clean kernel K, replace every component of K^- by a single vertex and join vertices v and w by as many edges as was the number of chains connecting components corresponding to v and w in K. In such a way we obtain from K a 3-regular multigraph which may contain loops and multiple edges. We shall consider vertices of this cubic multigraph to be *unlabelled* and call it the *kernel pattern* of a hypergraph.

4. Expanding kernel patterns

In this section we make a crucial step in the proof of Theorems 1 and 2: for a given 3-regular unlabelled multigraph G with $2k$ vertices we compute the number $c_d(m, G)$

of clean d-uniform complex hypergraphs H with the vertex set $[m] = \{1,\dots,m\}$ and the kernel pattern G.

Thus, in this sequel G is an unlabelled 3-regular graph on $2k$ vertices with λ_1 loops, λ_2 double edges and λ_3 triple edges, whose automorphism group has σ elements. We define the *compensation factor* of G setting

$$\mu(G) = 2^{\lambda_1+\lambda_2}6^{\lambda_3}\sigma\,. \tag{2}$$

(Note that, because we defined the kernel pattern as an unlabelled graph, $\mu(G)$ is slightly different from the compensation factor studied in [6].)

Lemma 3. *Let $d\geqslant 2$, $m = l(d-1)-k$ and $1\leqslant k = o(\sqrt{m})$ and G be an unlabelled cubic graph on $2k$ vertices. Then there exists*

$$v_d(m,G) = \frac{1+O(k^2/m)}{\mu(G)}\frac{m!\,l^{3k-1}}{(3k-1)!}\frac{(d-1)^{2k}}{[(d-2)!]^l}$$

smooth d-uniform hypergraphs $\hat{H}$ with the vertex set $[m] = \{1,\dots,m\}$, whose kernel patterns are isomorphic to G.

Remark. Throughout this note error term does not reflect the dependence on d which is always treated as a constant.

Proof of Lemma 3. We first consider the case when $d\geqslant 3$. Let us recall that then there are two types of vertices in a kernel pattern: some of them are obtained by contracting 3-stars while the others replace isolated hyperedges. Let us suppose that the kernel contains i 3-stars. Thus, to reconstruct the kernel from its pattern, we must choose vertices of G which will become centres of 3-stars in one of $\binom{2k}{i}$ possible ways, label them ($(m)_i$ possibilities) and pick hyperedges of the 3-stars in one of

$$\frac{1}{6^i}\binom{m-i}{d-1}\binom{m-i-(d-1)}{d-1}\cdots\binom{m-i-(3i-1)(d-1)}{d-1}$$
$$= \frac{1}{6^i}\frac{(m-i)!}{[(d-1)!]^{3i}[m-i-3i(d-1)]!}$$

ways. Finally, in each edge we choose a place when we attach a chain, which gives an additional factor $(d-1)^{3i}$. Now choose vertices of $2k-i$ isolated hyperedges (there are

$$\frac{1}{6^{2k-i}}\binom{m-i-3i(d-1)}{d}\cdots\binom{m-i-3i(d-1)-(2k-i-1)d}{d}$$
$$= \frac{1}{6^{2k-i}}\frac{[m-i-3i(d-1)]!}{(d!)^{2k-i}[m+2i-2d(k-i)]!}$$

ways to do so) and pick up the point of putting chains between them (we have $\binom{d}{3}^{2k-i}$ possibilities). Finally, we must decide how to put chains in our construction. If G is just a graph then it is enough to order already chosen ends of chains in every 3-stars and isolated hyperedges, which can clearly be done on 6^{2k} ways, but in the case of multigraphs this number decreases by a factor of $2^{\lambda_1}2^{\lambda_2}6^{\lambda_3}$.

After recovering the kernel we must expand it to $\hat{H}$, i.e. replace chains by the remaining $(m+2i-2d(k-i))$ vertices of $\hat{H}$, which correspond to $l-2(k+i)$ hyperedges. Thus, order all remaining vertices, split them into $3k$ nonempty paths in one of $\binom{l-2(k+i)-1}{3k-1}$ ways, and, since in each hyperedge of a proper path vertices of degree one are not ordered divide the product by $[(d-2)!]^{l-2k+1}$. Finally, replace chains of the kernel by the proper paths obtained in this way.

In order to conclude our argument, it is enough to observe that each hypergraph $\hat{H}$ whose kernel pattern is G appears as a result of our 'expanding' procedure precisely σ times. Thus, we arrive at the following formula for $v_d(m,G)$:

$$\begin{aligned} v_d(m,G) &= \sum_{i=0}^{2k}\binom{2k}{i}\frac{m!}{(m-i)!}\frac{1}{6^i}\frac{(m-i)!(d-1)^{3i}}{[(d-1)!]^{3i}[m-i-3i(d-1)]!} \\ &\quad\times\frac{[m-i-3i(d-1)]!\,[d(d-1)(d-2)]^{2k-i}}{6^{2k-i}(d!)^{2k-i}[m+2i-2d(k-i)]!}\frac{6^{2k}}{2^{\lambda_1}2^{\lambda_2}6^{\lambda_3}\sigma} \\ &\quad\times[m+2i-2d(k-i)]!\binom{l-2(k+i)-1}{3k-1}\frac{1}{[(d-2)!]^{l-2(k+i)}} \\ &= \frac{m!}{\mu(G)}\sum_{i=0}^{2k}\binom{2k}{i}\frac{[d(d-1)(d-2)]^{2k-i}}{(d!)^{2k-i}[(d-2)!]^{l-2(k+i)}}\binom{l-2(k+i)-1}{3k-1} \\ &= \frac{1+O(k^2/l)}{\mu(G)}\frac{m!}{[(d-2)!]^l}\binom{l}{3k-1}\sum_{i=0}^{2k}\binom{2k}{i}(d-2)^{2k-i} \\ &= \frac{1+O(k^2/l)}{\mu(G)}\frac{m!\,l^{3k-1}}{(3k-1)!}\frac{(d-1)^{2k}}{[(d-2)!]^l}. \end{aligned}$$

Calculations in the case $d=2$ are much simpler. We must choose labels for $2k$ vertices of degree three in the kernel pattern, divide the rest of vertices into $3k$ groups each of at least two vertices, and put them in place of paths of length three joining vertices of degree three in the kernel pattern. Thus,

$$v_2(m,G) = \frac{m!}{(m-2k)!}\frac{(m-2k)!}{2^{\lambda_1+\lambda_2}6^{\lambda_3}}\binom{l-3k-1}{3k-1}\frac{1}{\sigma} = \frac{1+O(k^2/l)}{\mu(G)}\frac{m!\,l^{3k-1}}{(3k-1)!}. \qquad \square$$

The next natural step is to expand a smooth hypergraph $\hat{H}$ to a complex hypergraph whose core is isomorphic to $\hat{H}$.

Lemma 4. *Let $d \geqslant 2$, $r = s(d-1)-k$ and $1 \leqslant k = o(r^{1/3})$ and G be an unlabelled cubic graph on $2k$ vertices. Then*

$$c_d(r,G) = \left(1+O\left(\sqrt{\frac{k^3}{r}}\right)\right)\frac{\sqrt{2\pi}}{\mu(G)}\frac{(d-1)^{k/2+1}}{[(d-2)!]^{k/(d-1)}}$$
$$\times \frac{1}{2^{3k/2}\Gamma(3k/2)}\left[\frac{e^{2-d}}{(d-2)!}\right]^{r/(d-1)} r^{r+(3k-1)/2}$$

clean d-uniform complex hypergraphs on the vertex set $[r]$ have the kernel pattern isomorphic to G.

Proof. Our proof will follow from Lemma 3 and the following combinatorial fact, shown by Selivanov [13], which generalized the well-known Cayley's formula for the number of rooted forests.

Lemma 5. *Let $f_{r,m}$ denote the number of all d-uniform forests on the vertex set $[r]$, which have precisely m components and are such that vertices $1,2,\ldots,m$, belong to different components. Furthermore, let t be the number of hyperedges in each such forests, so we have $r = t(d-1)+m$. Then*

$$f_{r,m} = \frac{m(r-m)!r^{t-1}}{t![(d-1)!]^t}.$$

Proof of Lemma 4. From Lemmas 3 and 5 we get

$$c_d(r,G) = \sum_{m=2kd+3k(d-1)}^{r}\binom{r}{m}v_d(m,G)f_{r,m}$$
$$= \sum_{m=2kd+3k(d-1)}^{r}\frac{r!}{m!(r-m)!}\frac{1+O(k^2/l)}{\mu(G)}\frac{m!\,l^{3k-1}}{(3k-1)!}\frac{(d-1)^{2k}}{[(d-2)!]^l}$$
$$\times\frac{m(r-m)!r^{s-l-1}}{(s-l)![(d-1)!]^{s-l}}$$
$$= \frac{r!}{\mu(G)}\frac{(d-1)^{2k}}{[(d-2)!]^s}\frac{1}{(3k-1)!}$$
$$\times\sum_{m=2kd+3k(d-1)}^{r}\frac{m\,l^{3k-1}r^{s-l-1}}{(s-l)!(d-1)^{s-l}}(1+O(k^2/l))$$
$$= \frac{r!}{\mu(G)}\frac{(d-1)^{2k}}{[(d-2)!]^s}\frac{1}{(3k-1)!}$$
$$\times\sum_{l=\lceil(5kd+2k)/(d-1)\rceil}^{s}\frac{l^{3k-1}[l(d-1)-k][s(d-1)-k]^{s-l-1}}{(s-l)!(d-1)^{s-l}}$$

$$\times(1+O(k^2/l))$$

$$=\frac{r!}{\mu(G)}\frac{(d-1)^{2k}}{[(d-2)!]^s}\frac{1}{(3k-1)!}$$

$$\times\sum_{l=\lceil(5kd+2k)/(d-1)\rceil}^{s}\frac{l^{3k}s^{s-l-1}}{(s-l)!}\exp\left(-\frac{k(s-l)}{s(d-1)}+O\left(\frac{k^2}{s}+\frac{k^2}{l}\right)\right)$$

$$=\frac{r!}{\mu(G)}\frac{(d-1)^{2k}}{[(d-2)!]^s}\frac{e^{-k/(d-1)}}{(3k-1)!}$$

$$\times\sum_{l=\lceil(5kd+2k)/(d-1)\rceil}^{s}\frac{l^{3k}s^{s-l-1}}{(s-l)!}(1+O(kl/s+k^2/l)).$$

Let

$$S=\sum_{l=\lceil(5kd+2k)/(d-1)\rceil}^{s}\frac{l^{3k}s^{s-l-1}}{(s-l)!}(1+O(kl/s+k^2/l))$$

Note that from Stirling's formula we get

$$\frac{s^{s-l-1}}{(s-l)!}=(1+O(1/(s-l))\frac{s^{s-l-1}e^{s-l}}{\sqrt{2\pi}(s-l)^{s-l+1/2}}$$

$$=(1+O(1/(s-l)))\frac{e^{s-l}}{s\sqrt{2\pi(s-l)}}\left(\frac{s}{s-l}\right)^{s-l}$$

$$=\frac{1}{s\sqrt{2\pi s}}\exp\left(s-\frac{l^2}{2s}+O\left(\frac{l^3}{(s-l)^2}\right)\right).$$

Thus

$$S=\frac{e^s}{s\sqrt{2\pi s}}\sum_{l=\lceil(5kd+2k)/(d-1)\rceil}^{s}l^{3k}\exp\left(-\frac{l^2}{2s}+O\left(\frac{kl}{s}+\frac{k^2}{l}+\frac{l^3}{(s-l)^2}\right)\right).$$

One can easily check that the main contribution to the above sum comes from the terms l for which $l\sim\sqrt{3ks}$ and thus the error term is, in fact, equal to $O(\sqrt{k^3/s})$. Moreover,

$$\sum_{l=\lceil(5kd+2k)/(d-1)\rceil}^{s}l^{3k}e^{-l^2/2s}(1+O(\sqrt{k^3/s}))$$

$$=(1+O(\sqrt{k^3/s}))\int_0^\infty x^{3k}e^{-x/2s}\,dx+O(1)$$

$$=(1+O(\sqrt{k^3/s}))(2s)^{3k/2}\sqrt{\frac{s}{2}}\Gamma\left(\frac{3k+1}{2}\right)+O(1)$$

$$=(1+O(\sqrt{k^3/s}))\sqrt{2s\pi}\left(\frac{s}{2}\right)^{3k/2}\frac{(3k-1)!}{\Gamma(3k/2)}$$

since $\Gamma(3k) = \frac{2^{3k-1}}{\sqrt{\pi}}\Gamma(3k/2)\Gamma((3k+1)/2)$. Hence,

$$S = \left(1 + \mathrm{O}\left(\sqrt{\frac{k^3}{s}}\right)\right)\frac{\mathrm{e}^s}{s}\left(\frac{s}{2}\right)^{3k/2}\frac{(3k-1)!}{\Gamma(3k/2)}, \tag{3}$$

and thus, keeping in mind that $r = s(d-1) - k$, we get

$$\begin{aligned} c_d(r,G) &= \left(1 + \mathrm{O}\left(\sqrt{\frac{k^3}{s}}\right)\right)\frac{r!}{\mu(G)}\frac{(d-1)^{2k}}{[(d-2)!]^s}\frac{\mathrm{e}^{-k/(d-1)}}{(3k-1)!}\frac{\mathrm{e}^s}{s}\left(\frac{s}{2}\right)^{3k/2}\frac{(3k-1)!}{\Gamma(3k/2)} \\ &= \left(1 + \mathrm{O}\left(\sqrt{\frac{k^3}{r}}\right)\right)\frac{\mathrm{e}^{-k/(d-1)}}{\mu(G)}\frac{(d-1)^{2k+1}}{\Gamma(3k/2)} \\ &\quad \times \frac{r!}{[(d-2)!]^{(r+k)/(d-1)}}\frac{\mathrm{e}^{(r+k)/(d-1)}}{r+k}\left(\frac{r+k}{2(d-1)}\right)^{3k/2} \\ &= \frac{1 + \mathrm{O}(\sqrt{\frac{k^3}{r}})}{\mu(G)}\frac{(d-1)^{k/2+1}}{[(d-2)!]^{k/(d-1)}}\frac{1}{2^{3k/2}\Gamma(3k/2)} \\ &\quad \times \frac{\mathrm{e}^{r/(d-1)}}{[(d-2)!]^{r/(d-1)}} r^{3k/2-1} r! \\ &= \left(1 + \mathrm{O}\left(\sqrt{\frac{k^3}{r}}\right)\right)\frac{\sqrt{2\pi}}{\mu(G)}\frac{(d-1)^{k/2+1}}{[(d-2)!]^{k/(d-1)}}\frac{1}{2^{3k/2}\Gamma(3k/2)} \\ &\quad \times \left[\frac{\mathrm{e}^{2-d}}{(d-2)!}\right]^{r/(d-1)} r^{r+(3k-1)/2}. \quad \square \end{aligned}$$

5. Excluding unclean hypergraphs

In the previous part of the paper we dealt with clean hypergraphs, now we show that, if k is small, there are only a few unclean hypergraphs with complexity k.

Lemma 6. *Let $d \geqslant 2$, $m = l(d-1) - k$ and $1 \leqslant k = \mathrm{o}(\log m/\log\log m)$. Then, for every m which is large enough, not more than*

$$\frac{m!}{m^{1/3}}\frac{l^{3k-1}}{[(d-2)!]^l}$$

smooth d-uniform complex hypergraphs $\hat{H}$ with the vertex set $[m]$ and the excess k are not clean.

Moreover, if $r = s(d-1) - k$ and $1 \leqslant k = \mathrm{o}(\log r/\log\log r)$, then there are less than

$$\left[\frac{\mathrm{e}^{2-d}}{(d-2)!}\right]^{r/(d-1)}\frac{r^{r+(3k-1)/2}}{r^{1/3}}$$

unclean d-uniform complex hypergraphs with the vertex set [r] and s hyperedges, provided r is large enough.

Proof. In order to verify the above statement we must just repeat arguments from the previous section. Let K be a kernel of some unclean d-uniform graph with a vertices and b links. Since K is a kernel of an unclean hypergraph, from Claim 2 we know that $a \leqslant 6kd$ and $b \leqslant 3k-1$. Thus, the number of smooth complex d-uniform graphs $\hat{H}$ with $\ker(\hat{H})$ isomorphic to K can be crudely bounded from above by

$$\frac{m!}{(m-a)!}(m-a)!\binom{l+b-1}{b-1}\frac{(6kd)^b}{[(d-2)!]^{l-a-k}} \leqslant \frac{m^{0.01}m!\,l^{b-1}}{[(d-2)!]^l}, \tag{4}$$

where here and below we claim that all inequalities holds only for values m, l, r and s which are large enough.

Now let us count the number $c_d(r,K)$ of all d-uniform hypergraphs which have the kernel isomorphic to K. Similarly, as in the proof of Lemma 4, (4) implies that $c_d(r,K)$ is bounded from above by

$$c_d(r,K) \leqslant \sum_{m=a}^{r}\binom{r}{m}\frac{m^{0.01}m!\,l^{b-1}}{[(d-2)!]^l}\frac{m(r-m)!\,r^{s-l-1}}{(s-l)![(d-1)!]^{s-l}}$$

$$\leqslant \frac{r!}{[(d-2)!]^s}\sum_{m=a}^{r}\frac{l^{b-0.98}s^{s-l-1}}{(s-l)!}.$$

Repeating the argument which led us to (3) and using the fact that $r = s(d-1)-k$, we conclude that

$$c_d(r,K) \leqslant \frac{(3k)!\,r!}{[(d-2)!]^s}\frac{e^s}{s}\left(\frac{s}{2}\right)^{0.5b-0.49} \leqslant \left[\frac{e^{2-d}}{(d-2)!}\right]^{r/(d-1)} r^{r+0.5b-0.98}$$

$$\leqslant \left[\frac{e^{2-d}}{(d-2)!}\right]^{r/(d-1)} r^{r+(3k-1)/2-0.47}. \tag{5}$$

To complete the proof we must estimate the number of nonisomorphic candidates for kernels of unclean complex d-uniform hypergraphs of complexity k. Since they have at most $6dk$ vertices, and thus not more than $6dk+k$ hyperedges of size d, there are not more than

$$\binom{6dk}{d}^{6dk+k} \leqslant (17k)^{7d^2k}$$

ways of choosing them. Furthermore, there are not more than

$$(36k^2d^2)^{3k-1} \leqslant 50\,000k^{6k}d^{6k}$$

possibilities of placing at most $3k-1$ chains which could appear in the kernel. But if $k = o(\log n/\log\log n)$ then clearly

$$50\,000d^{6k}(17k)^{7d^2k+6k} \leqslant n^{0.01}$$

so the assertion follows from (4) and (5). □

6. Counting complex hypergraphs

From the previous sections we know that when the excess is small compared to the size of the hypergraph, the contribution to the number of complex d-uniform hypergraphs which comes from unclean hypergraphs is negligible. Furthermore, we have estimated the number of clean uniform hypergraphs with a given kernel pattern. Thus, to count hypergraphs in a subfamily of complex hypergraphs, we need only to know the weighted sum of kernel patterns of members of the family. The following lemma provides its value for families of all complex hypergraphs and of all connected hypergraphs with excess k.

Lemma 7. *Let $\mathscr{U}_{2k}$ be the family of all unlabelled cubic graphs on $2k$ vertices, and let $\mathscr{C}_{2k}$ be the subfamily of $\mathscr{U}_{2k}$ which consists of all connected graphs of this kind. Then*

$$\sum_{G\in\mathscr{U}_{2k}} \frac{1}{\mu(G)} = \frac{(6k)!}{(3k)!\,(2k)!}\frac{1}{2^{5k}3^{2k}} \tag{6}$$

and

$$\sum_{G\in\mathscr{C}_{2k}} \frac{1}{\mu(G)} = \alpha_k\left(\frac{3}{2}\right)^k (k-1)!\,, \tag{7}$$

where weights $\mu(G)$ are defined as in (2) and coefficients α_k are determined by (1).

Proof. Let $V_1,\dots V_{2k}$ be disjoint sets, each containing three distinguishable points. Consider the family $\mathscr{M}$ of all $(6k)!/(3k)!\,2^{3k}$ possible perfect matchings of the set $V = \bigcup_{i=1}^{2k} V_i$. Furthermore, for every matching $M \in \mathscr{M}$ let $G(M)$ be a multigraph obtained from M by contracting all sets V_i, i.e. G is labelled graph with the vertex set $[2k]$ such that the vertices i,j, $1\leqslant i\leqslant j\leqslant 2k$, of $G(M)$ are joined by the same number of edges as the number of edges of M between V_i and V_j. We claim that in the family $\{G(M) : M \in \mathscr{M}\}$ each labelled multigraph G appears precisely $6^{2k-\lambda_3}2^{-\lambda_1-\lambda_2}$ times, where λ_1, λ_2 and λ_3 denote the number of loops, double and triple edges of G, respectively. Indeed, for a given matching $M \in \mathscr{M}$ we can modify M by 'switching' vertices in sets $V_1,\dots,V_{2k}$ in such a way that $G(M)$ remains unchanged. Thus, if G is a graph without loops, one matching M gives $(3!)^{2k}$ other matchings M' with $G(M) = G(M')$. When G is a multigraph, this number drops down by a factor of $(2!)^{\lambda_1+\lambda_2}(3!)^{\lambda_3}$ because some switches result in the same matchings. Now to complete the proof of (6) it is enough to observe that from each unlabelled graph G on $2k$ vertices can be labelled in precisely $(2k)!/\sigma$ ways, where σ is the number of elements in the automorphism group of G.

In order to show the second equation note that if we choose r much larger than k then Lemmas 4 and 6 imply that

$$c_2(r,k) = \left(1 + O\left(\sqrt{\frac{k^3}{r}}\right)\right)\left(\sum_{G\in\mathscr{C}_k}\frac{1}{\mu(G)}\right)\frac{\sqrt{2\pi}}{2^{3k/2}\Gamma(3k/2)}\, r^{r+(3k-1)/2}\,.$$

On the other hand the value of $c_2(r,k)$ is given in Wright's Theorem 1. Since we can make an error term as small as we wish by picking r large enough (7) follows. □

Remark. It is not hard to show that when k is large the contribution to the first sum coming from nonconnected multigraphs is negligible, i.e. the two sums must be asymptotically equal. This provides yet another proof for the fact which has already been observed independently by several authors (see [6, p. 262])

$$\lim_{k\to\infty} \alpha_k = \lim_{k\to\infty} \frac{(6k)!}{(3k)!(2k)!(k-1)!}\frac{1}{2^{4k}3^{3k}} = \frac{1}{2\pi}.$$

Now from Lemmas 3, 4, 6 and 7 and Stirling's formula we can easily obtain the numbers of complex d-uniform hypergraphs with a small excess.

Theorem 8. *Let $d \geqslant 2$ and $1 \leqslant k = o(\log m/\log\log m)$. Then the number of smooth complex d-uniform hypergraphs with m vertices and excess k is given by*

$$\bar{v}_d(m,k) = \left(1 + O\left(\sqrt{\frac{k^2}{m}}\right)\right)\frac{\sqrt{2}\pi}{(d-1)^{k-1}2^{5k}3^{2k}}\frac{(6k)!}{(3k)!(2k)!(3k-1)!}$$

$$\times\frac{1}{[(d-2)!]^{k/(d-1)}}\left[\frac{e^{1-d}}{(d-2)!}\right]^{m/(d-1)} m^{m+3k-1/2},$$

while for the number of connected smooth d-uniform hypergraphs with m vertices and excess k we have

$$v_d(m,k) = \left(1 + O\left(\sqrt{\frac{k^2}{m}}\right)\right)\frac{\sqrt{2}\pi\alpha_k}{(d-1)^{k-1}}\frac{3^k}{2^k}\frac{(k-1)!}{(3k-1)!}$$

$$\times\frac{1}{[(d-2)!]^{k/(d-1)}}\left[\frac{e^{1-d}}{(d-2)!}\right]^{m/(d-1)} m^{m+3k-1/2},$$

where α_k is defined as in (1).

In particular, if $k = o(\log m/\log\log m)$ but $k \to \infty$ then

$$\bar{v}_d(m,k) = (1 + O(1/k))v_d(m,k)$$

$$= \left(1 + O\left(\sqrt{\frac{k^2}{m}}\right)\right)\sqrt{\frac{3}{2\pi}}(d-1)$$

$$\times\left[\frac{1}{18(d-1)[(d-2)!]^{1/(d-1)}}\right]^k\left[\frac{e^{1-d}}{(d-2)!}\right]^{m/(d-1)} m^{m+3k-1/2}.$$

Theorem 9. *Let $d \geqslant 2$ and $1 \leqslant k = o(\log r/\log\log r)$. Then for the number of complex d-uniform hypergraphs with r vertices and excess k we have*

$$\bar{c}_d(r,k) = \left(1 + O\left(\sqrt{\frac{k^3}{r}}\right)\right)\frac{\sqrt{2\pi}}{\Gamma(3k/2)}\frac{(6k)!}{(2k)!(3k)!}\frac{1}{2^{13k/2}3^{2k}}$$

$$\times\frac{(d-1)^{k/2+1}}{[(d-2)!]^{k/(d-1)}}\left[\frac{e^{2-d}}{(d-2)!}\right]^{r/(d-1)} r^{r+(3k-1)/2},$$

and counting connected d-uniform hypergraphs with r vertices and excess k gives

$$c_d(r,k) = \left(1 + O\left(\sqrt{\frac{k^3}{r}}\right)\right)\frac{\sqrt{2\pi}\alpha_k}{\Gamma(3k/2)}\left(\frac{9}{32}\right)^{k/2}(k-1)!$$

$$\times\frac{(d-1)^{k/2+1}}{[(d-2)!]^{k/(d-1)}}\left[\frac{e^{2-d}}{(d-2)!}\right]^{r/(d-1)} r^{r+(3k-1)/2},$$

where α_k is defined as in (1).

In particular, if $k = o(\log r/\log\log r)$ but $k \to \infty$ then

$$\bar{c}_d(r,k) = (1 + O(1/k))c_d(r,k) = \left(1 + O\left(\sqrt{\frac{k^3}{r}}\right)\right)\sqrt{\frac{3}{4\pi}}\left(\frac{e}{12k}\right)^{k/2}$$

$$\times\frac{(d-1)^{k/2+1}}{[(d-2)!]^{k/(d-1)}}\left[\frac{e^{2-d}}{(d-2)!}\right]^{r/(d-1)} r^{r+(3k-1)/2}.$$

One may wonder whether Theorems 8 and 9 remain valid also for larger values of k. Clearly, the only problem is the elimination of unclean hypergraphs: the estimate for the number of clean hypergraphs holds as far as $k = o(\sqrt{m})$ and $k = o(r^{1/3})$. It is not very hard to replace the assumption $k = o(\log r/\log\log r)$ by $k = o(\log r)$. Indeed, one can estimate $c_d(r,K)$ in the proof of Lemma 6 more carefully, and show that if kernels K and K' has, respectively, b and b' chains, where $b \leqslant b' \leqslant 3k$, then

$$c_d(r,K) = O((6d)^{6dk}(3k)^{b'-b}c_d(r,K')r^{(b-b')/2}),$$

i.e. when the number of chains drops by one $c_d(K,r)$ decreases roughly by $\sqrt{r}$. On the other hand from our 'algorithmic' proof of Claim 2 it follows that the number of candidates for kernels with b chains differs from the number of possible kernels with b' chains by a factor which is a polynomial of k whose degree grows linearly with $b - b'$. Hence, if $k = o(\log r)$, one can prove that the contribution coming from unclean hypergraphs is negligible and the assertion follows.

Nonetheless, we believe that estimates given in Theorems 8 and 9 are valid whenever $k = o(\sqrt{m})$ and $k = o(r^{1/3})$. However, even if $k = o(m^{\varepsilon})$ or $k = o(r^{\varepsilon})$ for some $\varepsilon > 0$, the proof would require considerably more work, since counting unclean hypergraphs one must control factors of order d^k. Thus, probably one should rather look for a general argument, similar to that of Bender et al. [2], which could capture the behaviour of $v_d(m,k)$ and $c_d(r,k)$ in the whole range of k.

References

[1] G.N. Bagaev, Sluchaĭnye grafy so stepen'ıû svıâznosti 2 Diskretnyĭ Analiz 22 (1973) 3–14

[2] E.A. Bender, E.A. Canfield and B.D. McKay, The asymptotic number of labeled connected graphs with a given number of vertices and edges, Random Struct. Algorithms 1 (1990) 127–169

[3] B. Bollobás, The evolution of sparse graphs, in: B. Bollobás, Ed., Graph Theory and Combinatorics, Proc. Cambridge Combinatorial Conf. In honour of Paul Erdös (Academic Press, New York, 1984) 35–57.

[4] B. Bollobás, The evolution of random graphs, Trans. Amer. Math. Soc. 286 (1984) 257–274.

[5] B. Bollobás, Random Graphs (Academic Press, London 1985).

[6] S. Janson, D.E. Knuth, T. Łuczak and B. Pittel, The birth of the giant component, Random Struct. Algorithms 4 (1993) 233–358.

[7] L. Katz, Probability of indecomposability of a random mapping, Ann. Math. Statist. 26 (1955) 512–517.

[8] T. Łuczak, On the number of sparse connected graphs, Random Struct. Algorithms 1 (1990) 171–173.

[9] T. Łuczak, Cycles in a random graph near the critical point, Random Struct. Algorithms 2 (1991) 421–439

[10] T. Łuczak, B. Pittel and J.C. Wierman, The structure of the random graph at the double-jump threshold, Trans. Amer. Math. Soc. 341 (1994) 721–748.

[11] T. Łuczak and J.C. Wierman, The chromatic number of random graphs at the double-jump threshold, Combinatorica 9 (1989) 39–49.

[12] A. Rényi, On connected graphs I Magyar Tud. Akad. Mat. Kutató Int. Közl 4 (1959) 385–388.

[13] B.I. Selivanov, Perechislenie odnorodnykh hipergrafov c prostoĭ ciklovoĭ strukturoĭ, Kombinatornyĭ Analiz 2 (1972) 60–67.

[14] E.M. Wright, The number of connected sparsely edges graphs, J. Graph Theory 1 (1977) 317–334.

[15] E.M. Wright, The number of connected sparsely edges graphs. II. Smooth graphs and blocks, J. Graph Theory 2 (1978) 299–305.

[16] E.M. Wright, The number of connected sparsely edges graphs. III. Asymptotic results, J. Graph Theory 4 (1980) 393–408