ELSEVIER

Journal of Computational and Applied Mathematics 142 (2002) 125–135

JOURNAL OF COMPUTATIONAL AND APPLIED MATHEMATICS

www.elsevier.com/locate/cam

The phase transition in a random hypergraph

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

[a] *Department of Discrete Mathematics, Adam Mickiewicz University, Matejki 48/49, 60-769 Poznań, Poland*

[b] *Department of Mathematics and Computer Science, Emory University, Atlanta, GA, USA*

Received 26 October 2000; received in revised form 12 January 2001

Abstract

We show that in the evolution of the random d-uniform hypergraph $\mathbb{G}^d(n,M)$ the phase transition occurs when $M = n/d(d-1) + O(n^{2/3})$. We also prove local limit theorems for the distribution of the size of the largest component of $\mathbb{G}^d(n,M)$ in the subcritical and in the early supercritical phase. © 2002 Elsevier Science B.V. All rights reserved.

MSC: 05C80; 05C65

1. Introduction

A *hypergraph* $\mathscr{H}$ is a pair $(V,\mathscr{E})$, where V denotes the set of vertices of $\mathscr{H}$ and $\mathscr{E}$ is a family of subsets of V called edges. We say that $\mathscr{H}$ is *d-uniform*, or, simply, *uniform*, if $|E| = d$ for every $E \in \mathscr{E}$. The *random hypergraph* $\mathbb{G}^d(n,M)$ is defined as a hypergraph chosen uniformly at random from the family of all $\binom{\binom{n}{d}}{M}$ d-uniform labelled hypergraphs with vertex set $[n] = \{1,2,\ldots,n\}$ and M edges. (Note that $\mathbb{G}(n,M) = \mathbb{G}^2(n,M)$, i.e., for $d = 2$ the notion of a 2-uniform random hypergraph coincides with that of the *random graph*.) We study the behaviour of $\mathbb{G}^d(n,M)$ as $n \to \infty$, where the number of edges $M = M(n)$ may vary as a function of n. In particular, we say that for a given function $M = M(n)$ graph property holds for $\mathbb{G}^d(n,M)$ *asymptotically almost surely*, or, briefly, *a.a.s.*, if the probability that $\mathbb{G}^d(n,M)$ has this property tends to 1 as $n \to \infty$.

One of the most striking results of the seminal paper on random graphs by Erdős and Rényi [4] was the discovery of the abrupt change in the structure of $\mathbb{G}(n,M)$, when $M = cn$ and $c \sim \frac{1}{2}$. They proved that if $c < \frac{1}{2}$, then a.a.s. $\mathbb{G}(n,M)$ consists of many small components, while for $c > \frac{1}{2}$,

* Corresponding author. Department of Discrete Mathematics, Adam Mickiewicz University, Matejki 48/49, 60-769 Poznan, Poland.

E-mail address: tomasz@amu.edu.pl (T. Łuczak).

[1] Partially supported by KBN grant 2 P03A 021 17.

PII: S0377-0427(01)00464-2

it a.a.s. has one large component which dominates the whole graph. The component structure for random hypergraphs was studied by Schmidt-Pruzan and Shamir [9]. In particular, they proved that if $d \geqslant 2$, $M = cn$ and $c < 1/d(d-1)$, then a.a.s. the largest component of $\mathbb{G}^d(n,M)$ is of the order $\log n$, for $c = 1/d(d-1)$ it has $\Theta(n^{2/3})$ vertices, and finally, when $c > 1/d(d-1)$, a.a.s. $\mathbb{G}^d(n,M)$ contains the unique giant component of size $\Theta(n)$. Thus, as in the case of the random graph, the largest component of $\mathbb{G}^d(n,M)$ grows rapidly when the number of edges is roughly $n/d(d-1)$.

The study of the behaviour of the component structure of $\mathbb{G}(n,M)$ when $2M/n \to 1$ is much more difficult. Erdős and Rényi [4] suggested that in this case the largest component has a.a.s. $\Theta(n^{2/3})$ vertices. The fact that it is not true was first observed by Bollobás, who in his pioneering work [2] (see also [3, Chapter VI]) precisely described the structure of $\mathbb{G}(n,M)$ for $2M/n \to 1$ (his results were later supplemented by Łuczak [8]). Thus, in *the subcritical phase*, when $M = n/2 - m$ and $m/n^{2/3} \to \infty$, the largest components of $\mathbb{G}(n,M)$ have roughly similar size while for $M = n/2 + m$, where $m/n^{2/3} \to \infty$ as $n \to \infty$ (*supercritical phase*) a.a.s. $\mathbb{G}(n,M)$ contains a unique largest component significantly larger than all its competitors. (For a more detailed description of the phase transition phenomenon in $\mathbb{G}(n,M)$ see [5], [6, Chapter 5].)

In this paper, we study the asymptotic behaviour of the random hypergraph near the critical range, i.e., for $M \sim n/d(d-1)$. It turns out that in the subcritical phase, now determined by the condition that $M = n/d(d-1) - m$ and $m/n^{2/3} \to \infty$ as $n \to \infty$, the structure of $\mathbb{G}^d(n,M)$ is not hard to analyze. In this case, a.a.s. $\mathbb{G}^d(n,M)$ consists of hypertrees and unicyclic components and one can obtain a local limit distribution for the size of the largest component using elementary method of moments. The problem of describing the component structure of $\mathbb{G}^d(n,M)$ when $M = n/d(d-1) + m$ and $m/n^{2/3} \to \infty$, seems to be a much more challenging task. However, in Section 3 we observe that the asymptotic distribution of the size of the largest component can be deduced from the result on the number of connected hypergraphs with a given number of vertices and edges. As a matter of fact in this way one can obtain a surprisingly precise local limit result on the joint distribution of the two random variables which measure the number of vertices and edges in the largest component (Theorems 8 and 9), which has not been known even for random graphs, when $d = 2$.

2. Connected hypergraphs

Let H be a d-uniform hypergraph with r vertices and s edges. Define the *excess* of H as

$$\mathrm{ex}(H) = (d-1)s - r.$$

Note that from the definition of the excess it follows that if $\mathrm{ex}(H) = k$, then $(d-1) | (r+k)$. Observe also that if H is connected, then $\mathrm{ex}(H) \geqslant -1$. A connected hypergraph H for which $\mathrm{ex}(H) = -1$ we call a *hypertree*, or, briefly, a *tree*; if for a connected H we have $\mathrm{ex}(H) = 0$ we say that H is *unicyclic*. Finally, we call a connected hypergraph H *complex* if its excess is positive.

Let $C_d(s,k)$ denote the number of connected d-uniform hypergraphs with $r = s(d-1) - k$ vertices and s edges. In the case of graphs, i.e., when $d = 2$, the behaviour of $C_2(s,k)$ has been thoroughly studied by many authors, and finally settled down by Bender et al. (see [1] and references therein).

For $d \geqslant 2$ the value of $C_d(s,-1)$ is given by the following result (see [10]). We remark that all asymptotic estimates in this note are made under the assumption that d is fixed, i.e., the hidden constants in $O(\cdot)$ may, and typically do, depend on d.

Lemma 1. *Let $r = s(d-1)+1$. Then the number of connected d-uniform hypertrees with r vertices, s edges is given by*

$$C_d(s,-1) = \frac{[(s(d-1)]![s(d-1)+1]^{s-1}}{s![(d-1)!]^s}.$$

In particular, if $s \to \infty$, then

$$C_d(s,-1) = \left(1 + O\left(\frac{1}{s}\right)\right)\frac{1}{\sqrt{d-1}}\frac{s^{s(d-1)-1}}{e^{s(d-2)-1/(d-1)}}\left[\frac{(d-1)^{d-1}}{(d-2)!}\right]^s.$$

Selivanov [10] gave also the following formula for $C_d(s,0)$.

Lemma 2. *The number of connected d-uniform hypergraphs with $r = s(d-1)$ vertices and s edges is given by*

$$C_d(s,0) = \frac{[s(d-1)]!}{2[(d-2)!]^s s^{s-1}}\sum_{j=2}^{s}\frac{1}{s^j(s-j)!}.$$

Thus, for $s \to \infty$,

$$C_d(s,0) = \left(1 + O\left(\frac{1}{s}\right)\right)\sqrt{\frac{\pi(d-1)}{8}}\frac{s^{s(d-1)-1/2}}{e^{s(d-2)}}\left[\frac{(d-1)^{d-1}}{(d-2)!}\right]^s.$$

Finally, for a given d and $k = o(\log s/\log\log s)$, the asymptotic value of $C_d(s,k)$ was determined by the following result of Karoński and Łuczak [7].

Lemma 3. *Let $d \geqslant 2$ and let $k = k(s)$ be a function of s such that $k \to \infty$ but $k\log\log s/\log s \to 0$ as $s \to \infty$. Then*

$$C_d(s,k) = \left(1 + O\left(\frac{1}{k} + \frac{k^2}{s} + \sqrt{\frac{k^3}{r}} + \frac{k^{100d^2k}}{r}\right)\right)\sqrt{\frac{3}{4\pi}}\left(\frac{e}{12k}\right)^{k/2}$$
$$\times\frac{(d-1)^{s(d-1)+k+1/2}}{[(d-2)!]^s}s^{s(d-1)+(k-1)/2}e^{s(2-d)-k/(d-1)}.$$

3. Subcritical phase

As in the case of the random graph $\mathbb{G}(n,M)$, the random hypergraph $\mathbb{G}^d(n,M)$ has a particularly simple structure whenever $M = n/d(d-1) - m$, and $m/n^{2/3} \to \infty$.

Theorem 4. *Let $M = n/d(d-1) - m$, where $m/n^{2/3} \to \infty$ as $n \to \infty$. Then, a.a.s. $\mathbb{G}^d(n,M)$ consists of hypertrees and unicyclic components.*

Proof. We say that a sequence of edges $e_1,\dots,e_t$, $t\geqslant 1$, is a *path* if $|e_i\cap e_{i+1}|=1$ for $i=1,2,\dots,t-1$, and $e_i\cap e_j=\emptyset$ whenever $|i-j|\geqslant 2$. Our argument is based on the observation that any component of a hypergraph which is neither a hypertree, nor a unicyclic component contains a structure of one of the two following types.

Type 1: There is a path $e_1,\dots,e_t$, $t\geqslant 1$, and a edge f such that $f\cap e_1\neq\emptyset$, $f\cap e_t\neq\emptyset$ and

$$\left|f\cap\bigcup_{i=1}^{t}e_i\right|\geqslant 3.$$

Type 2: There is a path $e_1,\dots,e_{t-1}$, $t\geqslant 2$, and edges f_1,f_2 such that $f_1\cap e_1\neq\emptyset$, $f_2\cap e_{t-1}\neq\emptyset$ and

$$\left|f_j\cap\bigcup_{i=1}^{t-1}e_i\right|\geqslant 2\quad\text{for } j=1,2.$$

Observe that the number of hypergraphs $w(t)$ of one of the above types with precisely $t+1$ edges which are contained in the complete d-uniform hypergraph on n vertices is bounded above by

$$\begin{aligned} w(d)&\leqslant\frac{d}{d-1}\binom{n}{d}\left[(d-1)\binom{n}{d-1}\right]^{t-1}td^3\binom{n}{d-3}\\ &\quad+\frac{d}{d-1}\binom{n}{d}\left[(d-1)\binom{n}{d-1}\right]^{t-2}d^4t^2\binom{n}{d-2}^2\\ &\leqslant n^{(d-1)(t+1)-1}\frac{8t^2d^4}{[(d-2)!]^t}. \end{aligned}$$

Let Y denote the number of structures of types 1 and 2 which are contained in $\mathbb{G}^d(n,M)$, where $M=n/d(d-1)-m$ and $m/n^{2/3}\to\infty$. Then,

$$\mathbb{P}(Y>0)\leqslant\mathbb{E}Y=\sum_{r=1}^{n+1}w(d)\binom{\binom{n}{d}-t-1}{M-t-1}\Big/\binom{\binom{n}{d}}{M}.$$

Observe that, for t large enough,

$$\begin{aligned} &\binom{\binom{n}{d}-t-1}{M-t-1}\Big/\binom{\binom{n}{d}}{M}\\ &\quad\leqslant\left(\frac{M-t-1}{\binom{n}{d}-t-1}\right)^t\leqslant\frac{[(d-2)!]^{t+1}}{n^{(d-1)(t+1)}}\left(1-\frac{m+t}{n}\right)^{t+1}\leqslant\frac{[(d-2)!]^{t+1}}{n^{(d-1)(t+1)}}\exp\left(-\frac{mt}{n}\right). \end{aligned}$$

Hence,

$$\mathbb{P}(Y>0) \leqslant 8d^5 \sum_{t=1}^{n+1} \frac{t^2}{n} \exp\left(-\frac{mt}{n}\right) \leqslant 8d^5 \int_0^\infty \frac{x^2}{n} \mathrm{e}^{-mx/n}\,\mathrm{d}x = 16d^5 \frac{n^2}{m^3}.$$

Since $m/n^{2/3} \to \infty$, the above sum tends to 0 as $n \to \infty$, i.e., a.a.s. $Y=0$ and the assertion follows. $\square$

In order to study the phase transition phenomenon, we need precise estimates on the number of complex components at different stages of the evolution of a random uniform hypergraph. Thus, let $X_{n,M}(s,k)$ denote the random variable which counts components on $r=s(d-1)-k$ vertices and s edges of $\mathbb{G}^d(n,M)$. Then, for the expectation of $X_{n,M}(r,k)$, we have

$$\mathbb{E}X_{n,M}(r,k) = \binom{n}{r} C_d(s,k) \left(\binom{\binom{n-r}{d}}{M-s}\right) \Big/ \left(\binom{\binom{n}{d}}{M}\right).$$

Now, from Stirling's formula,

$$\binom{n}{r} = \frac{1}{\sqrt{2\pi r}} \frac{n^r \mathrm{e}^r}{r^r} \exp\left(-\frac{r^2}{2n} - \frac{r^3}{6n^2} + O\left(\frac{r^4}{n^3} + \frac{1}{r}\right)\right).$$

Furthermore,

$$\begin{aligned}
&\left(\binom{\binom{n-r}{d}}{M-s}\right) \Big/ \left(\binom{\binom{n}{d}}{M}\right) \\
&\quad = \frac{(n-r)^{d(M-s)}}{n^{d(M-s)}} \frac{(M)_s (d!)^s}{n^{ds}} \exp\left(O\left(\frac{s}{n}\right)\right) \\
&\quad = \exp\left(-\left(\frac{r}{n} + \frac{r^2}{2n^2} + \frac{r^3}{3n^3}\right) d(M-s) - \frac{s^2}{2M} - \frac{s^3}{6M^2} + O\left(\frac{s}{n} + \frac{s^4}{n^3}\right)\right) \left(\frac{d!\,M}{n^d}\right)^s.
\end{aligned}$$

Let $M = n/d(d-1) - m$, where $m/n^{2/3} \to \infty$ but $m = o(n)$. Then

$$-\frac{s^2}{2M} - \frac{s^3}{6M^2} = -\frac{s^2 d(d-1)}{2n} - \frac{ms^2 d^2 (d-1)^2}{2n^2} - \frac{s^3 d^2 (d-1)^2}{6n^2} + O\left(\frac{ms^3}{n^3}\right)$$

and

$$\begin{aligned}
\left(\frac{d!\,M}{n^d}\right)^s &= \frac{[(d-2)!]^s}{n^{s(d-1)}} \left(1 + \frac{d(d-1)m}{n}\right)^s \\
&= \frac{[(d-2)!]^s}{n^{s(d-1)}} \exp\left(-\frac{msd(d-1)}{n} - \frac{m^2 s d^2 (d-1)^2}{2n^2} + O\left(\frac{m^3 s}{n^4}\right)\right).
\end{aligned}$$

Combining the above formulae and substituting $r = s(d-1) - k$ we get

$$\binom{n}{r}\binom{\binom{n-r}{d}}{M-s}\Big/\binom{\binom{n}{d}}{M}$$

$$\sim \frac{1}{\sqrt{2\pi}}\frac{[(d-2)!]^s}{(d-1)^{s(d-1)-k+1/2}}\frac{n^{-k}}{s^{s(d-1)-k+1/2}}$$

$$\times \exp\left(s(d-2) + \frac{k}{d-1} - \frac{ms^2d(d-1)^3}{2n^2} - \frac{m^2sd^2(d-1)^2}{2n^2} - \frac{s^3(d-1)^4}{6n^2}\right), \quad (1)$$

where $\sim$ means that the asymptotic equation holds up to a factor of

$$1 + O\left(\frac{1}{s} + \frac{s}{n} + \frac{s^4 + ms^3 + m^3s}{n^4} + \frac{ks}{n} + \frac{k^2}{s}\right).$$

Theorem 4 states that in the subcritical phase, when $M = n/d(d-1) - m$ and $m/n^{2/3} \to \infty$, a.a.s. $\mathbb{G}^d(n,M)$ contains no complex components. Since for $k = -1, 0$ the asymptotic value of $C_d(s,k)$ is given by Lemmas 1 and 2, from (1) we get

$$\mathbb{E}X_{n,M}(s,0) \sim \frac{1}{4s}\exp\left(-\frac{m^2sd^2(d-1)^2}{2n^2} - \frac{ms^2d(d-1)^3}{2n^2} - \frac{s^3(d-1)^4}{6n^2}\right) \quad (2)$$

and

$$\mathbb{E}X_{n,M}(s,-1) \sim \frac{1}{\sqrt{2\pi}}\frac{1}{(d-1)^2}\frac{n}{s^{5/2}}\exp\left(-\frac{m^2sd^2(d-1)^2}{2n^2} - \frac{ms^2d(d-1)^3}{2n^2} - \frac{s^3(d-1)^4}{6n^2}\right),$$

where in both of the above cases we omitted the factor

$$1 + O\left(\frac{1}{s} + \frac{s}{n} + \frac{s^4 + ms^3 + m^3s}{n^4}\right).$$

For a natural number ℓ let $U_\ell = U_\ell(n,M)$ denote the number of edges of the ℓth largest unicyclic component of $\mathbb{G}^d(n,M)$. The following theorem describes the limit distribution of U_ℓ in the subcritical phase.

Theorem 5. *Let $\ell \geqslant 1$ be a fixed natural number and let $M = n/d(d-1) - m$, where $m/n^{2/3} \to \infty$ but $m/n \to 0$ as $n \to \infty$. Then, for every function $u = u(n)$ such that $u(n) \to x > 0$ as $n \to \infty$,*

$$\lim_{n\to\infty} \mathbb{P}\left(U_\ell \geqslant \frac{2un^2}{d^2(d-1)^2m^2}\right) = \sum_{i=0}^{\ell-1}\frac{\mu^i}{i!}e^{-\mu} \quad (3)$$

and

$$\mathbb{P}\left(U_\ell = \left\lfloor\frac{2un^2}{d^2(d-1)^2m^2}\right\rfloor\right) = (1+o(1))\frac{m^2}{n^2}\frac{d^2(d-1)^2\mu^{\ell-1}}{8x(\ell-1)!}e^{-x-\mu},$$

where

$$\mu = \mu(x) = \int_x^\infty \frac{e^{-t}}{4\sqrt{t}}\, dt.$$

Proof. Let $Z(u) = Z_{n,M}(u) = \sum_{s \geqslant u} X_{n,M}(s,0)$ denote the number of unicyclic components with at least $a(u) = \lfloor 2un^2/m^2d^2(d-1)^2 \rfloor$ edges. Then, from (2), we get

$$\mathbb{E}Z(u) = \sum_{s \geqslant a(u)} \mathbb{E}X_{n,M}(s,0) = (1+o(1)) \sum_{s \geqslant a(u)} \frac{1}{4s} \exp\left(-\frac{sm^2d^2(d-1)^2}{2n^2}\right),$$

where the quantity $o(1)$ tends to 0 uniformly for all u such that, say, $1/\log(m^3/n^2) \leqslant u \leqslant \log(m^3/n^2)$. Hence,

$$\mathbb{E}Z(u) = (1+o(1)) \int_x^\infty \frac{e^{-t}}{4\sqrt{t}}\, dt = (1+o(1))\mu.$$

Furthermore, it is easy to check that, for every $j \geqslant 1$, the jth factorial moment $\mathbb{E}_j Z(u)$ of $Z(u)$ converges to μ^j. Thus, $Z(u)$ converges in distribution to a random variable with Poisson distribution with the expectation μ and (3) follows.

Finally, note that

$$\begin{aligned}\mathbb{P}(U_\ell = u) &= \binom{n}{u(d-1)} C_d(u,0) \binom{\binom{n}{d}}{M-u} \Big/ \binom{\binom{n}{d}}{M} \mathbb{P}(Z(u) = \ell - 1) + o(1)\\ &= (1+o(1))\mathbb{E}X_{n,M}(u,0) \frac{\mu^{\ell-1}}{(\ell-1)!} e^{-\mu}\\ &= (1+o(1)) \frac{m^2}{n^2} \frac{d^2(d-1)^2 \mu^{\ell-1}}{8x(\ell-1)!} e^{-x-\mu},\end{aligned}$$

where in the first line of the above equation the quantity $o(1)$ stands for the probability that $\mathbb{G}^d(n,M)$ contains two unicyclic components of size s. □

Arguing in a similar way one can prove an analogous result for the number of edges $L_\ell = L_\ell(n,M)$ contained in the ℓth largest component of $\mathbb{G}^d(n,M)$.

Theorem 6. *Let $\ell \geqslant 1$ be a fixed natural number and $M = n/d(d-1) - m$, where $m/n^{2/3} \to \infty$ but $m/n \to 0$ as $n \to \infty$. Then, a.a.s. the ℓth largest component of $\mathbb{G}^d(n,M)$ is a hypertree.*

Furthermore, let $t = t(n)$ be a function which tends to y, $-\infty < y < \infty$ as $n \to \infty$. Then

$$\lim_{n\to\infty} \mathbb{P}\left(L_\ell \leqslant \frac{2n^2}{d^2(d-1)^2m^2}\left(\log\frac{m^3}{n^2} - \frac{5}{2}\log\log\frac{m^3}{n^2} + t\right)\right) = \sum_{i=0}^{\ell-1} \frac{\lambda^\ell}{\ell!} e^{-\lambda}$$

and

$$\mathbb{P}\left(L_\ell = \left\lfloor \frac{2n^2}{d^2(d-1)^2m^2}\left(\log\frac{m^3}{n^2} - \frac{5}{2}\log\log\frac{m^3}{n^2} + t\right)\right\rfloor\right)$$

$$= (1+o(1))\frac{m^2}{n^2}\frac{d^5(d-1)^3}{8\sqrt{\pi}}\frac{\lambda^{\ell-1}}{(\ell-1)!}\mathrm{e}^{-y-\lambda},$$

where

$$\lambda = \lambda(y) = \frac{d^3(d-1)\mathrm{e}^{-y}}{4\sqrt{\pi}}. \quad \square$$

As an immediate consequence of Theorems 4–6 we get the following fact.

Corollary 7. *Let $M = n/d(d-1) - m$, where $m/n^{2/3} \to \infty$ as $n \to \infty$. Then a.a.s. $G(n,M)$ contains no components with more than $n^{2/3}$ edges.*

4. Supercritical phase

In this section, we prove the main result concerned with the number of vertices and the number of edges in the largest component of $\mathbb{G}^d(n,M)$ in the supercritical case, i.e., when $M = n/d(d-1)+m$ and $m/n^{2/3} \to \infty$ as $n \to \infty$. Unfortunately, we are able to do it only under the additional assumption that $m/n^{2/3}$ tends to infinity slowly enough, more precisely that $m = o(n^{2/3}\log n/\log\log n)$.

Let $p_{s,k} = p_{s,k}(n,M)$ denote the probability that the lexicographically first largest component of $\mathbb{G}^d(n,M)$ contains $r = s(d-1)-k$ vertices and s edges. (We remark that for this range of M the largest component of $\mathbb{G}^d(n,M)$ is a.a.s. always unique; thus the words "lexicographically first" we are using to make $p_{s,k}$ well defined are not very relevant.) The main result of this section gives us the precise joint distribution of s and k in the early supercritical phase.

Theorem 8. *Let $M = n/d(d-1)+m$, where $m^3/n^2 \to \infty$ but $m^3\log\log n/n^2\log n \to 0$ as $n \to \infty$. Then the largest component of $\mathbb{G}^d(n,M)$ a.a.s. contains $(1+o(1))2dm/(d-1)$ edges and has excess $(1+o(1))2(d-1)^3m^3/3n^2$. Furthermore, let $x = x(n)$, $y = y(n)$ be functions such that $x(n) \to a$, $y(n) \to b$ as $n \to \infty$. Set*

$$s = \left\lfloor \frac{2dm}{d-1} + \frac{x}{d-1}\sqrt{\frac{2n^2}{d(d-1)m}}\right\rfloor \tag{4}$$

and

$$k = \left\lfloor \frac{2(d-1)d^3m^3}{3n^2} + y\sqrt{\frac{10(d-1)d^3m^3}{3n^2}}\right\rfloor. \tag{5}$$

Then

$$p_{s,k} = (1+o(1))\frac{\sqrt{6}}{8\pi}\frac{d-1}{dm}\exp\left(-\frac{5}{4}a^2 + \frac{\sqrt{15}}{2}ab - \frac{5}{4}b^2\right). \tag{6}$$

Proof. Let M, s, and k be defined as above and let $r=s(d-1)-k$. In order to construct a d-uniform hypergraph on n vertices in which the largest component has $r=s(d-1)-k$ vertices and s edges, first choose the vertices and the edges of the largest component in one of $\binom{n}{r}C_d(s,k)$ possible ways and then pick the remaining $M-s$ edges such that no components of more than r vertices emerges. One can easily check that

$$\left(M-s-\frac{n-r}{d(d-1)}\right)\Big/ n^{2/3}\to-\infty$$

as $n\to\infty$. Thus, Corollary 7 implies that the probability that the largest component of $G(n-r,M-s)$ is larger than $n^{2/3}=o(s)$ tends to 0 as $n\to\infty$ uniformly for the range of s and k we consider. Consequently,

$$p_{s,k}=(1+o_{s,k}(1))\binom{n}{r}C_d(s,k)\left(\begin{pmatrix}n-r\\d\end{pmatrix}\atop M-s\right)\Big/\left(\begin{pmatrix}n\\d\end{pmatrix}\atop M\right),$$

where here and below $o_{s,k}(1)$ denotes the value which tends to 0 as $n\to\infty$ uniformly for every $s=s(n,M)$ such that $dm\leqslant s(d-1)\leqslant 3dm$ and $k=k(n,M)$ for which $(d-1)d^3m^3/2n^2\leqslant k\leqslant (d-1)d^3m^3/n^2$.

The asymptotic value of

$$\binom{n}{r}\left(\begin{pmatrix}n-r\\d\end{pmatrix}\atop M-s\right)\left(\begin{pmatrix}n\\d\end{pmatrix}\atop M\right)$$

is given by formula (1) (note that since $M=n/d(d-1)+m$, the sign of m in (1) must be changed). Furthermore, Lemma 3 provides the value for $C_d(s,k)$. Thus, we arrive at

$$p_{s,k}=(1+o_{s,k})\frac{\sqrt{6}}{4\pi s}\left(\frac{e(d-1)^4s^3}{12kn^2}\right)^{k/2}\exp\left(-\frac{m^2sd^2(d-1)^2}{2n^2}+\frac{ms^2d(d-1)^3}{2n^2}-\frac{s^3(d-1)^4}{6n^2}\right). \tag{7}$$

Routine but not very exciting calculations show that for every function $\omega=\omega(n)$ which tends to infinity as $n\to\infty$,

$$s_\pm=\lfloor 2dm/(d-1)\pm\omega n/\sqrt{m}\rfloor$$

and

$$k_\pm=\lfloor 2(d-1)d^3m^3/3n^2\pm\omega m^{3/2}/n\rfloor,$$

we have

$$\sum_{s=s_-}^{s_+}\sum_{k=k_-}^{k_+}p_{s,k}=1+o_{s,k}(1).$$

Finally, if we put into (7) the value of s and k given by (4) and (5) then, after tedious computations, it reduces to (6). □

As an immediate corollary of the above result we get the following.

Theorem 9. *Let $M = n/d(d-1) + m$, where $m^3/n^2 \to \infty$ but $m^3 \log\log n/n^2 \log n \to 0$ as $n \to \infty$. Furthermore, let X_n and Y_n denote the number of edges and the excess in the lexicographically first largest component of $\mathbb{G}^d(n,M)$, and*

$$\tilde{X}_n = \left(X_n - \frac{2dm}{d-1}\right) \Big/ \sqrt{\frac{2n^2}{d(d-1)m}}$$

and

$$\tilde{Y}_n = \left(Y_n - \frac{2(d-1)d^3m^3}{3n^2}\right) \Big/ \sqrt{\frac{10(d-1)d^3m^3}{3n^2}}.$$

Then the random variable $(\tilde{X}, \tilde{Y})$ converges in distribution to (X, Y), where (X, Y) has the standarized normal distribution with correlation $\sqrt{15}/5$. □

The structure of $\mathbb{G}^d(n,M)$ can be easily deduced from Theorems 5 and 6, Corollary 7 and Theorem 8. Let us call a component of $\mathbb{G}^d(n,M)$ *large* if it contains more than $n^{2/3}$ edges and *small* otherwise. Then, in the supercritical phase, a.a.s. $\mathbb{G}^d(n,M)$ contains precisely one large component, whose size and excess are characterized by Theorem 8. Furthermore, the distribution of the sizes of the small components can be characterized in a similar way as in Theorems 5 and 6; since we would not like to repeat lengthy and complicated formulae we give the local limit theorem only for the size of the ℓth largest component.

Theorem 10. *Let $M = n/d(d-1) + m$, where $m^3/n^2 \to \infty$ but $m^3 \log\log n/n^2 \log n \to 0$ as $n \to \infty$. Then a.a.s. $\mathbb{G}^d(n,M)$ consists of one large complex component and some number of small components which are either hypertrees or unicyclic.*

Furthermore, let $\ell \geqslant 2$ be a fixed number and let $t = t(n)$ be a function which tends to y, $-\infty < y < \infty$ as $n \to \infty$. Then the ℓth largest component of $\mathbb{G}^d(n,M)$ is a hypertree with L_ℓ edges, where

$$\mathbb{P}\left(L_\ell = \left\lfloor \frac{2n^2}{d^2(d-1)^2m^2}\left(\log\frac{m^3}{n^2} - \frac{5}{2}\log\log\frac{m^3}{n^2} + t\right)\right\rfloor\right)$$
$$= (1+o(1))\frac{m^2}{n^2}\frac{d^5(d-1)^3}{8\sqrt{\pi}}\frac{\lambda^{\ell-2}}{(\ell-2)!}\mathrm{e}^{-y-\lambda}$$

and

$$\lambda = \lambda(y) = \frac{d^3(d-1)\mathrm{e}^{-y}}{4\sqrt{\pi}}.$$

Proof. Let $M = n/d(d-1) + m$, where $m^3/n^2 \to \infty$ but $m^3 \log\log n/n^2 \log n \to 0$ as $n \to \infty$. Let us remove from $\mathbb{G}^d(n,M)$ the vertices of the largest component. Then, from Theorem 8 we infer that a.a.s. the random graph $\hat{\mathbb{G}}^d(n,M)$ obtained in this way has

$$n' = n - 2dm + O(n/\sqrt{m})$$

vertices, and

$$M' = \frac{n}{d(d-1)} + m - \frac{2dm}{d-1} + O\left(\frac{n}{\sqrt{m}}\right) = \frac{n'}{d(d-1)} - m + O\left(\frac{n}{\sqrt{m}}\right)$$

edges. Note that, if we fix n' and M', then each such hypergraph with largest component smaller than, say, $n^{2/3}$, is equally likely to appear as $\hat{\mathbb{G}}^d$. Furthermore, from Corollary 7 it follows that a.a.s. the largest component of $\hat{\mathbb{G}}^d$ has at most $n^{2/3}$ edges. Thus, to complete the proof it is enough to observe that the limit distributions given in Theorem 6 remain unchanged if we replace n by $n' = n - 2dm - O(n/\sqrt{m})$ and m by $m' = m + O(n/\sqrt{m})$. $\square$

Theorems 8 and 9 describe the structure of the largest component of $\mathbb{G}^d(n,M)$ only for the early supercritcal phase, when $M = n/d(d-1)+m$, and $m/n^{2/3} = o(\log n/\log\log n)$. We conjecture however that a similar result holds for every m such that $m/n^{2/3} \to \infty$ but $m = o(n)$; i.e., then the appropriately standarized random variables X_n and Y_n in Theorem 9 converge in distribution to the standarized bivariate normal distribution with correlation coefficient $\sqrt{15}/5$.

References

[1] E.A. Bender, E.A. Canfield, 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.
[2] B. Bollobás, The evolution of random graphs, Trans. Amer. Math. Soc. 286 (1984) 257–274.
[3] B. Bollobás, Random graphs, Academic Press, London, 1985.
[4] P. Erdős, A. Rényi, On the evolution of random graphs, Publ. Math. Sci. Hung. Acad. Sci. 5 (1960) 17–61.
[5] S. Janson, D.E. Knuth, T. Łuczak, B. Pittel, The birth of the giant component, Random Struct. Algorithms 4 (1993) 233–358.
[6] S. Janson, T. Łuczak, A. Ruciński, Random Graphs, Wiley, New York, 2000.
[7] M. Karoński, T. Łuczak, The number of connected sparsely edged uniform hypergraphs, Discrete Math. 171 (1997) 153–167.
[8] T. Łuczak, Component behavior near the critical point of the random graph process, Random Struct. Algorithms 1 (1990) 287–310.
[9] J. Schmidt-Pruzan, E. Shamir, Component structure in the evolution of random hypergraphs, Combinatorica 5 (1985) 81–94.
[10] B.I. Selivanov, Perechislenie odnorodnykh hipergrafov c prostoĭ ciklovoĭ strukturoĭ , Kombinatornyĭ Analiz 2 (1972) 60–67.