Cores of Random r-Partite Hypergraphs

Fabiano C. Botelho[a,c], Nicholas Wormald[b], Nivio Ziviani[c]

[a] *Data Domain an EMC company, Santa Clara, CA, USA*
[b] *Department of Combinatorics and Optimization, University of Waterloo, Waterloo ON, Canada*
[c] *Department of Computer Science, Federal University of Minas Gerais, Belo Horizonte, Brazil*

Abstract

We show that the threshold $c_{r,k}$ (in terms of the average degree of the graph) for appearance of a k-core in a random r-partite r-uniform hypergraph $G_{r,n,m}$ is the same as for a random r-uniform hypergraph with cn/r edges without the r-partite restriction, where $r, k \geq 2$. In both cases, the average degree is c. The case $r, k = 2$ was analyzed in [4] but the general case $r \geq 3, k \geq 2$ is still open. Besides the proof for the general case, we have also provided a simpler proof for the case $r, k = 2$. This problem was provided without a proof (but with strong experimental evidence) in the analysis of the algorithm presented in [2]. This algorithm constructs a family of minimal perfect hash functions based on random r-partite r-uniform hypergraphs with an empty k-core subgraph, for $k \geq 2$. For an input key set S with m keys, the family of minimal perfect hash functions generated by the algorithm can be stored in $O(m)$ bits, where the hidden constant is within a factor of two from the information theoretical lower bound.

Key words: Minimal Perfect Hashing, cores of hypergraphs

1. Introduction

The study of random graphs started with Erdős and Rényi [7, 8, 10, 9]. A modern treatment is given in [1, 15]. Many results describing statistical properties of random graphs were obtained [11, 13, 14, 18, 19, 22, 23, 24]. For instance, distribution of component sizes, existence and size of a giant component, vertex degree distributions, arising of cycles, existence and size of specific subgraphs, among others.

Random graphs have been extensively used by perfect hashing methods as a model to construct a data structure that compactly stores a static key set and supports queries to locate keys in one probe, see, e.g., [12, 16]. The algorithm presented in [2] constructs a family of perfect hash functions or collision free hash functions based on random r-partite r-uniform hypergraphs with an empty k-core subgraph, for $k \geq 2$ (see the definitions in Section 1.1).

The objective of this paper is to prove that the threshold $c_{r,k}$ (in terms of the average degree of the graph) for appearance of a k-core in a random r-partite r-uniform hypergraph is the same as for a random hypergraph without the r-partite restriction, where $r, k \geq 2$. This problem came up in the analysis of the algorithm presented in [2], where this claim was not proved but was provided with strong experimental evidence. The case $r, k = 2$ was analyzed in [4] but the general case $r \geq 3, k \geq 2$ is still open. Besides the proof for the general case, we have also provided a simpler proof for the case $r, k = 2$. The first determination of the threshold of existence of a k-core in a random graph without the r-partite restriction was given by Pittel, Spencer and Wormald [20].

In Section 1.1, we present basic concepts and definitions that are used in the other parts of the text. In Section 1.2, we formalize the objective of this paper and outline our results and contributions. In Section 2, we present the proof that the threshold $c_{r,k}$ for appearance of a k-core in a random r-partite r-uniform hypergraph is the same as for a random hypergraph without the r-partite restriction, where $r, k = 2$. In Section 3, we present the proof of the same case of Section 2, but allowing a single cycle with length multiple of four per connected component. In Section 4, we present the proof of the same case of Section 2, but now for the general case where $r, k \geq 2$.

1.1. Preliminaries

Our random hypergraph models are based on the set $\Omega_{r,n}$ of r-partite r-uniform hypergraphs whose vertex set V is a disjoint union of r parts $U_1, \ldots, U_r$, $|U_i| = n$, for $i = 1, \ldots, r$. Each edge contains r vertices, one from each part of the partition. Multiple edges are permitted.

Definition 1 *Let* $\mathcal{G}_{r,n,p}$, $0 \leq p \leq 1$, *denote the model of random hypergraphs from* $\Omega_{r,n}$ *whose* n^r *possible edges occur independently of each other, each with probability* p.

Definition 2 *Let* $\mathcal{G}_{r,n,m}$ *denote the probability space of hypergraphs from* $\Omega_{r,n}$ *generated via a stochastic process which starts with a set of* rn *vertices, and* m *times repeats the step of adding uniformly at random any edge, permitting repetitions. Let* $\mathcal{G}'_{r,n,m}$ *denote the uniform probability space containing those hypergraphs from* $\Omega_{r,n}$ *that have* m *distinct edges.*

Email addresses: fabiano@datadomain.com (Fabiano C. Botelho), nwormald@uwaterloo.ca (Nicholas Wormald), nivio@dcc.ufmg.br (Nivio Ziviani)

Definition 3 *We use $G_{r,n,m}$ to denote a random r-partite r-uniform hypergraph in $\mathcal{G}_{r,n,m}$, and c to denote m/n.*

In this paper we will concentrate on proving results about $\mathcal{G}_{r,n,m}$. It is easy to see by simple counting that if we condition on the edges being distinct, each hypergraph occurs with the same probability, and hence we obtain precisely the model $\mathcal{G}'_{r,n,m}$. Note that if $m = cn$ where c is constant, there will be no multiple edges with probability $(n^r)_m/n^{rm}$, where $(n^r)_m = n^r(n^r-1)\ldots(n^r-m+1)$. Standard analysis (using for example the method in Section 2) shows this is $e^{-O(1)}$ for $r \geq 2$. Hence the m edges are distinct with probability bounded away from 0. So if we show something is a.a.s. (asymptotically almost surely) true in $\mathcal{G}_{r,n,m}$, it is also true for the model $\mathcal{G}'_{r,n,m}$ in which we condition on the edges being distinct.

It is to be expected that many results for $\mathcal{G}_{r,n,p}$ are equivalent to results for $\mathcal{G}'_{r,n,m}$ whenever $p = c/n^{r-1}$ and $n \to \infty$. In particular, by Proposition 1.12 of [14], for a monotone increasing property P (which means one preserved by taking supergraphs), if we show that $\mathcal{G}'_{r,n,m}$ a.a.s. has P, then so does $\mathcal{G}_{r,n,p}$ for p just slightly above c/n^{r-1}.

Definition 4 *The k-core of a hypergraph is the largest subgraph of minimum degree at least k.*

Definition 5 *A minimal perfect hash function is a bijection from a static key set S of size m to $\{0, 1, \ldots, m-1\} = [m]$.*

In the following we introduce some definitions found in [5] in order to use the same approach to state and prove the results summarized in Section 1.2. For $k \geq 0$, denote by $Z(k, \lambda)$ a random variable which has a k-truncated Poisson distribution with parameter λ, that is:

$$\mathbf{P}(Z(k,\lambda) = j) = \begin{cases} \dfrac{\lambda^j}{j!\,f_k(\lambda)} & j \geq k \\ 0 & j < k \end{cases} \tag{1}$$

where f_k is defined as:

$$f_k(\lambda) = e^{\lambda} - \sum_{i=0}^{k-1} \frac{\lambda^i}{i!} = \sum_{i \geq k} \frac{\lambda^i}{i!}.$$

Let $h_k(\mu) = \mu/(e^{-\mu} f_{k-1}(\mu))$ and define

$$c_k = \inf\{h_k(\mu) : \mu > 0\}. \tag{2}$$

Take $k \geq 3$. Then c_k is a positive real because $h_k(\mu)$ tends to ∞ if μ tends to 0 or ∞. Note that $\frac{d}{d\mu} f_{k-1} = f_{k-2}$ and hence the logarithmic derivative of $h_k(\mu)$ with respect to μ is:

$$\mu^{-1} + 1 - f_{k-2}/f_{k-1} = \mu^{-1} - \frac{\mu^{k-2}}{(k-2)!\,f_{k-1}}.$$

For this to be 0 we need $1/(k-2)! = f_{k-1}/\mu^{k-1} = \sum_{i\geq 0} \mu^i/(i+k-1)!$). Since the latter function is monotonically increasing, $h_k(\mu)$ has a unique stationary point for positive μ (at which point it has value c_k), and for $c > c_k$ the equation $h_k(\mu) = c$ has two positive roots. Define $\mu_{k,c}$ to be the larger one. Also define $c_2 = 1$.

For $k \geq 2$ and $r \geq 3$, we are interested in k-cores of the random r-partite r-uniform hypergraph as presented in Definition 3. Note that the k-cores of $\mathcal{G}_{r,n,m}$ form a probability space, which we denote by $\mathcal{K}_{r,n,m,k}$. Then, we can generalize the definition of c_k to:

$$c_{r,k} = \inf\{h_{r,k}(\mu) : \mu > 0\},$$

where $h_{r,k}(\mu) = \mu/(e^{-\mu} f_{k-1}(\mu))^{r-1}$. As with $h_k(\mu)$, $h_{r,k}(\mu)$ tends to ∞ if μ tends to 0 or ∞, so $c_{r,k}$ is a positive real (and this applies even when $k = 2$). Define $\mu_{r,k,c}$ to be the larger solution of $h_{r,k}(\mu) = c$ for $c > c_{r,k}$. We note that $c_{r,k}$ is the threshold of appearance of a k-core in a random r-uniform hypergraph without the r-partite restriction.

1.2. Results

More formally, the objective of this paper is to prove that the threshold $c_{r,k}$ for appearance of a k-core in $G_{r,n,m}$ is the same as for a random r-uniform hypergraph with cn/r edges without the r-partite restriction, where $r, k \geq 2$. In both cases, the average degree of the hypergraph is c.

The threshold $c_{r,k}$ for appearance of a k-core for a random r-uniform hypergraph with cn/r edges was first analyzed in [6]. In this paper we instead analyze it for r-partite r-uniform hypergraphs. So these results will finish the analysis of the algorithm presented in [2].

We first consider the case when $r = 2$ and $k = 2$, which has been analyzed in [4]. We have used a simpler method here to get to the same result and also to define what we need to prove Theorem 2. The result is summarized in Theorem 1.

Theorem 1 ([4], Theorem 3.5) *Let $G_{2,n,m} = (V, E)$ be a random bipartite 2-uniform hypergraph with $2n$ vertices and $m = cn$ edges. Then, if $c = m/n$ holds for $c \in (0, 1)$ and $n \to \infty$, the probability that $G_{2,n,m}$ is simple and has an empty 2-core component is:*

$$\mathbf{P}_a = \sqrt{1 - c^2}. \tag{3}$$

Although it was not mentioned in [2], the version of the algorithm presented there for $G_{2,n,m}$ can be sped up by allowing a single cycle with length multiple of four per connected component. For details on the optimization the reader can refer to [3, Chapter 6]. This happens because the probability of generating $G_{2,n,m}$ over the condition that it does not have any cycle with a length that is not a multiple of four is 58% higher than the probability of generating it when cycles are not permitted (see Sections 2 and 3) and the expected runtime of the algorithm is inversely proportional to this probability. To analyze this version of the algorithm it is required the following theorem.

Theorem 2 *Let $G_{2,n,m}$ and c be as in Theorem 1. Then, the probability that $G_{2,n,m}$ is simple and has no cycle of length congruent to 2 (mod 4), for $n \to \infty$, is:*

$$\mathbf{P}_b = \frac{\sqrt{1 - c^2}}{(1 - c^4)^{\frac{1}{4}}}. \tag{4}$$

We remark that for $c \geq 1$ there are a.a.s. many cycles of all short even lengths.

We now consider the case when r and k are not both 2. Note that in $\mathcal{G}_{r,n,m}$ multiple edges are permitted. However, as we discussed in Section 1.1, even if they were forbidden, the same results would hold because the probability that there are no multiple edges is bounded away from 0. Hence, once we prove the theorem for the hypergraphs that lie in $\mathcal{G}_{r,n,m}$, it also follows for the other variations where multiple edges are forbidden.

Theorem 3 *Let $c > 0$ and integers $r \geq 2$, $k \geq 2$ be fixed, where r and k are not both 2. Suppose that $m \sim cn$, and $G \in \mathcal{G}_{r,n,m}$. For $c < c_{r,k}$, G has empty k-core a.a.s. For $c > c_{r,k}$, the k-core of G a.a.s. has $e^{-\mu_{r,k,c}} f_k(\mu_{r,k,c}) rn(1 + o(1))$ vertices and $\mu_{r,k,c} e^{-\mu_{r,k,c}} f_{k-1}(\mu_{r,k,c}) n(1+o(1))$ hyperedges. Moreover, let $j \geq k$ be fixed, and assume $c > c_{r,k}$. Then the number of vertices of degree j in $\mathcal{K}(r,n,m,k)$ is a.a.s. $rne^{-\mu}\mu^j/j! + o(n)$, where $\mu = \mu_{r,k,c}$.*

2. Proof of Theorem 1

We start the proof by showing that the probability distribution of cycles in bipartite random graphs can be approximated by a Poisson distribution. For this we use a consequence of Brun's sieve or the Bonferroni inequalities; see [1] or [25, Lemma 2.8]. The same method was used for cycle distributions in the first papers on random graphs by Erdős and Rényi.

Consider $G_{r,n,p}$, where $r = 2$ and $p = c/n$, a bipartite random graph as in Definition 1. Note that by using the $\mathcal{G}_{r,n,p}$ model multiple edges are forbidden. For a non-negative integer i and real x, $(x)_i$ denotes $x(x-1)\dots(x-i+1)$. Let X_{2l} be the (random variable) number of cycles of length $2l$, $l \geq 2$. We only need to show that:

$$\lim_{n\to\infty} \mathbb{E}\left(\prod_{l=2}^{k} (X_{2l})_{j_{2l}}\right) = \prod_{l=2}^{k} \lambda_{2l}^{j_{2l}} \tag{5}$$

for each fixed sequence $j_4, j_6, \dots, j_{2k}$ and fixed k where $\lambda_{2l} = \frac{c^{2l}}{2l}$.

The left hand side of Eq. (5) is exactly the number of ordered sets of j_4 distinct 4-cycles, j_6 distinct 6-cycles, up to j_{2k} $2k$-cycles in the random graph. Let's first consider the case the cycles are disjoint. There are asymptotically $\prod_{l=2}^{k}(n^{2l}/(4l))^{j_{2l}}$ ways to place the vertices of such cycles, since specifying the vertices of some of them leaves $n - o(n)$ vertices for the next choice (and a cycle of length $2l$ has $4l$ automorphisms). Then, given the placement of the cycles, the probability that all their edges occur in the random graph is $\prod_{l=2}^{k}((c/n)^{2l})^{j_{2l}}$. The product of these factors is asymptotic to the right side of Eq. (5).

For the case that the ordered set of cycles share any vertices, notice first that from elementary graph theoretic considerations, the set of cycles induces a subgraph with more vertices than edges. For each such subgraph, there is a bounded number of ways of specifying the ordered set of cycles. Moreover, the expected number of such subgraphs with a given number k of vertices, and at least $k+1$ edges, is $O(n^k p^{k+1}) = o(1)$. Hence the contribution from this case is $o(1)$.

We deduce that Eq. (5) holds and X_{2l}, $2 \leq l \leq k$, are asymptotically independent Poisson, each one with parameter $\lambda_{2l} = \frac{c^{2l}}{2l}$. Then to deduce the probability of no even cycles given that there is no multiple edges as $n \to \infty$, just use:

$$\begin{aligned}\mathbf{P}(X_e = 0) &= \mathbf{P}(X_4 = \cdots = X_{2k} = 0) + O\left(\sum_{l>k} \mathbf{E}(X_{2l})\right) \\ &= e^{-\sum_{l=2}^{\infty} \frac{1}{2l} c^{2l}} + O(\epsilon_k) = e^{\frac{1}{2}\ln(1-c^2) + \frac{1}{2}c^2} + O(\epsilon_k)\end{aligned} \tag{6}$$

where ϵ_k is the upper tail of the summation and goes to 0 as $k \to \infty$. We have used Maclaurin's expansion $\sum_{l=1}^{\infty} \frac{1}{2l} x^l = -\frac{1}{2}\ln(1-x)$ above, where $x = c^2$.

The algorithm in [2] iteratively generates a bipartite random graph until one with no cycles and no multiple edges is obtained. Let C denote the event that the bipartite random graph has no cycles and $\mathcal{M}$ denote the event that it has no multiple edges. Thus, to finalize the proof we need to show that:

$$\mathbf{P}_a = \mathbf{P}(C \cap \mathcal{M}) = \mathbf{P}(C|\mathcal{M})\mathbf{P}(\mathcal{M}). \tag{7}$$

Eq. (6) gives $\mathbf{P}(C|\mathcal{M})$ and $\mathbf{P}(\mathcal{M}) = (n^2)_m / n^{2m}$. Thus:

$$\lim_{n\to\infty} \mathbf{P}(\mathcal{M}) = e^{-\frac{1}{2}c^2}. \tag{8}$$

To get this we used standard calculus to approximate $f(x) = 1 - x$ by $g(x) = e^{-x}$ for a small real $x \in (0,1)$. By taking k large enough and multiplying equations (6) and (8) we obtain:

$$\mathbf{P}_a = \sqrt{1 - c^2}. \tag{9}$$

Note that c is restricted to be in the range $(0,1)$ and, therefore, $c_{2,2} = 1$. ∎

This matches the experimental results presented in [2]. For instance, when $c = 0.957$ we have $\mathbf{P}_a = 0.29$. This is very close to 0.294 that is the value obtained in [2] by generating 1,000 random bipartite 2-graphs with $n = 10^7$ edges.

3. Proof of Theorem 2

Let X_{2l}, $2 \leq l \leq k$, and $G_{r,n,p}$, where $r = 2$ and $p = c/n$ be defined as in the proof of Theorem 1 presented in Section 2. Those cycles with a length that is not a multiple of four (i.e, $l = 3, 5, 7 \dots$) cannot be used to build MPHFs in the algorithm presented in [3, Chapter 6]. Accordingly, we define such cycles to be *bad*, and let X_b be the random variable that denotes the number of bad cycles in $G_{r,n,p}$.

Then, to deduce the probability of no bad cycles given that there is no multiple edges as $n \to \infty$, just use:

$$\begin{aligned}\mathbf{P}(X_b = 0) &= \mathbf{P}(X_6 = X_{10} \cdots = X_{2(2k+1)} = 0) + O\left(\sum_{l>k} \mathbf{E}(X_{2(2l+1)})\right) \\ &= e^{-\sum_{l=3,5,7,\dots}^{\infty} \frac{1}{2l} c^{2l}} + O(\epsilon_k) \\ &= e^{-\left(\sum_{l\geq 2} \frac{1}{2l} c^{2l} - \sum_{l\geq 1} \frac{1}{4l} c^{4l}\right)} + O(\epsilon_k) \\ &= e^{\frac{1}{2}\ln(1-c^2) + \frac{1}{2}c^2 - \frac{1}{4}\ln(1-c^4)} + O(\epsilon_k)\end{aligned} \tag{10}$$

for a fixed $k \geq 1$, where ϵ_k is the upper tail of the summation and goes to 0 as $k \to \infty$.

Therefore, the probability that the algorithm in [3, Chapter 6] generates a random graph with no bad cycles and no multiple edges is obtained by multiplying Equations (10) and (8) and fixing k in a large enough value. Thus:

$$\mathbf{P}_b = \frac{\sqrt{1-c^2}}{(1-c^4)^{\frac{1}{4}}}.$$

Note that c is restricted to be in the range $(0, 1)$. ∎

For $c = 0.957$ we have $\mathbf{P}_b = 0.458$. Experimentally, we obtained $\mathbf{P}_b = 0.463$ by generating 1,000 random bipartite 2-graphs with $n = 10^7$ edges, which is very close to the theoretical value.

4. Proof of Theorem 3

Analogous results were proved for ordinary (not multipartite) hypergraphs in [5, 6, 17]; for our present goal we will follow the proof in [5]. The argument given there defines a related model of hypergraphs, analyses a greedy algorithm that finds the core using a differential equation argument, and deduces the size of the core a.a.s. This argument requires very few modifications for the present setting, and in particular, exactly the same system of differential equations arises. So we will refer the reader to [5] for the details of some parts of the argument that apply with little modification to the present setting.

The set-up for the argument in [5] is as follows. A *heavy* vertex is one of degree at least k, and a vertex is *light* otherwise. The greedy algorithm starts with a hypergraph and repeatedly deletes light vertices (in any order); when all remaining vertices are heavy, the k-core of the hypergraph is what remains. First, the *allocation model* for graphs was defined. This consists of a mapping of $2m$ 'points' $1, \ldots, 2m$ to n vertices, chosen uniformly at random. Then a pseudograph (we use 'pseudo-' to denote that the graph possibly has loops and multiple edges) results when one regards $\{2i-1, 2i\}$ as an edge ($1 \leq i \leq m$). A result of Chvátal (see [5, Lemma 2]) shows that the probability the pseudograph is simple is $\Omega(1)$. It is also clear that restricting the pseudograph to being simple gives the standard random graph model $\mathcal{G}_{n,m}$.

It is routine to check that the same statements carry over to $\mathcal{G}_{r,n,m}$ if we now define a random r-partite pseudohypergraph model in the obvious way from a random allocation of rm points to rn vertices in sets $U_1, \ldots, U_r$ with $|U_j| = n$ such that $ri - r + j$ is mapped to a vertex in U_j ($1 \leq i \leq n$, $1 \leq j \leq r$). This pseudohypergraph, which we call the *allocation model for a random pseudohypergraph*, by design has no repeated vertices in any edge. Also, the fact that it has no multiple edges with probability $\Omega(1)$ can be shown by a straightforward application of the method in Section 2. So, by proving that a property is a.a.s. true of the k-core of the random pseudohypergraph arising from the allocation model, we may deduce that it is a.a.s. true of the k-core of $\mathcal{G}_{r,n,m}$. Henceforth we study this allocation model.

We have to pay attention to the number and total degree of the heavy vertices in each of the parts. So, for $j = 1, \ldots, r$, let t_j denote the number of heavy vertices in U_j, let s_j denote the sum of their degrees, and let ℓ_j denote the sum of the degrees of the light vertices in U_j.

Given the vectors $\mathbf{t} = (t_1, \ldots, t_r)$, $\mathbf{s} = (s_1, \ldots, s_r)$ and $\mathbf{e} = (\ell_1, \ldots, \ell_r)$, we consider a slightly different model of random hypergraphs, called (in this case) the multipartite pairing-allocation model $\mathcal{P}$. The motivation behind this definition is that splitting each light vertex of $G_{r,n,m} \in \mathcal{G}_{r,n,m}$ of degree i into i vertices of degree 1 does not change the k-core, and it enables us to 'forget' the degrees of the light vertices. $\mathcal{P}$ has disjoint vertex sets $V_1, \ldots, V_r$ with $|V_j| = t_j$, and also separate vertex sets $L_1, \ldots, L_r$ with $|L_j| = \ell_j$. Now select the m edges randomly by choosing for each edge one vertex in each of $L_j \cup V_j$, $j = 1, \ldots, r$, conditioning on the total degree of vertices in V_j being s_j, for each j, and each vertex in L_j having degree exactly 1, and also conditioning on all vertices in V_j receiving degree at least k (for each j). This produces a random pseudohypergraph $G_{\mathcal{P}}$ in which each vertex in L_j represents the end of an edge at a light vertex of $G_{r,n,m}$. The model does not record which of the light vertices in $G_{\mathcal{P}}$ the edges are actually incident with.

For use in analysing a process which determines the k-core, we need to note the following. Let $\tilde{G}_{\mathcal{P}}$ denote $G_{\mathcal{P}}$ conditioned on the hypergraph being simple (i.e. having no multiple edges). Conditioned on $\mathbf{t}$, $\mathbf{s}$ and $\mathbf{e}$, the k-core of $\tilde{G}_{\mathcal{P}}$ has the same distribution as the k-core of $G_{r,n,m}$. The proof of this is an easy exercise by counting; c.f. [5, Lemma 3]. Moreover, since the model for $G_{\mathcal{P}}$ involves a random allocation of the ends of edges to the vertices, subject to fixed sum of the heavy vertex degrees, we see that its distribution of vertex degrees in the vertices in V_j is precisely multinomial (with parameters being $|V_j| = t_j$ and the total degree s_j of its vertices) conditioned on all degrees being at least k. We call this distribution truncated multinomial, denoted by $\mathbf{Multi}(t, s)|_{\geq k}$ where $t = t_j$ and $s = s_j$. To be specific, in this distribution, the probability of a nonnegative integer vector $(d_1, \ldots, d_t)$ whose entries sum to s is

$$\frac{s!}{t^s \prod_{i=1}^{t} d_i!}.$$

It is easy to see that the multinomial degree distributions of the vertices in $V_1, \ldots, V_r$ are independent. Moreover, for use shortly, we note the following relation between $\mathbf{Multi}(t, s)|_{\geq k}$ and the truncated Poisson random variable defined in (1). From [5, Lemma 1], if X is any particular element of a vector distributed as $\mathbf{Multi}(t, s)|_{\geq k}$, then for fixed j and k, and $s - kt = \Theta(t)$,

$$\mathbf{P}(X = j) \sim \frac{\lambda_b^j}{f_k(\lambda_b) j!} \tag{11}$$

as $t \to \infty$, where $b = s/t$ and λ_b is the positive root λ of the equation

$$\frac{\lambda f_{k-1}(\lambda)}{f_k(\lambda)} = b. \tag{12}$$

It is easily seen that λ_b exists uniquely provided $b > k$. For instance, as in the proof of [21, Lemma 1], using $\frac{d}{d\lambda} f_k(\lambda) = f_{k-1}(\lambda)$, note that

$$\frac{d}{d\lambda} \frac{\lambda f_{k-1}(\lambda)}{f_k(\lambda)} = \frac{1}{\lambda} \mathbf{Var}\, Z(k, \lambda) > 0$$

for $Z(k,\lambda)$ defined in (1).

The process of iteratively deleting light vertices from $G_{r,n,m}$ we will model by a process applied to $G_{\mathcal{P}}$ whereby, at each step, one light vertex is randomly selected from each of the sets $L_1,\dots,L_r$, and deleted along with its incident edge. Insisting on one from each set means this process terminates as soon as any one of the sets of light vertices becomes empty, but as we shall see, this suffices for our purpose. If in such a step, the degree of any vertex drops to $k-1$, that vertex immediately fragments into $k-1$ light vertices (added to the appropriate set L_i). It is easy to check by induction that each eligible distribution of edge-ends into vertices is equally likely, and as a consequence we have the following.

Observation 1 *The random pseudograph after a given number of deletions, conditioned on the current values of* **t**, **s** *and* **e**, *has precisely the distribution of $G_{\mathcal{P}'}$ where $\mathcal{P}'$ is defined with those same current parameters.*

(For essentially the same argument applied to the nonmultipartite case, see the proof of [5, Lemma 6].) This gives the following.

Observation 2 *The degree distribution in V_j, conditioned on the values of s_j, t_j and e_j at each step of the process, is always truncated multinomial (and independent of the degree distributions in the other V_i).*

During the deletion process applied to $G_{\mathcal{P}}$, let $S_{j,i}$ and $T_{j,i}$ denote the random values of s_j and t_j respectively after i steps. Then deleting the edge containing a light vertex in L_j will have an effect on each of the other $r-1$ parts of the partition, in each part either deleting one light vertex or reducing the degree of a heavy vertex by 1, or fragmenting such a vertex into $k-1$ light ones if its degree drops to $k-1$. Each step deletes r edges. Thus, after i steps, each part (including vertices in $V_i\cup L_i$) will have total degree $m-ri$.

We will study the sequence of random values $(T_{j,i},S_{j,i})_{j=1,\dots,r}$ for $i=0,\dots,i_{\text{fin}}$ where i_{fin} is the (random) step in which the light vertices first disappear from one of the parts. Our analysis will stop when S_i-T_ik falls to ϵn for some predefined $\epsilon>0$ (which implies $m-ri>\epsilon n$). Initially, $(T_{j,i},S_{j,i})$ have random values which are sharply concentrated in the allocation model. The number of vertices of degree i in any one of the r parts is a.a.s. $ne^{-c}c^i/i!+o(n)$, and it follows (using exactly the same argument as in [5, Section 4]) that a.a.s.

$$T_{j,0}=e^{-c}f_k(c)n+o(n),\quad S_{j,0}=ce^{-c}f_{k-1}(c)n+o(n).$$

We next interpret the argument around equations (10–12) of [5] in our current setting. Just after the ith step of the algorithm, the probability that the edge containing the light vertex chosen for deletion from $L_{j'}$ intersects V_j ($j\neq j'$) is, by Observation 1,

$$\frac{S_{j,i}}{m-i+1}+o(1). \tag{13}$$

(We note that the $o(1)$ term accounts for the effect of the change in the values of the variables resulting when a number of the r light vertices have been deleted during this step, and assumes $m-i=\Omega(n)$.) Furthermore, conditional upon this event, if $v\in V_j$ is the vertex intersected, the probability that v has degree k (which causes S_j to decrease by a further $k-1$ and T_j by 1) is $k/S_{j,i}$ times the number of vertices in V_j of degree k, and hence, by Observation 2 and (11), is asymptotic to

$$\frac{k}{S_{j,i}}\times\frac{T_{j,i}\lambda_i^k}{f_k(\lambda_i)k!}$$

where $\lambda_i=\lambda_{S_i/T_i}$ as defined in (12), using $S_i-T_ik=\Theta(n)$. This is equal to

$$1-\frac{\lambda_iT_i}{S_i}+o(1)$$

using $\lambda^{k-1}/(k-1)!=f_{k-1}-f_k$ and (12).

It follows that the expected value of $T_{j,i+1}-T_{j,i}$, conditioned on the values of **s** and **t** at step i, is (with the above assumptions involving ϵ) asymptotically:

$$-\frac{(r-1)S_{j,i}}{m-ri}\left(1-\frac{\lambda_{j,i}T_{j,i}}{S_{j,i}}\right)$$

and for $S_{j,i+1}-S_{j,i}$ the expected change is asymptotically

$$-\frac{(r-1)S_{j,i}}{m-ri}\left(k-(k-1)\frac{\lambda_{j,i}T_{j,i}}{S_{j,i}}\right).$$

The argument of [5] can now be followed from just below its Equation (13). This applies the differential equation method, which calls for writing a real variable for each random variable, scaling i and the random variables by a factor n, and writing a differential equation to correspond to each calculated expected change above. The general differential equation theorem guarantees that the random variables a.a.s. track close to the solutions of the resulting differential equation system. In this case, we let $x=i/n$, $y_j=S_{i,j}/n$ and $z_j=T_{i,j}/n$. In place of Equations (15) and (16) of [5] we now have for $1\le j\le r$

$$z_j' = -\frac{(r-1)y_j}{c-rx}\left(1-\frac{\mu_jz_j}{y_j}\right), \tag{14}$$

$$y_j' = -\frac{(r-1)y_j}{c-rx}\left(k-(k-1)\frac{\mu_jz_j}{y_j}\right)=(k-1)z_j'-1 \tag{15}$$

where $\mu_j=\lambda_{y_j/z_j}$. Now [24, Theorem 1] gives that a.a.s.

$$S_{j,i}=ny_j(i/n)+o(n)\quad\text{and}\quad T_{j,i}=nz_j(i/n)+o(n)$$

for i in an appropriate range (see the argument in [5]). If we replace each y_j by y and z_j by z, these are exactly the same equations as occur for hypergraphs in the proof of [5, Theorem 3]. Hence (by uniqueness of the solutions of such differential equations), the equations have the same solution in any region where they are Lipschitz, in which all y_i are identical (for each i) to a function y, and similarly all z_i are identical to a function z. As shown in [5], the solution reaches a point where $c-rx-y=0$ before $y=0$ if and only if $c>c_{r,k}$. It follows that for $c>c_{r,k}$ the k-core exists a.a.s. Its size in each of the r parts is given asymptotically by yn, and the value of y when $c-rx-y=0$

is the same as for the standard hypergraph case. Moreover, the asymptotic distribution of degrees is also the same, which gives the other statements in the theorem on the number of edges and number of vertices of given degrees.

For the case $c < c_{r,k}$, the argument in [5] shows that for $c < c_{r,k}$, where $c-rx-y = 0$ is reached, this corresponds to $z = O(\epsilon)$. In this case the k-core, if it exists, a.a.s. has size at most ϵn for any fixed $\epsilon > 0$. One can also observe that if ϵ is sufficiently small, then a random r-partite hypergraph with $cn + o(n)$ edges a.a.s. has no k-core of size less than ϵn. This comes from a simple first moment calculation, which shows that the expected number of sets of at most ϵn vertices that contain at least $k\epsilon n/r$ hyperedges tends to 0 as $n \to \infty$. ∎

Acknowledgements

We thank the partial support given by the Brazilian National Institute of Science and Technology for the Web (grant MCT/CNPq 573871/2008-6), Canada Research Chairs Program and NSERC (Nicholas Wormald) and CNPq Grant 30.5237/02-0 (Nivio Ziviani).

References

[1] Bollobás, B., 2001. Random graphs, 2nd Edition. Vol. 73 of Cambridge Studies in Advanced Mathematics. Cambridge University Press, Cambridge.

[2] Botelho, F., Pagh, R., Ziviani, N., 2007. Simple and space-efficient minimal perfect hash functions. In: Proceedings of the 10th Workshop on Algorithms and Data Structures (WADs'07). Springer LNCS vol. 4619, pp. 139–150.

[3] Botelho, F., September 2008. Near-optimal space perfect hashing algorithms. Ph.D. thesis, Federal University of Minas Gerais, supervised by Nivio Ziviani, `http://hdl.handle.net/1843/RVMR-7LAMRS`.

[4] Botelho, F., Pagh, R., Ziviani, N., 2007. Practical Perfect Hashing in Nearly Optimal Space. Information Systems (Submitted in 2010).

[5] Cain, J., Wormald, N. C., 2006. Encores on cores. Electronic Journal of Combinatorics 13 (1).

[6] Cooper, C., 2004. The cores of random hypergraphs with a given degree sequence. Random Structures & Algorithms 25 (4), 353–375.

[7] Erdős, P., Rényi, A., 1959. On random graphs I. Pub. Math. Debrecen 6, 290–297.

[8] Erdős, P., Rényi, A., 1960. On the evolution of random graphs. Magyar Tud. Akad. Mat. Kutató Int. Közl. 5, 17–61.

[9] Erdős, P., Rényi, A., 1961. On the evolution of random graphs. Bull. Inst. Internat. Statist. 38, 343–347.

[10] Erdős, P., Rényi, A., 1961. On the strength of connectedness of a random graph. Acta Mathematica Scientia Hungary 12, 261–267.

[11] Flajolet, P., Knuth, D. E., Pittel, B., 1989. The first cycles in an evolving graph. Discrete Math 75, 167–215.

[12] Fox, E. A., Heath, L. S., Chen, Q., and Daoud, A. M. 1992. Practical Minimal Perfect Hash Functions For Large Databases. Communications of the ACM, 35 (1), 105–121.

[13] Janson, S., 1987. Poisson convergence and poisson processes with applications to random graphs. Stochastic Processes and their Applications 26, 1–30.

[14] Janson, S., Łuczak, T., Ruciński, A., 1990. An exponential bound for the probability of nonexistence of a specified subgraph in a random graph. In: Random graphs '87 (Poznań, 1987). Wiley, Chichester, pp. 73–87.

[15] Janson, S., Łuczak, T., Ruciński, A., 2000. Random graphs. Wiley-Inter.

[16] Majewski, B., Wormald, N. C., Havas, G., Czech, Z., 1996. A family of perfect hashing methods. The Computer Journal 39 (6), 547–554.

[17] Molloy, M., 2005. Cores in random hypergraphs and boolean formulas. Random Structures & Algorithms 27 (1), 124–135.

[18] Molloy, M., Reed, B., 1995. A critical point for random graphs with a given degree sequence. Random Structures and Algorithms 6, 161–179.

[19] Molloy, M., Reed, B., 1998. The size of the giant component of a random graph with a given degree sequence. Combinatorics, Probability and Computing 7, 295–305.

[20] Pittel, B., Spencer, J., Wormald, N. C., 1996. Sudden emergence of a giant k-core in a random graph. Journal of Combinatorial Theory Series B 67 (1), 111–151.

[21] Pittel, B., Wormald, N. C., 2003. Asymptotic enumeration of sparse graphs with a minimum degree constraint, *J. Combinatorial Theory, Series A* **101**, 249–263.

[22] Pittel, B., Wormald, N. C., 2005. Counting connected graphs inside-out. Journal of Combinatorial Theory Series B 93 (2), 127–172.

[23] Radhakrishnan, J., 1992. Improved bounds for covering complete uniform hypergraphs. Information Processing Letters 41, 203–207.

[24] Wormald, N. C., 1995. Differential equations for random processes and random graphs. Annals of Applied Probability (5), 1217–1235.

[25] Wormald, N. C., 1999. Models of random regular graphs. In Surveys in Combinatorics. London Mathematical Society Lecture Note Series 267, edited by J.D. Lamb and D.A. Preece, Cambridge University Press, Cambridge, 239–298.