Load balancing and orientability thresholds for random hypergraphs

Pu Gao[*] and Nicholas Wormald[†]
Department of Combinatorics and Optimization
University of Waterloo
p3gao@math.uwaterloo.ca, nwormald@math.uwaterloo.ca

Abstract

Let $h > w > 0$ be two fixed integers. Let H be a random hypergraph whose hyperedges are uniformly of size h. To *w-orient* a hyperedge, we assign exactly w of its vertices positive signs with respect to the hyperedge, and the rest negative. A (w, k)-orientation of H consists of a w-orientation of all hyperedges of H, such that each vertex receives at most k positive signs from its incident hyperedges. When k is large enough, we determine the threshold of the existence of a (w, k)-orientation of a random hypergraph. The (w, k)-orientation of hypergraphs is strongly related to a general version of the off-line load balancing problem. The graph case, when $h = 2$ and $w = 1$, was solved recently by Cain, Sanders and Wormald and independently by Fernholz and Ramachandran, thereby settling a conjecture made by Karp and Saks. Motivated by a problem of cuckoo hashing, the special hypergraph case with $w = k = 1$, was solved in three separate preprints dating from October 2009, by Frieze and Melsted, by Fountoulakis and Panagiotou, and by Dietzfelbinger, Goerdt, Mitzenmacher, Montanari, Pagh and Rink.

[*]The research reported in this paper forms part of this author's Ph.D. thesis "Generation and properties of random graphs and analysis of randomized algorithms" submitted to the University of Waterloo in 2009.

[†]Research supported by the Canadian Research Chairs Program and NSERC

1 Introduction

In this paper we consider a generalisation to random hypergraphs of a commonly studied orientation problem on graphs. An *h-hypergraph* is a hypergraph whose hyperedges are of all of cardinality h. Let $h > w$ be two given positive integers. We consider $\mathcal{G}_{n,m,h}$, the probability space of the set of all h-hypergraphs on n vertices and m hyperedges with the uniform distribution. A hyperedge is said to be *w-oriented* if exactly w distinct vertices in it are marked with positive signs with respect to the hyperedge. The *indegree* of a vertex is the number of positive signs it receives. Let k be a positive integer. A (w, k)-orientation of an h-hypergraph is a w-orientation all hyperedges such that each vertex has indegree at most k. If such a (w, k)-orientation exists, we say the hypergraph is (w, k)-*orientable*; for $w = 1$ we simply say *k-orientable.* Of course, being able to determine the (w, k)-orientability of an h-hypergraph H for all k solves the optimisation problem of minimising the maximum indegree of a w-orientation of H. If a graph (i.e. the case $h = 2$) is $(1, k)$-oriented, we may orient each edge of the graph in the normal fashion towards its vertex of positive sign, and we say the graph is *k-oriented*.

Note that a sufficiently sparse hypergraph is easily (w, k)-orientable. On the other hand, a trivial requirement for (w, k)-orientability of an h-hypergraph with m edges is $m \leq kn/w$, since any w-orientation has average indegree mw/n. In this paper, we show the existence and determine the value of a sharp threshold (defined more precisely in Section 2) at which a random h-hypergraph from $\mathcal{G}_{n,m,h}$ fails to be (w, k)-orientable, provided k is a sufficiently large constant. We show that the threshold is the same as the threshold at which a certain type of subhypergraph achieves a critical density. In the above, as elsewhere in this paper, the phrase "for k sufficiently large" means for k larger than some constant depending only on w and h.

1.1 Applications to load balancing, and some previous results

The hypergraph orientation problem is motivated by classical load balancing problems which have appeared in various guises in computer networking. A seminal result of Azar, Broder, Karlin and Upfal [2] is as follows. Throw n balls sequentially into n bins, with each ball put into the least full of $h \geq 2$ randomly chosen boxes. Then, with high probability, by the time all balls are allocated, no bin contains many more than $(\ln \ln n)/\ln h$ balls. If, instead, each ball is placed in a random bin, a much larger maximum value is likely to occur, approximately $\ln n/(\ln \ln n)$. This surprisingly simple method of reducing the maximum is widely used for load balancing. It has become known as the multiple-choice paradigm, the most common version being two-choice, when $h = 2$.

One application of load balancing occurs when work is spread among a group of computers, hard drives, CPUs, or other resources. In the on-line version, the jobs arrive sequentially and are assigned to separate machines. To save time, the load balancer decides which machine a job goes to after checking the current load of only a few (say h) machines. The goal is to minimise the maximum load of a machine. Mitzenmacher, Richa and Sitaraman [19] survey the history, applications and techniques relating to this. In particular, Berenbrink, Czumaj, Steger and Vőcking [2, 3] show an achieveable maximum load is $m/n + O(\log \log n)$ for m jobs and n machines when $h \geq 2$.

Another application of load balancing mentioned by Cain, Sanders and the second author [5] is the disk scheduling problem, in which any w out of h pieces of data are needed to reconstruct a logical data block. Individual pieces can be initially stored on different disks. Such an arrangement has advantageous fault tolerance features to guard against disk failures. It is also good for load balancing: when a request for a data block arrives, the scheduler can choose any w disks among the h relevant ones. See Sanders, Egner and Korst [20] for further references.

These load balancing problems correspond to the (w, k)-orientation problem for h-uniform hypergraphs, with $w = 1$ in the case of the job scheduling problem. The machines (bins) are vertices and a job (ball) is an edge consisting of precisely the set of machines to which it can be allocated. A job is allocated to a machine by assigning a positive sign to that vertex. The maximum load is then equal to the maximum indegree of a vertex in the (w, k)-oriented hypergraph.

1.2 Off-line version

In the on-line versions, the edges of the hypergraph arrive one by one, but the off-line version, in which the edges are all exposed at the start, is also of interest. For instance, in the disk scheduling problem, the scheduler may be able to process a large number of requests together, to balance the load better. Of course, the on-line and off-line versions are the same if $h = 1$, i.e. there is no choice. For $h = 2$, the off-line version experiences an even better improvement than the on-line one. If $m < cn$ items are allocated to n bins, for c constant, the expected maximum load is bounded above by some constant c' depending on c.

To our knowledge, previous theoretical results concern only the case $w = 1$ (this also applies to on-line problem). For $w = 1$ it is well known that an optimal off-line solution, i.e. achieving minimum possible maximum load, can be found in polynomial time ($O(m^2)$) by solving a maximum flow problem. As explained in [5], it is desirable to achieve fast algorithms that are close to optimal with respect to the maximum load. There are linear time algorithms that achieve maximum load $O(m/n)$ [10, 16, 18].

A central role in solutions of the off-line orientation problem with $(w, h) = (1, 2)$ is played by the *k-core* of a graph, being the largest subgraph with minimum degree at least k. The sharp threshold for the k-orientability of the random graph $\mathcal{G}(n, m) = \mathcal{G}_{n,m,2}$ was found in [5], and simultaneously by Fernholz and Ramachandran [13]. These were proofs of a conjecture of Karp and Saks that this threshold coincides with the threshold at which the $(k + 1)$-core has average degree at most $2k$. (It is obvious that a graph cannot be k-oriented if it has a subgraph of average degree greater than $2k$.) In each case, the proof analysed a linear time algorithm that finds a k-orientation *asymptotically almost surely* (a.a.s.; i.e. with probability $1 - o(1)$ as $n \to \infty$) when the mean degree of the $(k + 1)$-core is slightly less than $2k$. In this sense, the algorithms are asymptotically optimal since the threshold for the algorithms succeeding coincides with the threshold for existence of a k-orientation. The proof in [13] was significantly simpler than the other, which was made possible because a different algorithm was employed. It used a trick of "splitting vertices" to postpone decisions and thereby reduced the number of variables to be considered.

Since the preparation of the original version of the present paper, three preprints appeared by Frieze and Melsted [11], Fountoulakis and Panagiotou [12], and by Dietzfelbinger, Goerdt, Mitzenmacher, Montanari, Pagh and Rink [9] which independently study the threshold of $(1, 1)$-orientability of $\mathcal{G}_{n,m,h}$, i.e. the case $w = k = 1$. This has applications to cuckoo hashing. However, there seems to be no easy way to extend the proofs in [9, 11, 12] to solve for the case $k > 1$, even when $w = 1$.

1.3 Our contribution

We solve the generalisation of the conjecture of Karp and Saks mentioned above, for fixed $h > w > 0$, provided k is sufficiently large. That is, we find the threshold of (w, k)-orientation of random h-hypergraphs in $\mathcal{G}_{n,m,h}$. The determination of this threshold helps to predict loads in the off-line w-out-of-h disk scheduling problem, where the randomness of the hypergraph is justified by the random intial allocation of file segments to disks. We believe furthermore that the characterisation of the threshold in terms of density of a type of core, and possibly our method of proof, will potentially help lead to fast algorithms for finding asymptotically optimal orientations.

Our approach has a significant difference from that used in the graph case when $(w, h) = (1, 2)$. The algorithm used in [13] does not seem to apply in the hypergraph case, at least, splitting vertices cannot be done without creating hyperedges of larger and larger size. The algorithm used by [5], on the other hand, generalises in an obvious way, but it is already very complicated to analyse in the graph case, and the extension of the analysis to the hypergraph case seems formidable. However, in common with both approaches, we first find what we call the $(w, k+1)$-core in the hypergraph, which is an analogue of the $(k+1)$-core in graphs. We use the differential equation method to determine the size and density of this core. From here, we use the natural representation of the orientation problem in terms of flows. It is quite easy to generalise the network flow formulation from the case $h = 2$, $w = 1$ to the arbitrary case, giving a problem that can be solved in time $O(m^2)$ for $m = \Theta(n)$. Unlike the approaches for the graph case, we do not study an algorithm that solves the load balancing problem. Instead, we use the minimum cut characterisation of the maximum flow to show that a.a.s. the hypergraph can be (w, k)-oriented if and only if the density of its $(w, k+1)$-core is below a certain threshold. When the density of the $(w, k+1)$-core is above this threshold, it is trivially too dense to be (w, k)-oriented.

Even the case $w = 1$ of our result gives a significant generalisation of the known results. We prove that the threshold of the orientability coincides with the threshold at which certain type of density (in the case $w = 1$, this refers to the average degree divided by h) of the $(w, k+1)$-core is at most k, and also the threshold at which a certain type of induced subgraph (in the case $w = 1$, this refers to the standard induced subgraph) does not appear. For the graph case, our method provides a new proof (for sufficiently large k) of the Karp-Saks conjecture that we believe is simpler than the proofs of [5] and [13].

We give precise statements of our results, including definition of the $(w, k+1)$-core, in Section 2. In Section 3 we study the properties of the $(w, k+1)$-core. In Section 4, we formulate the appropriate network flow problem, determine a canonical minimum cut for a network corresponding to a non-(w, k)-orientable hypergraph, and give conditions under which such a minimum cut can exist. Finally, in Section 5, we show that for k is sufficiently large, such a cut a.a.s. does not exist when the density of the core is below a certain threshold. Proofs not given in this paper will be provided in a longer version [14].

2 Main results

Let $h > w > 0$ and $k \geq 2$ be given constants. For any h-hypergraph H, we examine whether a (w, k)-orientation exists. We call a vertex *light* if the degree of the vertex is at most k. For any light vertex v, we can give v the positive sign with respect to every hyperedge x that is incident to v (we call this *partially orienting* x towards to v), without violating the condition that each vertex has indegree at most k. Remove v from H, and for each hyperedge x incident to v, simply update x by removing v. Then the size of x decreases by 1, and it has one less vertex that needs to be given a positive sign. If a hyperedge becomes of size $h - w$, we can simply remove that hyperedge from the hypergraph. Repeating this until no light vertex exists, we call the remaining hypergraph $\widehat{H}$ the $(w, k+1)$-*core* of H. Every vertex in $\widehat{H}$ has degree at least $k+1$, and every hyperedge in $\widehat{H}$ of size $h - j$ requires a $(w - j)$-orientation in order to complete a w-orientation of the original hyperedge in H.

Instead of considering the probability space $\mathcal{G}_{\bar{n},\bar{m},h}$, we may consider $\mathcal{M}_{\bar{n},\bar{m},h}$, the probability space of random uniform multihypergraphs with $\bar{n}$ vertices and $\bar{m}$ hyperedges, such that each hyperedge x is of size h, and each vertex in x is chosen independently, uniformly at random from

$[\bar{n}]$. Actually $\mathcal{M}_{\bar{n},\bar{m},h}$ may be a more accurate model for the off-line load balancing problem in some applications, and as we shall see, results for the non-multiple edge case can be deduced from it. For a nonnegative integer vector $\mathbf{m} = (m_2, \dots, m_h)$, we also define the probability space $\mathcal{M}_{n,\mathbf{m}}$, being the obvious generalisation of $\mathcal{M}_{n,m,h}$ to non-uniform multihypergraphs in which m_i is the number of hyperedges of size i.

All our asymptotic notation refers to $n \to \infty$. For clarity, we consider $H \in \mathcal{M}_{\bar{n},\bar{m},h}$. We use n, m_{h-j} and μ for the number of vertices, the number of hyperedges of size $h-j$ and the average degree of $\widehat{H}$. We parametrise the number $\bar{m}$ of edges in the hypergraphs under study by letting $\bar{\mu} = \bar{\mu}(n)$ denote $h\bar{m}/\bar{n}$, the average degree of $H \in \mathcal{M}_{\bar{n},\bar{m},h}$ (or of $H \in \mathcal{G}_{\bar{n},\bar{m},h}$).

Our first observation concerns the distribution of $\widehat{H}$ and its vertex degrees. Let $\mathrm{Multi}(n, m, k{+}1)$ denote the multinomial distribution of n integers summing to m, restricted to each of the integers being at least $k+1$. We call this *truncated multinomial*.

Proposition 2.1 *Let $h > w \geq 1$ be two fixed integers. Let $H \in \mathcal{M}_{\bar{n},\bar{m},h}$ and let $\widehat{H}$ be its $(w, k{+}1)$-core. Conditional on its number n of vertices and numbers m_{h-j} of hyperedges of size $h-j$ for $j = 0, \dots, w-1$, the random hypergraph $\widehat{H}$ is distributed uniformly randomly on multihypergraphs with all vertices of degree at least $k+1$ and having the same parameters n and m_{h-j} for each j. Furthermore, the degree distribution of $\widehat{H}$ is the truncated multinomial distribution* $\mathrm{Multi}(n, m, k{+}1)$ *where $m = \sum (h-j) m_{h-j}$.*

The following theorem shows that the size and the number of hyperedges of $\widehat{H}$ are highly concentrated around the solution of a system of differential equations. It covers any $h > w \geq 2$ for sufficiently large k. For special case $w = 1$, the concentration result was already known for all $k \geq 0$, for example in [5, Theorem 3]. The particular system of differential equations is given in the long version of the paper [14].

Theorem 2.2 *Let $h > w \geq 2$ be two fixed integers. Assume $ck \leq \bar{\mu} := h\bar{m}/\bar{n}$ for some constant $c > 1$. Let $H \in \mathcal{M}_{\bar{n},\bar{m},h}$ and let $\widehat{H}$ be its $(w, k+1)$-core. Let n be the number of vertices and m_{h-j} the number of hyperedges of size $h-j$ of $\widehat{H}$. Then, provided k is sufficiently large, there are constants $\alpha > 0$ and $\beta_{h-j} > 0$, which are determined by the solution of a certain differential equation system that depends only on $\bar{\mu}$, k, w and h, for which a.a.s. $n \sim \alpha \bar{n}$ and $m_{h-j} \sim \beta_{h-j} \bar{n}$ for $0 \leq i \leq w-1$. The same conclusion (with the same constants) holds if $H \in \mathcal{G}_{\bar{n},\bar{m},h}$.*

The differential equation system is rather complicated so it is not included here.

Let $\mathcal{P}$ be a hypergraph property and let $\mathcal{M}_{n,m,h} \in \mathcal{P}$ denote the event that a random hypergraph from $\mathcal{M}_{n,m,h}$ has the property $\mathcal{P}$. Following [1, Section 10.1], we say that $\mathcal{P}$ has a *sharp threshold function* $f(n)$ if for any constant $\epsilon > 0$, $\mathbf{P}(\mathcal{M}_{n,m,h} \in \mathcal{P}) \to 1$ when $m \leq (1-\epsilon) f(n)$, and $\mathbf{P}(\mathcal{M}_{n,m,h} \in \mathcal{P}) \to 0$ when $m \geq (1+\epsilon) f(n)$.

Recall that n denotes the number of vertices and m_{h-j} denotes the number of hyperedges of size $h-j$ in $\widehat{H}$. Let $\kappa(\widehat{H})$ denote $\sum_{j=0}^{w-1} (w-j) m_{h-j}/n$. Then $\kappa(\widehat{H})$ defines a certain type of density of $\widehat{H}$. We say that a hypergraph H has *property $\mathcal{T}$* if its $(w, k+1)$-core $\widehat{H}$ satisfies the condition that $\kappa(\widehat{H})$ is at most k. (Since w and k are fixed, we often drop them from the notation.) The following theorem, proved using Theorem 2.2, immediately gives the corollary that there is a sharp threshold function for the property $\mathcal{T}$.

Theorem 2.3 *Let $H \in \mathcal{M}_{\bar{n},\bar{m},h}$. Let $\bar{\mu}$ be the average degree of H and let $\widehat{H}$ be the $(w, k+1)$-core of H. Then for all sufficiently k, there exists a strictly increasing function $c(\bar{\mu})$ of $\bar{\mu}$, such that for any fixed $c_2 > c_1 > 1$ and for any $c_1 k < \bar{\mu} < c_2 k$, a.a.s. $\kappa(\widehat{H}) \sim c(\bar{\mu})$.*

Corollary 2.4 *There exists a sharp threshold function $f(\bar{n})$ for the hypergraph property $\mathcal{T}$ in $\mathcal{M}_{\bar{n},\bar{m},h}$ provided k is sufficiently large.*

The function $c(\bar{\mu})$ in the theorem, and the threshold function in the corollary, are determined by the solution of the differential equation system referred to in Theorem 2.1.

We have defined a (w, k)-orientation of a uniform hypergraph in Section 1. We can similarly define a (w, k)-orientation of a non-uniform hypergraph G with sizes of hyperedges between $h-w+1$ and h to be a simultaneous $(w-j)$-orientation of each hyperedge of size $h-j$ such that every vertex has indegree at most k. By counting the positive signs in orientations, we see that if property $\mathcal{T}$ fails, there is no (w, k)-orientation of $\widehat{H}$, and hence there is no (w, k)-orientation of H.

For a nonnegative integer vector $\mathbf{m} = (m_{h-w+1}, \dots, m_h)$, let $\mathcal{M}(n, \mathbf{m}, k + 1)$ denote $\mathcal{M}_{n,\mathbf{m}}$ restricted to multihypergraphs with minimum degree at least $k+1$. By Proposition 2.1, $\mathcal{M}(n, \mathbf{m}, k+1)$ has the distribution of the $(w, k + 1)$-core of $H \in \mathcal{M}_{\bar{n},\bar{m},h}$ conditioned on the number of vertices being n and the number of hyperedges of each size being given by $\mathbf{m}$. To emphasise the difference, we will use G to denote a not-necessarily-uniform hypergraph in cases where we might use H for a uniform hypergraph.

Given a vertex set S, we say a hyperedge x is *partially contained in S* if $|x \cap S| \geq 2$.

Definition 2.5 *Let $0 < \gamma < 1$. We say that a multihypergraph G has property $\mathcal{A}(\gamma)$ if for all $S \subset V(G)$ with $|S| < \gamma|V(G)|$ the number of hyperedges partially contained in S is strictly less than $k|S|/2w$.*

In the following theorem, $\mathbf{m} = \mathbf{m}(n)$ denotes an integer vector for each n.

Theorem 2.6 *Let γ be any constant between 0 and 1. Then there exists a constant $N > 0$ depending only on γ, such that for all $k > N$ and any $\epsilon > 0$, if $\mathbf{m}(n)$ satisfies $\sum_{j=0}^{w-1}(w-j)m_{h-j}(n) \leq kn - \epsilon n$ for all n, then $G \in \mathcal{M}(n, \mathbf{m}(n), k + 1)$ a.a.s. either has a (w, k)-orientation or does not have property $\mathcal{A}(\gamma)$.*

We show in the forthcoming Corollary 4.2 that for certain values of γ, a.a.s. $\widehat{H}$ has property $\mathcal{A}(\gamma)$ if $\bar{\mu}$ is constrained to be at most hk/w. We will combine this with Corollary 2.4 and Theorem 2.6 and a relation we will show between $\mathcal{M}_{\bar{n},\bar{m},h}$ and $\mathcal{G}_{\bar{n},\bar{m},h}$ (Lemma 3.1), to obtain the following.

Corollary 2.7 *Let $h > w > 0$ be two given integers and k be a sufficiently large constant. Let $f(\bar{n})$ be the threshold function of property $\mathcal{T}$ whose existence is asserted in Corollary 2.4. Then $f(\bar{n})$ is a sharp threshold for the (w, k)-orientability of $\mathcal{M}_{\bar{n},\bar{m},h}$ and $\mathcal{G}_{\bar{n},\bar{m},h}$.*

For any vertex set $S \subset V(H)$, define *the subgraph induced by S with parameter w* to be the subgraph of G on vertex set S whose set of hyperedges is $\{x' = x \cap S : x \in H,\ s.t.\ |x'| \geq h-w+1\}$. Call this hypergraph H_S. Let $d(H_S)$ denote the degree sum of vertices in the hypergraph H_S and let $e(H_S)$ denote the number of hyperedges in H_S. From the above results we will obtain the following.

Corollary 2.8 *The following three graph properties have the same sharp threshold in $\mathcal{M}_{\bar{n},\bar{m},h}$ and in $\mathcal{G}_{\bar{n},\bar{m},h}$.*

(i) H is (w, k)-orientable.

(ii) H has property $\mathcal{T}$.

(iii) There exists no $H' \subset H$ as an induced subgraph with parameter w such that $d(H') - (h - w)e(H') > ks$.

The proofs of Proposition 2.1, Theorem 2.2, 2.3 and 2.6 will appear in [14].

3 Analysing the size and density of the $(w, k+1)$-core

A model of generating random graphs via multigraphs, used by Bollobás and Frieze [4] and Chvatál [7], is described as follows. Let $\mathcal{P}_{\bar{n},\bar{m}}$ be the probability space of functions $g : [\bar{m}] \times [2] \to [\bar{n}]$ with the uniform distribution. Equivalently, $\mathcal{P}_{\bar{n},\bar{m}}$ can be described as the uniform probability space of allocations of $2\bar{m}$ balls into $\bar{n}$ bins. A probability space of random multigraphs can be obtained by taking $\{g(i,1), g(i,2)\}$ as an edge for each i. This model can easily be extended to generate non-uniformly random multihypergraphs by letting $\mathbf{m} = (m_2, \dots, m_h)$ and taking $\mathcal{P}_{\bar{n},\mathbf{m}} = \{g : \cup_{i=2}^{h}[m_i] \times [i] \to [\bar{n}]\}$. Let $\mathcal{M}_{\bar{n},\mathbf{m}}$ be the probability space of random multihypergraphs obtained by taking each $\{g(j,1), \dots, g(j,i)\}$ as a hyperedge, where $j \in [m_i]$ and $2 \le i \le h$. (Loops and multiple edges are possible.) Note that $\mathcal{M}_{\bar{n},\mathbf{m}}$, where $\mathbf{m} = (m_2) = (\bar{m})$, is a random multigraph; it was shown in [7] that if this is conditioned on being simple (i.e. no loops and no multiple edges), it is equal to $\mathcal{G}_{\bar{n},\bar{m},2}$, and that the probability of a multigraph in $\mathcal{M}_{\bar{n},(\bar{m})}$ being simple is $\Omega(1)$ if $\bar{m} = O(\bar{n})$. This result is easily extended to the following result, using the same method of proof.

Lemma 3.1 *Assume $h \ge 2$ is a fixed integer and $\mathbf{m} = (m_2, \dots, m_h)$ is a non-negative integer vector. Assume further that $\sum_{i=2}^{h} m_i = O(\bar{n})$. Then the probability that a multihypergraph in $\mathcal{M}_{\bar{n},\mathbf{m}}$ is simple is $\Omega(1)$.*

Cain and the second author [6] recently introduced a new model to analyse the k-core of a random (multi)graph or (multi)hypergraph, including its size and degree distribution. This model is called the *pairing-allocation* model. A generalisation of this, the *partition-allocation* model, is defined below. We will use this model to prove Theorem 2.6 and to analyse a randomized algorithm called the RanCore algorithm, defined later in this section, which outputs the $(w, k+1)$-core of a given h-hypergraph.

Given $h \ge 2$, n, $\mathbf{m} = (m_2, \dots, m_h)$, $\mathbf{L} = (l_2, \dots, l_h)$ and a nonnegative integer k such that $D - \ell \ge kn$, where $D = \sum_{i=2}^{h} i m_i$ and $\ell = \sum_{i=2}^{h} l_i$, let V be a set of n bins, and $\mathbf{M}$ a collection of pairwise disjoint sets $\{M_1, \dots, M_h\}$, where M_i is a set of $i m_i$ balls partitioned into parts, each of size i, for all $2 \le i \le h$. Let Q be an additional bin aside from V. It may assist the reader to know that Q 'represents' all the hyperedge incidences at vertices of degree less than k, and l_i is the number of these incidences in edges of size i. The *partition-allocation* model $\mathcal{P}(V, \mathbf{M}, \mathbf{L}, k)$ is the uniform probability space whose elements are the allocations of balls to bins defined as follows. Let $\mathcal{C} = \{c_2, \dots, c_h\}$ be a set of colours. Colour the balls in M_i with c_i. (The role of the colours is only to denote the size of the part a ball lies in.) Then allocate the D balls uniformly at random (u.a.r.) into the bins in $V \cup \{Q\}$, such that the following constraints are satisfied:

(i) Q contains exactly ℓ balls;

(ii) each bin in V contains at least k balls;

(iii) for any $2 \le i \le h$, the number of balls with colour c_i that are contained in Q is l_i.

We call Q the *light* bin and all bins in V *heavy*. To assist with the analysis in some situations, we consider the following algorithm which clearly generates a probability space equivalent to $\mathcal{P}(V, \mathbf{M}, \mathbf{L}, k)$. We call this alternative the *allocation-partition* algorithm since it allocates before partitioning the balls. First, allocate D balls randomly into bins $\{Q\} \cup V$ with the restriction that Q contains exactly ℓ balls and each bin in V contains at least k balls. Then colour the balls u.a.r. with the following constraints:

(i) exactly $i m_i$ balls are coloured with c_i;

(ii) for each $i = 2, \dots, h$, the number of balls with colour c_i contained in Q is exactly l_i.

Finally, take u.a.r. a partition of the balls such that for each $i = 2, \dots, h$, all balls with colour c_i are partitioned into parts of size i.

To prove Theorem 2.2, we will convert the problem to a question about $\mathcal{P}(V, \mathbf{M}, \mathbf{L}, k+1)$, in particular the $(w, k+1)$-core of the hypergraph induced in the obvious way by the bins containing at least $k+1$ balls.

A deletion algorithm producing the k-core of a random multigraph was analysed in [6]. The differential equation method [22] was used to analyse the size and the number of hyperedges of the final k-core. The degree distribution of the k-core was shown to be a truncated multinomial. We now extend this deletion algorithm to find the $(w, k+1)$-core of H in $\mathcal{M}_{\bar{n},\bar{m},h}$ and $\mathcal{G}_{\bar{n},\bar{m},h}$, analysing it using the partition-allocation model. We also prove Proposition 2.1 by considering the allocation-partition algorithm which generates $\mathcal{P}(V, \mathbf{M}, \mathbf{0}, k+1)$.

The deletion algorithm referred to above is the following, and is expressed in the setting of representing multihypergraphs using bins for vertices, where each hyperedge x is a set $h(x)$ of $|x|$ balls. Initially let LV be the set of all light vertices/bins, and let $\overline{LV} = V(H) \setminus LV$ be the set of heavy vertices. A *light* ball is any ball contained in LV.

RanCore Algorithm to derive the $(w, k+1)$-core
Input: an h-hypergraph H. Set $t := 0$.
 While neither LV nor $\overline{LV}$ is empty,
  t:=t+1;
  Remove all empty bins;
  U.a.r. choose a light ball v, let x be the hyperedge that contains v, and let u be the vertex/bin that contains v;
  If $|x| \geq h - w + 2$, update x with $x \setminus \{u\}$,
   otherwise, remove this hyperedge x from the current hypergraph. If any vertex/bin $u \in \overline{LV}$ becomes light, move u to LV together with all balls in it;
 If LV is empty, ouput the remaining hypergraph, otherwise, output the empty graph.

The output is clearly the $(w, k+1)$-core of the input hypergraph H.

Theorems 2.2 and 2.3 are proved in [14] by analysing the RanCore algorithm and the differential equations that result. These equations can be numerically solved when the values of h, w, k and μ are given. Table 3 gives the results of some computations, where h, w and k are given, $\widetilde{\mu}$ is the average degree of the hypergraph at the threshold for $\mathcal{T}$ given in Corollary 2.4, and $\widehat{\mu}$ is the corresponding average degree of its core $\widehat{H}$. By Corollary 2.7, discussed in the next section, $\widetilde{\mu}$ is also our main target, the threshold for orientability. Note that $\widehat{\mu}$ must be at least hk/w by the definition of property $\mathcal{T}$, and that it follows from the trivial upper bound of the orientability threshold given in the introduction part that $\widetilde{\mu}$ is at most hk/w.

h	w	k	$\widetilde{\mu}$	$\widehat{\mu}$
3	2	4	5.485	6.65086
3	2	10	14.766	15.5872
3	2	40	59.991	60.0773
10	2	4	19.99999	20.0003

Table 1: Some numerical computation results

4 The (w, k)-orientability of the $(w, k+1)$-core

In this section we prove Corollary 2.8 assuming Theorem 2.6, and study the basic network flow formulation of the problem that is used in the next section to prove Theorem 2.6.

The proof of the following lemma is provided in [14].

Lemma 4.1 *Let $H \in \mathcal{M}_{\bar{n},\bar{m},h}$ and let $\widehat{H}$ be the $(w, k+1)$-core of H. Let c_1 be a constant that can depend on k, with the constraint that $2 \le c_1 < h^2 e^2 \bar{\mu}$. Let $0 < \gamma = \varphi(k, c_1) = \big(c_1/h^2 e^2 \bar{\mu}\big)^2$. Then a.a.s. for all $S \subset V(\widehat{H})$ with $|S| < \gamma n$, the number of hyperedges partially contained in S is less than $c_1|S|$.*

We next show that for any fixed $0 < \gamma < 1$, $\widehat{H}$ a.a.s. has property $\mathcal{A}(\gamma)$, defined in Definition 2.5.

Corollary 4.2 *Assume that $H \in \mathcal{M}_{\bar{n},\bar{m},h}$ with $\bar{m} \le hk/w$, and that $\widehat{H}$ is the $(w, k+1)$-core of H. Let $\gamma = e^{-4}h^{-6}/4$. Then provided $k \ge 4w$, a.a.s. either $\widehat{H}$ is empty or $\widehat{H}$ has property $\mathcal{A}(\gamma)$.*

Proof. Apply Lemma 4.1 with $c_1 = k/2w$. Clearly $c_1 < ch^2e^2k$, and $c_1 \ge 2$ provided $k \ge 4w$. Then $\gamma \le \phi(k, c_1)$. By Definition 2.5, $\widehat{H}$ a.a.s. has property $\mathcal{A}(\gamma)$. ∎

Proof of Corollary 2.7 Let $\widehat{H}$ be the $(k+1)$-core of the random multihypergraph $H \in \mathcal{M}_{\bar{n},\bar{m},h}$. Let $\epsilon > 0$ be any constant. By Theorem 2.3, there exists a constant $\delta > 0$, such that a.a.s. if $\bar{m} \le f(\bar{m}) - \epsilon\bar{n}$, then $\sum_{j=0}^{w-1}(w-j)m_{h-j} \le kn - \delta n$. By Theorem 2.6 and Corollary 4.2, there exists a constant N depending only on h and w such that provided $k > N$, $\widehat{H}$ a.a.s. has a (w, k)-orientation. On the other hand, if $\bar{m} \ge f(\bar{m}) + \epsilon\bar{n}$, then a.a.s. $\sum_{j=0}^{w-1}(w-j)m_{h-j} \ge kn + \delta n$, and hence clearly $\widehat{H}$ is not (w, k)-orientable. Therefore $f(\bar{n})$ is a sharp threshold function for the (w, k)-orientation of $\mathcal{M}_{\bar{n},\bar{m},h}$. By Lemma 3.1, $f(\bar{n})$ is also a sharp threshold function for the (w, k)-orientation of $\mathcal{G}_{\bar{n},\bar{m},h}$. ∎

Let G be a non-uniform multihypergraph with the sizes of its hyperedges between $h-w+1$ to h. In the rest of the paper, we will use the following notations. Let $E_{h-j} = \{x \in E(G) : |x| = h-j\}$. For any given $S \subset [n]$, let $m_{h-j,i}(S) := |\{x \in E_{h-j} : |x \cap S| = i\}|$ for any $0 \le i \le h-j$. When the context is clear we may drop S from the notation. Let $\overline{S}$ denote the set $[n] \setminus S$ and let $d(S)$ denote the sum of degrees of vertices in S.

Recall the definition of induced subgraph with parameter w above the statement of Corollary 2.8 in Section 2. The following Lemma generalises Hakimi's theorem [15, Theorem 4] for graphs, and is proved in [14] using network flows and the max-flow min-cut theorem in a fashion similar to that used for numerous combinatorial problems (see [8, 21] for examples). A similar network flow setting was used in [20, Section 3.3] to study the load balancing problem with $w = 1$ and $h = 2$.

Lemma 4.3 *A multihypergraph G with sizes of hyperedges between $h-w+1$ and h has a (w, k)-orientation if and only if*

$$d(H_S) - (h-w)e(H_S) \le k|S|, \quad \textit{for all} \;\; S \subset V(G).$$

The following corollary is immediate.

Corollary 4.4 *A hypergraph $H \in \mathcal{G}_{\bar{n},\bar{m},h}$ has a (w, k)-orientation if and only if for any $S \subset V(H)$,*

$$d(H_S) - (h-w)e(H_S) \le k|S|.$$

Proof of Corollary 2.8. This follows directly from Corollary 2.7 and Corollary 4.4. ∎

For any vertex set S, define

$$\partial^*(S) = d(S) - \sum_{j=0}^{w-1} \sum_{i=w-j+1}^{h-j} (i-(w-j))m_{h-j,i}, \tag{4.1}$$

which measures a type of expansion in the hypergraph. For each hyperedge x of size $h-j$ which intersects S with i vertices, its contribution to $\partial^*(S)$ is $w-j \geq 0$ if $i \geq w-j+1$ and $i \geq 0$ otherwise. Therefore $\partial^*(S) \geq 0$ for any S. The following lemma characterises the existence of the (w,k)-orientation of G in terms of $\partial^*(S)$. (For the proof, see [14].)

Lemma 4.5 *The following two graph properties of a multihypergraph G with sizes of hyperedges between $h-w+1$ and h are equivalent.*

$$\begin{array}{ll} (i) & d(H_S) - (h-w)e(H_S) \leq k|S|, \quad \textit{for all} \;\; S \subset V(G); \\ (ii) & \partial^*(S) \geq k|S| + \left(\displaystyle\sum_{j=0}^{w-1} (w-j)m_{h-j} \right) - kn, \quad \textit{for all} \;\; S \subset V(G). \end{array}$$

It follows from Lemma 4.3 and Lemma 4.5 that G is (w,k)-orientable if and only if Lemma 4.5 (ii) holds.

For any $S \subset V(G)$, let

$$q_{h-j}(S) = \sum_{i=1}^{h-j} i m_{h-j,i}, \quad \eta(S) = \sum_{j=0}^{w-1} \sum_{i=1}^{h-j-1} m_{h-j,i}. \tag{4.2}$$

In other words, $q_{h-j}(S)$ denotes the contribution to $d(S)$ from hyperedges of size $h-j$ and $\eta(S)$ denotes the number of hyperedges which intersect both S and $\overline{S}$. When the context is clear, we may use q_{h-j} and η instead to simplify the notation.

Recall that given a vertex set S, a hyperedge x is partially contained in S if $|x \cap S| \geq 2$. Let $\rho(S)$ denote the number of hyperedges partially contained in S and let $\nu(S)$ denote the number of hyperedges intersecting S.

The following lemma, proved in [14], shows that, instead of checking Lemma 4.5 (ii), we can check that certain other events do not occur.

Lemma 4.6 *Suppose that for some $S \subset V(G)$,*

$$\partial^*(S) < k|S| + \left(\sum_{j=0}^{w-1} (w-j)m_{h-j} \right) - kn. \tag{4.3}$$

Then all of the following hold:

$$\begin{array}{ll} (i) & \rho(\overline{S}) > k|\overline{S}|/w; \\ (ii) & \nu(S) < k|S|; \\ (iii) & (h-w)\rho(S) > d(S) - k|S|; \\ (iv) & \textit{If, in addition,} \;\; \displaystyle\sum_{j=0}^{w-1} \frac{w-j}{h-j} q_{h-j}(S) \geq (1-\delta)k|S| \; \textit{for some} \; \delta > 0, \;\; \textit{then} \; \eta(S) < h^2 \delta k|S|. \end{array}$$

5 Proof of Theorem 2.6

Recall from Section 2 that $\mathcal{M}(n, \mathbf{m}, k+1)$ is $\mathcal{M}_{n,\mathbf{m}}$, which is a random multihypergraph with given edge sizes, restricted to multihypergraphs with minimum degree at least $k+1$. In this section we discuss the only remaining proof, that of Theorem 2.6. This theorem relates the orientability of $\mathcal{M}(n, \mathbf{m}, k+1)$ to a kind of density. Recall that this probability space was important because, by Proposition 2.1, it gives the distribution of the $(w, k+1)$-core $\widehat{H}$ of $H \in \mathcal{M}_{\bar{n},\bar{m},h}$ conditioned on the values of n, the number of vertices and m_{h-j}, the number of hyperedges of size $h-j$ for each j, in the core.

It is clear, that given values of n and $\mathbf{m}$, the probability space of random multihypergraphs generated by $\mathcal{P}([n], \mathbf{M}, \mathbf{0}, k+1)$, with $|M_{h-j}| = (h-j)m_{h-j}$ $(h = 0, \ldots, w-1)$, is equivalent to $\mathcal{M}(n, \mathbf{m}, k+1)$. So we may, and do, make use of the partition-allocation model for proving results about $\mathcal{M}(n, \mathbf{m}, k+1)$. For the full proof of Theorem 2.6 we refer to [14]. An outline is as follows.

In this setting, $\epsilon > 0$ and $k \geq 2$ are fixed. Let G be a random multihypergraph from the probability space $\mathcal{M}(n, \mathbf{m}, k+1)$. Without loss of generality, we may assume that $\epsilon < \frac{1}{2}$ since ϵ may be taken arbitrarily small. By the hypothesis of Theorem 2.6, we consider only $\mathbf{m}$ such that $\sum_{j=0}^{w-1}(w-j)m_{h-j} \leq kn - \epsilon n$. We may also assume that $\sum_{j=0}^{w-1}(w-j)m_{h-j} \geq kn - 2\epsilon n$ since otherwise, by Theorem 2.3, we can simply add a set of random hyperedges so that the assumption holds. This is valid because (w, k)-orientability is a decreasing property (i.e. it holds in all subgraphs of G whenever G has the property).

Let $q_{h-j}(S)$ and $\eta(S)$ be as defined in (4.2). The partition-allocation model gives a good foundation for proving that a.a.s. has certain properties concerning the distribution of vertex degrees and intersections of hyperedge sets with vertex sets. Using this and various other probabilistic tools, we show that

(a) the probability that $G \in \mathcal{M}(n, \mathbf{m}, k+1)$ has property $\mathcal{A}(\gamma)$ and contains some set S with $|S| < \gamma n$ for which both Lemma 4.6(ii) and (iii) holds is $o(1)$;

(b) there exists $\delta > 0$, such that when k is large enough, a.a.s. $\sum_{j=0}^{w-1} \frac{w-j}{h-j} q_{h-j} \geq (1-\delta)k|S|$, and the probability of $G \in \mathcal{M}(n, \mathbf{m}, k+1)$ containing some set S with $\gamma n \leq |S| \leq (1-\gamma)n$ and $\eta(S) < h^2 \delta k |S|$ is $o(1)$.

We also show the deterministic result that

(c) no multihypergraph G with property $\mathcal{A}(\gamma)$ contains any sets S with $|S| > (1-\gamma)n$ for which Lemma 4.6(i) holds.

It follows that the probability that G has property $\mathcal{A}(\gamma)$ and contains some set S for which all parts (i)–(iv) of Lemma 4.6 hold is $o(1)$. Then by Lemmas 4.5 and 4.6,

$$\mathbf{P}(G \in \mathcal{A}(\gamma) \wedge G \text{ is not } (w, k)\text{-orientable}) = o(1).$$

Finally, Lemma 3.1 shows that the result applies to random (simple) hypergraphs as well.

References

[1] N. Alon and J. Spencer. *The probabilistic method.* Wiley-Interscience Series in Discrete Mathematics and Optimization. John Wiley & Sons, Inc., Hoboken, NJ. xviii+352 pp., third edition, 2008.

[2] Y. Azar, A. Z. Broder, A. R. Karlin, and E. Upfal. Balanced allocations. *SIAM J. Comput.*, 29(1):180–200, 1999.

[3] P. Berenbrink, A. Czumaj, A. Steger, and B. Vőcking. Balanced allocations: The heavily loaded case. *32th Annual ACM Symposium on Theory of Computing*, pages 745–754, 2000.

[4] Béla Bollobás and Alan M. Frieze. On matchings and hamiltonian cycles in random graphs. In *Random graphs '83 (Poznan', 1983)*, pages 23–46. North-Holland Math. Stud., 118, North-Holland, Amsterdam, 1985.

[5] J. Cain, P. Sanders, and N. Wormald. The random graph threshold for k-orientiability and a fast algorithm for optimal multiple-choice allocation. In *Proc. ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 469–476, January 2007.

[6] J. Cain and N. Wormald. Encores on cores. *Electronic Journal of Combinatorics*, 13:RP 81, 2006.

[7] V. Chvátal. Almost all graphs with $1.44n$ edges are 3-colorable. *Random Structures Algorithms*, 2(1):11–28, 1991.

[8] William J. Cook, William H. Cunningham, William R. Pulleyblank, and Alexander Schrijver. *Combinatorial optimization. (English summary).* Wiley-Interscience Series in Discrete Mathematics and Optimization. A Wiley-Interscience Publication. John Wiley & Sons, Inc., New York, x+355 pp., 1998.

[9] M. Dietzfelbinger, A. Goerdt, M. Mitzenmacher, A. Montanari, R. Pagh and M. Rink. Tight Thresholds for Cuckoo Hashing via XORSAT. preprint available at `http://arxiv.org/abs/0912.0287`.

[10] M. Dietzfelbinger and F. Meyer auf der Heide. Simple, efficient shared memory simulations. In *5th ACM Symposium on Parallel Algorithms and Architectures, Velen, Germany, SIGACT and SIGARCH*, pages 110–119, June 30–July 2, 1993.

[11] A. Frieze and P. Melsted. Maximum Matchings in Random Bipartite Graphs and the Space Utilization of Cuckoo Hashtables. preprint available at `http://arxiv.org/abs/0910.5535`.

[12] N. Fountoulakis and K. Panagiotou. Sharp Load Thresholds for Cuckoo Hashing. preprint available at `http://arxiv.org/abs/0910.5147`.

[13] D. Fernholz and V. Ramachandran. The k-orientability thresholds for $G_{n,p}$. In *Proc. ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 459–468 January 2007.

[14] P. Gao and N. Wormald. Orientability thresholds for random hypergraphs. 2010, preprint.

[15] S. L. Hakimi. On the degrees of the vertices of a directed graph. *J. Franklin Inst.*, 279:290–308, 1965.

[16] R. M. Karp, M. Luby, and F. Meyer auf der Heide. Efficient pram simulation on a distributed memory machine. In *24th ACM Symp. on Theory of Computing*, pages 318–326, May 1992.

[17] B.D. McKay and N.C. Wormald. Asymptotic enumeration by degree sequence of graphs with degrees $o(n^{1/2})$. *Combinatorica*, 11(4):369–382, 1991.

[18] F. Meyer auf der Heide, C. Scheideler, and V. Stemann. Exploiting storage redundancy to speed up randomized shared memory simulations. *Theoret. Comput. Sci.*, 162(2):245–281, August 1996.

[19] M. Mitzenmacher, A. Richa, and R. Sitaraman. The power of two random choices: a survey of techniques and results. In *Handbook of randomized computing*, volume I, II, pages 255–312. Comb. Optim., 9, Kluwer Acad. Publ., Dordrecht, 2001.

[20] P. Sanders, S. Egner, and J. Korst. Fast concurrent access to parallel disks. *Algorithmica*, 35(1):21–55, 2003.

[21] Alexander Schrijver. *Combinatorial optimization. Polyhedra and efficiency*, volume B. Algorithms and Combinatorics, 24, Springer-Verlag, Berlin, 2003. pp. i–xxxiv.

[22] N.C. Wormald. Differential equations for random processes and random graphs. *The Annals of Applied Probability*, 5(4):1217–1235, November 1995.