Cores in random hypergraphs and boolean formulas

Michael Molloy*
Department of Computer Science
University of Toronto
Toronto, ON, Canada
and Microsoft Research
Redmond, WA, USA

August 7, 2003

Abstract

We describe a technique for determining the thresholds for the appearance of cores in random structures. We use it to determine (i) the threshold for the appearance of a k-core in a random r-uniform hypergraph for all $r, k \geq 2, r + k > 4$, and (ii) the threshold for the pure literal rule to find a satisfying assignment for a random instance of r-SAT, $r \geq 3$.

1 Introduction

In [22], Pittel, Spencer and Wormald determined that $c = 3.35...$ is the threshold for the appearance of a k-core, i.e. a non-empty maximal subgraph of minimum degree at least k, in the random graph $G_{n,p=c/n}$. This threshold had been sought after for over a decade, partly because it is a lower bound on the threshold for k-colourability (see [6, 10, 21, 3, 16]). The proof was a lengthy analysis, focussing on a system of differential equations which modelled a vertex-stripping process that determines whether any graph has a k-core. In the preliminary material, they provided a short and elegant heuristic argument which gave a good intuitive explanation for the location of the threshold. They remarked that this argument could easily be expanded to give a short proof that 3.35... is a lower bound on the threshold.

In this paper, we demonstrate a technique for determining the threshold for the appearance of k-cores and similar structures. It is essentially an extension of the aforementioned heuristic argument from [22]. The technique is considerably easier to apply than the technique used for the full proof in [22], and it provides much shorter proofs. It does not provide

*This work is supported by an NSERC Research Grant and a Sloan Research Fellowship. Part of the work was done while the author was a visitor at the Fields Institute.

quite as much information about the core as the technique from [22] does, but if one just wants the location of the threshold, then it can establish that location fairly quickly. It also provides the (almost sure) size of the core.

We illustrate this technique with two applications: (i) we determine the threshold for the appearance of a k-core in a random r-uniform hypergraph; (ii) we determine the threshold for the pure literal rule to succeed on a random instance of r-SAT.

We use $H^r_{n,p}$ to denote the random r-uniform hypergraph on n vertices where each of the $\binom{n}{r}$ potential edges is present with probability p, independently of all other potential edges. Note that when $p = c/n^{r-1}$, the expected number of edges is $(c/r!)n$, which is linear in n when c, r are both constant. We say that a property P holds *almost surely (a.s.)* for $H^r_{n,p}$ if the limit as $n \to \infty$ of the probability of P is equal to 1. A k-core is a subhypergraph where every vertex has degree at least k. If $k, r = 2$ then a k-core is simply a cycle; cycles are well-known to appear with probability bounded away from 0 for any $c > 0$ and with probability $1 - o(1)$ for any $c \geq 1$ (see, for example, [16]). For all other pairs $k, r \geq 2$, the threshold for the appearance of a 3-core is as follows:

Theorem 1 *Consider any $r, k \geq 2$ with r, k not both equal to 2. Define*

$$c^* = \min_{x>0} \frac{x(r-1)!}{\left(1 - e^{-x} \sum_{i=0}^{k-2} \frac{x^i}{i!}\right)^{r-1}}.$$

(a) For any $c < c^*$, $H^r_{n,p=c/n^{r-1}}$ a.s. has no non-empty k-core.

(b) For any $c > c^*$, $H^r_{n,p=c/n^{r-1}}$ a.s. has a k-core of size $\alpha(c)n + o(n)$, with

$$\alpha(c) = 1 - e^{-x} \sum_{i=0}^{k-1} \frac{x^i}{i!},$$

where x is the greatest solution to

$$c = \frac{x(r-1)!}{\left(1 - e^{-x} \sum_{i=0}^{k-2} \frac{x^i}{i!}\right)^{r-1}}.$$

The case $r = 2$ of this theorem is the main result of [22]. Majewski, Wormald, Havas and Czech[15] gave some initial analysis towards proving Theorem 1 for the case $k = 2$, but they did not provide a complete proof. The authors of [22] remarked that a complete proof could probably be obtained using a lengthy analysis along the lines of that in [22].

An instance of r-SAT is a boolean formula in conjunctive normal form, where all clauses have size exactly r. We use $F^r_{n,p}$ to denote the random instance of r-SAT on n variables where each of the $2^r\binom{n}{r}$ potential clauses is present with probability p, independently of all other potential clauses. Note that when $p = c/n^{r-1}$, the expected number of clauses is $(2^r c/r!)n$, which is linear for c, r constant. We say that a property P holds almost surely (a.s.) for $F^r_{n,p}$

if the limit as $n \to \infty$ of the probability of P is equal to 1. It is straightforward to show that $F^r_{n,p=c/n^{r-1}}$ is equivalent in a sense to the more commonly studied model $F^r_{n,M}$ where we pick exactly M random clauses on the n variables (see, eg, [1]), so long as $M = (2^r c/r!)n$; in particular, the main theorem of this paper also applies to the latter model.

The *pure literal rule* is a method to search for a satisfying solution of a boolean formula in conjunctive normal form. Each variable x in a boolean formula has two literals: $x, \overline{x}$, and we say that those literals are negations of each other. A literal is *pure* if it's negation does not appear in the formula. The pure literal rule method repeatedly sets a pure literal to be true, until there are no more pure literals. Note that the order in which the pure literals are chosen does not affect whether the procedure succeeds. It is well-known that the threshold for the pure literal rule to a.s. succeed on a random instance of 2-SAT is $c = 1/2$, which corresponds to $d = 1$ for $F^2_{n,M=dn}$ (see, eg. [1]). Here, we establish the threshold for all larger values of r.

Theorem 2 *Consider any $r \geq 3$. Define*

$$c^+ \;=\; \min_{y>0} \frac{y(r-1)!}{2^{r-1}\left(1-\mathrm{e}^{-y}\right)^{r-1}}.$$

(a) For any $c < c^+$, the pure literal rule a.s. succeeds on $F^r_{n,p=c/n^{r-1}}$.

(b) For any $c > c^+$, the pure literal rule a.s. fails on $F^r_{n,p=c/n^{r-1}}$ and leaves $\beta(c)n + o(n)$ unset variables, with

$$\beta(c) = \left(1-\mathrm{e}^{-y}\right)^2,$$

where y is the greatest solution to

$$c = \frac{y(r-1)!}{2^{r-1}\left(1-\mathrm{e}^{-y}\right)^{r-1}}.$$

Thus, for example, when $r = 3$ we have a threshold of $c^+ = 1.2277...$, which yields for $F^3_{n,M=dn}$ a threshold of $d^+ = 2^3c^+/3! = 1.636....$ The location of the threshold for this particular case was approximated to be between 1.63 and 1.7 in [9]. Later, [20] and [18] each provided a heuristic argument which predicts without formal proof that d^+ is indeed the threshold; the first is based on the solution of a set of differential equations, and the second is very similar to the heuristic argument from [22].

The technique that we use here was inspired heavily by the heuristic argument in [22] and also by work on a similar problem in [8]. This technique was used by the author and Goerdt[14] to provide the threshold for a k-core to appear in the graph obtained by randomly removing edges from a random regular graph. Recently, and independently of the work in this paper, Kim[17] reports a rather different technique for determining similar thresholds; his technique can also provide some more detailed information, such as the width of the scaling window for the threshold.

2 k-cores of hypergraphs

2.1 The key lemmas

Given a hypergraph H, we can determine whether H contains a k-core with the following simple procedure.

CORE:

Set $H_0 := H$ and $i := 0$.
While H_i has any vertices of degree less than k
  Set H_{i+1} to be the hypergraph remaining after deleting from H_i
    all vertices of degree less than k and all hyperedges
    containing such vertices.
  Increment $i := i + 1$.

It is straightforward to confirm that H contains a k-core iff CORE does not delete every vertex from H.

We will prove Theorem 1 by analyzing an application of CORE to $H^r_{n,p=c/n^{r-1}}$. To do so, we focus on what happens before and after iteration I for some particular large constant I. Recall the definitions of $c^*, \alpha(c)$ from the statement of Theorem 1.

Lemma 3 *For any $\epsilon, \delta > 0$, there exists $I = I(\epsilon, \delta)$ such that if $c \leq c^* - \epsilon$ then a.s. $|H_I| < \delta n$.*

Lemma 4 *For any $\epsilon, \delta > 0$, there exists $I = I(\epsilon, \delta)$ such that if $c \geq c^* + \epsilon$ then:*

(a) a.s. $\alpha(c)n < |H_I| < \alpha(c)n + \delta n$; and

(b) a.s. CORE will remove at most δn vertices after iteration I.

We also require the following standard lemma:

Lemma 5 *For any fixed $c > 0$, there exists $\delta = \delta(c)$ such that a.s. $H^r_{n,p=c/n^{r-1}}$ has no subgraph which has fewer than δn vertices and with average degree at least k.*

Lemma 4 implies that for $c > c^*$, a.s. $H^r_{n,M=cn}$ has a k-core of size $\alpha(c)n + o(n)$. Applying Lemma 3 with $\delta = \delta(c^* - \epsilon)$ from Lemma 5 implies that for $c < c^*$, a.s. the output of CORE is either empty or has average degree less than k; i.e. a.s. $H^r_{n,M=cn}$ has no k-core. This establishes Theorem 1.

2.2 Proofs

We begin with the proof of Lemma 5. The first lemma of this type was due to Luczak[19]; this is one of countless similar lemmas that followed, all using essentially the same proof.

Proof of Lemma 5 Consider any $r \le a < \delta n$, where δ is to be named later. If H has a subgraph on a vertices with average degree at least k, then that subgraph has at least $\frac{k}{r}a$ edges, and by deleting edges, we see that H has a subgraph on a vertices with exactly $\lceil \frac{k}{r}a \rceil$ edges. The expected number of such subgraphs in $H^r_{n,p=c/n^{r-1}}$ is at most

$$\begin{aligned}
\binom{n}{a}\binom{\binom{a}{r}}{\lceil \frac{k}{r}a \rceil} p^{\lceil \frac{k}{r}a \rceil} &\le \left(\frac{en}{a}\right)^a \left(\frac{e(\frac{a^r}{r!})}{\frac{k}{r}a}\right)^{\lceil \frac{k}{r}a \rceil} \left(\frac{c}{n^{r-1}}\right)^{\lceil \frac{k}{r}a \rceil} \\
&\le \left(\frac{en}{a}\right)^a \left(\frac{ec}{(r-1)!k} \times \frac{a^{r-1}}{n^{r-1}}\right)^{\lceil \frac{k}{r}a \rceil} \\
&\le \left(C\frac{a}{n}\right)^{(r-1)\lceil \frac{k}{r}a \rceil - a},
\end{aligned}$$

where $C = C(r,k,c)$. We set $\delta = \frac{1}{2C}$, and note that for any integers r, k both at least 2 and not both equal to 2, $(r-1)\lceil \frac{k}{r}a \rceil - a \le ((r-1)\frac{k}{r} - 1)a \le a/3$. Therefore, this expected value is at most $\left(\frac{1}{2}\right)^{a/3}$. Furthermore, it is also less than $O\left(n^{-1/3}\right)$. Therefore, the expected number of such subgraphs is at most

$$\sum_{a=1}^{n} \min(O(n^{-2/3}), \left(\frac{1}{2}\right)^{a/3}) = o(1).$$

Therefore, by Markov's inequality, there is a.s. no such subgraph. $\Box$

We now turn our attention to the proofs of Lemmas 3 and 4.

We say that a vertex *survives* after i rounds of CORE if $v \in H_i$. Consider any vertex v. We will focus on the probability that v survives after I rounds. We use $Z(x)$ to denote a Poisson variable with mean x. Consider the following recursive definition:

$$\begin{aligned}
\rho_0 &= 1 \\
\rho_i &= \mathbf{Pr}(Z(\rho_{i-1}^{r-1}c/(r-1)!) \ge k-1),
\end{aligned}$$

and define

$$\lambda_i = \mathbf{Pr}(Z(\rho_{i-1}^{r-1}c/(r-1)!) \ge k).$$

Lemma 6 *For any constant t, the probability that v survives after t rounds is $\lambda_t + o(1)$.*

Proof For each $i \ge 0$, let D_i denote the set of vertices of distance i from v. Let E_1 be the event that fewer than $\log^2 n$ vertices are within distance t of v, and none of these vertices lie in a cycle of length less than $2t+1$; if E_1 holds then $\cup_{0 \le i \le t} D_i$ induces a hypertree in H which is rooted at v and has size less than $\log^2 n$. For any $i \ge 0$ and $u \in D_i$ a *child* *edge* of u is a hyperedge of the form $(u, x_1, ..., x_{r-1})$ where each $x_j \in D_{i+1}$. Note that if E_1 holds and $u \neq v$, then exactly one hyperedge containing u is not a child edge of u.

We consider a different procedure, which we call **CORE2**: At each iteration $1 \leq i \leq t-1$, we only consider vertices in D_{t-i} and we delete a vertex if it has fewer than $k-1$ remaining child edges. Then, in iteration t, we delete v if it has fewer than k remaining edges. It is easy to verify that if E_1 holds then v remains after all t iterations of CORE2 iff v survives t iterations of CORE.

We show by induction on i, that for each $0 \leq i \leq t-1$, the probability that a particular vertex $u \in D_{t-i}$ survives CORE2 is $\rho_i + o(1)$. For $i = 0$, the assertion is trivial, since no vertices from D_t are ever removed. Assuming the statement is true for $i-1$, expose the vertices of $D_0, D_1, ..., D_{t-i}$ and consider some vertex $u \in D_{t-i}$. Suppose that $|D_0 \cup ... \cup D_{t-i}| = x$; if E_1 holds then $x \leq \log^2 n$. The number of child edges of u which remain after $i-1$ iterations of PURE2 has expectation $\mu = \binom{n-x}{r-1} p \rho_{i-1}^{r-1}$. If $x \leq \log^2 n$, then $\mu = \rho_{i-1}^{r-1} c/(r-1)! + o(1)$. It is straightforward to argue that the distribution of the number of such edges is asymptotic to a Poisson variable (one way to do this is to consider the ℓth moments of the variable for each integer ℓ). The probability that it is at least $k-1$ decreases with μ. Therefore that probability is at most $\mathbf{Pr}(Z(\rho_{i-1}^{r-1} c/(r-1)! + o(1)) \geq k-1) + o(1) = \rho_i + o(1)$ and at least $\mathbf{Pr}(x \leq \log^2 n) \times (\mathbf{Pr}(Z(\rho_{i-1}^{r-1} c/(r-1)! + o(1)) \geq k-1) + o(1)) = \rho_i + o(1)$.

v survives the final iteration of PURE2 iff it has at least k child edges that remain after $t-1$ iterations. The same analysis shows that this happens with probability $\lambda_i + o(1)$. Since $\mathbf{Pr}(E_1) = o(1)$, v survives t rounds of PURE with probability $\lambda_i + o(1)$. □

Lemma 7 *For any constant t, a.s. $||H_t| - \lambda_t n| < n^{2/3}$.*

Proof This can be proved with a straightforward concentration argument using, for example, Azuma's Inequality or an easy second moment calculation. We omit the details. □

Proof of Lemma 3 We consider $\rho = \lim_{t\to\infty} \rho_t$. Since ρ_t is easily seen to be positive and non-increasing, this limit exists. Furthermore, it must satisfy:

$$\rho = \mathbf{Pr}(Z(\rho^{r-1} c/(r-1)!) \geq k-1).$$

Suppose $\rho > 0$. Then substituting $x = \rho^{r-1} c/(r-1)!$ and rearranging, we obtain:

$$c = \frac{x(r-1)!}{(\mathbf{Pr}(Z(x) \geq k-1))^{r-1}} = \frac{x(r-1)!}{(1 - \mathrm{e}^{-x} \sum_{i=0}^{k-2} \frac{x^i}{i!})^{r-1}}.$$

But, $c < c^*$ and so there is no x for which this equation can hold. Therefore, $\rho = 0$, which implies $\lim_{t\to\infty} \lambda_t = 0$, and so we can choose a constant $I = I(\epsilon, \delta)$ such that $\lambda_I < \delta/2$. Thus by Lemma 7 a.s. $|H_I| < \delta n$. □

Proof of Lemma 4 Since $c > c^*$, the equation

$$c = \frac{x(r-1)!}{(1 - \mathrm{e}^{-x} \sum_{i=0}^{k-2} \frac{x^i}{i!})^{r-1}}$$

has a solution for x. In fact, it is easily seen to have 2 solutions; let x_1 be the largest one. It is straightforward to extend the arguments from the proof of Lemma 3 to verify that $\rho = \lim_{t\to\infty} \rho_t$ satisfies $x_1 = \rho^{r-1}c/(r-1)!$ and that $\lim_{t\to\infty} \lambda_t = \alpha(c) > 0$. Therefore, for I sufficiently large in terms of ϵ, δ, we have $\alpha(c) < \lambda_I < \alpha(c) + \delta$. This, along with Lemma 7, proves part (a).

To prove part (b), we begin by exposing the vertices of H_I and their degrees.

Claim 1: Every hypergraph with that vertex set and degree sequence is equally likely to be H_I.

Proof: Consider two hypergraphs H^a and H^b both with that vertex set and degree sequence. Consider any initial hypergraph H^A such that applying CORE to H^A yields $H_I = H^a$. Modify H^A by replacing the edges of H^a with the edges of H^b, and call the resulting hypergraph H^B. It is easy to see that applying CORE to H^B yields $H_I = H^b$. Furthermore, H^A and H^B have the same number of edges, and so the probability that our initial random hypergraph was equal to H^A is the same as the probability that it was equal to H^B. This implies that the probability that $H_I = H^a$ is equal to the probability that $H_I = H^b$, as required. $\square$

By Claim 1, it is valid for us to draw H_I from the configuration model (see [5, 7, 23, 12]). In this model, we take $\deg(v)$ copies of each vertex v, and then take a uniformly random partition of the set of all copies into sets of size r. Each set corresponds to a hyperedge containing the vertices of which the set contains copies. It is possible that this will form a multigraph, but the probability of forming a simple hypergraph is sufficiently high that results holding a.s. for the random configuration must also hold a.s. for H_I. (See [12] for the details.)

For each $d \geq 0$, let the number of vertices of degree d be $\gamma_d n$. For any vertex v, the probability that $v \in H_I$ and the degree of v in H_I is less than k is equal to the probability that v is in H_I but not H_{I+1}. By Lemma 7, a.s the number of such vertices is at most $(\lambda_I - \lambda_{I+1})n + o(n)$. Since $\lim_{i\to\infty} \lambda_i$ exists, we can make $\sum_{d=0}^{k-1} \gamma_d$ arbitrarily small by taking I sufficiently large. For any $d \geq k$, the same analysis as in the proofs of Lemmas 6 and 7 yields that a.s.

$$\gamma_d = \mathbf{Pr}(Z(\rho_{I-1}^{r-1}c/(r-1)!) = d) + o(1).$$

and by taking I sufficiently large, we can make this value arbitrarily close to

$$\gamma'_d = \mathrm{e}^{-x}\frac{x^d}{d!}.$$

We now consider yet another variation on our procedure, which we call CORE3. This can be thought of as a slower version of CORE. CORE3 deletes only 1 vertex per iteration, each time selecting a vertex of degree less than k uniformly at random from amongst all remaining such vertices. We consider switching from CORE to CORE3 after iteration I of CORE. Note that, since the order in which small degree vertices are deleted will not affect the actual core that remains, CORE3 will delete the same number of vertices that CORE deletes after iteration I. Thus, to prove part (b) of our lemma, it suffices to prove that a.s. CORE3 deletes at most δn vertices.

We let $X_0 = \sum_{d<k} \gamma_d n$, i.e. the number of vertices of degree less than k in H_I, and for each $j > 0$, we let X_j be the number of such vertices remaining after the jth iteration of CORE3..

During the first iteration of CORE3, we delete up to $k-1$ hyperedges. Expose the vertices in those hyperedges. Each has degree exactly k with probability $k\gamma_k / \sum_{d\geq 1} d\gamma_d + o(1)$ which, by taking I sufficiently large, can be made arbitrarily close to

$$\frac{k\gamma'_k}{\sum_{d\geq k} d\gamma'_d} = \frac{k\mathrm{e}^{-x}\frac{x^k}{k!}}{\sum_{d\geq k} d\mathrm{e}^{-x}\frac{x^d}{d!}} = \frac{\frac{x^{k-1}}{(k-1)!}}{\sum_{d\geq k-1}\frac{x^d}{d!}}. \tag{1}$$

.

Our next step is to bound this quantity. To do so, we first recall that at $c = c^*$, x minimizes

$$f(x) = x(r-1)!/\left(1 - \mathrm{e}^{-x}\sum_{i=0}^{k-2} x^i/i!\right)^{r-1}.$$

Setting the derivative of $f(x)$ to 0 yields:

$$\begin{aligned}
x\frac{\partial}{\partial x}\left(1 - \mathrm{e}^{-x}\sum_{i=0}^{k-2}\frac{x^i}{i!}\right)^{r-1} &= \left(1 - \mathrm{e}^{-x}\sum_{i=0}^{k-2}\frac{x^i}{i!}\right)^{r-1} \\
x(r-1)\left(1 - \mathrm{e}^{-x}\sum_{i=0}^{k-2}\frac{x^i}{i!}\right)^{r-2} \times \left(\mathrm{e}^{-x}\sum_{i=0}^{k-2}\frac{x^i}{i!} - \mathrm{e}^{-x}\sum_{i=1}^{k-2}\frac{x^{i-1}}{(i-1)!}\right) &= \left(1 - \mathrm{e}^{-x}\sum_{i=0}^{k-2}\frac{x^i}{i!}\right)^{r-1} \\
(r-1)(k-1)\mathrm{e}^{-x}\frac{x^{k-1}}{(k-1)!} &= 1 - \mathrm{e}^{-x}\sum_{i=0}^{k-2}\frac{x^i}{i!}
\end{aligned}$$

Therefore, for $c > c^*$, we have

$$(r-1)(k-1)\frac{x^{k-1}}{(k-1)!} < \mathrm{e}^x - \sum_{i=0}^{k-2}\frac{x^i}{i!} = \sum_{i\geq k-1}\frac{x^i}{i!},$$

which implies that expression (1) is less than $1/(r-1)(k-1)$. Therefore, there is some $\kappa = \kappa(x) > 0$ such that for a sufficiently large constant I, the expected number of degree k vertices whose degree decreases during the first iteration of CORE3 is at most $(1-\kappa)$.

Each such deletion alters the degrees of at most $(r-1)(k-1)$ vertices. It follows that there is some $\psi > 0$ such that for each $j < \psi n$, even if the adversary can deterministically choose the degrees of the $j(r-1)(k-1)$ vertices whose degrees are altered during the first j iterations, repeating the above calculations for the resulting degree sequence will yield that the expected number of degree k vertices whose degree decreases during iteration $j+1$ is at most $(1-\kappa/2)$. (The proof of Claim 1 easily extends to show that it is valid to use the configuration model to analyze the hypergraph remaining after the first j iterations.)

If the degree of a vertex y decreases by at least 2 during the deletion of a vertex x, then y must occur twice in the union of the hyperedges containing x; for a fixed x, the probability of this occuring for any vertex y is easily computed to be $O(1/n)$. Therefore $\mathbf{Exp}(X_1 - X_0) \leq -\kappa$ for some $\kappa > 0$, and furthermore, for each $j < \psi n$, if $X_j > 0$ then

$\mathbf{Exp}(X_{j+1} - X_j) \le -\kappa/2$. We pick I large enough that $\psi < \delta$ and $\lambda_I - \lambda_{I+1} < \psi\kappa/8$; thus a.s. $X_0 < (\psi\kappa/4)n$ by Lemma 7. As we explain below, standard results concerning random walks on the integers show that a.s. the sequence $X_0, X_1, X_2, \ldots$ will drift to $X_j = 0$ for some $j \le \psi n$. At this point, CORE3 will halt and we will have found a k-core, after removing at most $\psi n < \delta n$ vertices. This proves part (b).

To prove that $X_0, X_1, \ldots$ drifts to zero quickly, we couple it to a more conventional random sequence: $Y_0, Y_1, \ldots$. For each $-1 \le i \le (r-1)(k-1) - 1$, the probability that $X_1 = X_0 + i$ is equal to $o(1)$ plus the probability that precisely $i+1$ of the up to $(r-1)(k-1)$ neighbours of the vertex being deleted are of degree k. This latter probability is $o(1)$ plus

$$\zeta_i = \binom{(r-1)(k-1)}{i+1} \left(\frac{k\gamma_k}{\sum_{d\ge1} d\gamma_d} \right)^{i+1} \left(1 - \frac{k\gamma_k}{\sum_{d\ge1} d\gamma_d} \right)^{(r-1)(k-1)-i-1}.$$

From our calculations above, it follows that $\sum i\zeta_i \le -\kappa$. Furthermore, by the same reasoning as at the beginning of the previous paragraph, for each $0 \le i \le (r-1)(k-1) - 1$ there exists ζ_i' such that (a) for every $j < \psi n$, if $X_j > 0$ then the probability that $X_{j+1} = X_j + i$ is less than ζ_i', and (b) setting $\zeta_{-1}' = 1 - \sum_{i=0}^{(r-1)(k-1)-1} \zeta_i'$, we have $\sum i\zeta_i' \le -\kappa/2$.

Now define $Y_0 = X_0$ and for each $-1 \le i \le (r-1)(k-1)-1$, $Y_{j+1} = Y_j + i$ with probability ζ_i'. We can couple the sequences so that for each $j < \psi n$, if $X_j > 0$ and $X_{j+1} - X_j \ne -1$ then $Y_{j+1} - Y_j = X_{j+1} - X_j$. Thus, for each $j \le \psi n$, $X_j \le \max\{Y_j, 0\}$. It is easy to show that a.s. $Y_{\psi n} \le 0$; indeed, $\mathbf{Exp}(Y_{\psi n}) \le Y_0 - (\kappa/2)\psi n \le (\psi\kappa/4)n$ and since any one choice affects $Y_{\psi n}$ by at most $(r-1)(k-1)$, Azuma's Inequality implies that a.s. $Y_{\psi n} \le (\psi\kappa/8)n$. Therefore, a.s. $X_{\psi n} = 0$ as required. □

3 The pure literal rule

It is easy to see that the order in which we set pure literals has no effect on whether the pure literal rule succeeds, nor on which variables will remain unset. Thus, to prove Theorem 2, it suffices to consider the following procedure:

PURE:

Set $F_0 := F$ and $i := 0$.
While F_i has any pure literals
    Set all the pure literals of F_i to True, and remove all clauses in which they appear.
    Let F_{i+1} be the remaining formula.
    Increment $i := i + 1$.

Proof of Theorem 2 The proof follows the same structure as that of Theorem 1, so we only outline the key places where they differ.

The analog of Lemma 5 can be found in [2] amongst other places. Specifically, for every $c > 0$ there is a $\delta = \delta(c)$ such that a.s. every subformula of $F^r_{n,p=c/n^{r-1}}$ with fewer than δn

variables has a variable of degree at most 1 and hence a pure literal. The proof is a simple variation of the proof of Lemma 5.

We say that a variable *survives* i rounds of PURE if it is not set during those rounds.

Given our r-SAT formula F, the *underlying hypergraph* H of F has the variables of F as vertices, and its hyperedges are those r-sets of variables which lie in clauses of F. Consider any particular variable v, and set D_i to be the set of variables who are at distance i from v in H. Fix any constant t. With high probability, the hypergraph induced by $\cup_{i=0}^{t} D_i$ is a hypertree, and for each non-leaf variable u of degree d in the tree, the $d-1$ clauses which correspond to hyperedges belonging to the subtree rooted at u are called the *child clauses* of u; if $u \neq v$ then the other clause u lies in is called the *parent clause* of u.

We consider the procedure PURE2: at each iteration $1 \leq i \leq t-1$, we only consider variables in D_{t-i}. For each such variable u, we remove u and all of its clauses if the negation of the literal of u in its parent clause does not appear in any of its child clauses; i.e. if u is pure. Then, in iteration t, we delete v if at least one of v's literals does not appear in any of the remaining child clauses of v. We say that a variable *survives* i rounds of PURE if it is not set during those rounds. It is easy to see that v survives t rounds of PURE iff v is not removed by PURE2.

We recursively define:

$$\begin{aligned}
\alpha_0 &= 1 \\
\alpha_i &= \mathbf{Pr}(Z(\alpha_{i-1}^{r-1} 2^{r-1} c/(r-1)!) \geq 1) \\
\beta_i &= \mathbf{Pr}(Z(\alpha_{i-1}^{r-1} 2^{r-1} c/(r-1)!) \geq 1)^2
\end{aligned}$$

If follows as in the proof of Lemma 6 that a variable $u \in D_{t-i}, i < t$ survives PURE2 with probability $\alpha_i + o(1)$ and that v survives PURE2, and hence PURE, with probability β_t. It follows as in the proof of Lemma 7 that a.s. the number of variables that survive t rounds of PURE is $\beta_t n + o(n)$.

We define $\alpha = \lim_{i \to \infty} \alpha_i$, and so $\alpha = \mathbf{Pr}(Z(\alpha^{r-1} 2^{r-1} c/(r-1)!) \geq 1)$. Setting $y = \alpha^{r-1} 2^{r-1} c/(r-1)!$ and rearranging, we obtain:

$$c = \frac{y(r-1)!}{2^{r-1}\left(1 - \mathrm{e}^{-y}\right)^{r-1}}$$

This implies, as in the proof of Lemma 3, that if $c < c^+$ then for any constant $\delta > 0$ there is a sufficiently large constant I such that we will a.s. have fewer than δn unset variables after I iterations. This in turn, by the aforementioned analog of Lemma 5, implies that there is no pure core, i.e. the pure literal rule a.s. succeeds. This proves part (a).

To prove part (b), we consider the *literal degree sequence* of the formula remaining after I iterations of PURE. That is, the sequence formed by writing down the degree of each literal and then erasing any record of which degree corresponds to which literal. The only notable consequence of this erasure is that there is no record as to which numbers are paired to form the degrees of a literal and its negation.

Claim: Every formula with that literal degree sequence is equally likely to be the one remaining after I iterations of PURE.

The proof is nearly identical to that of the corresponding claim for Lemma 4.

Thus, we can expose the literal degree sequence of the formula and then assume that it is generated as follows: (1) Take a random matching amongst the degrees to obtain a pair of degrees for each variable. (2) Generate the random formula using the CNF analog of the configuration model. That is, for each variable x, take $\deg(x)$ copies of x and $\deg(\overline{x})$ copies of $\overline{x}$; then partition the set of literal-copies into sets of r; those sets of r form the 3-clauses in the obvious way.

For $c > c^+$, we set $\beta = \lim_{i\to\infty} \beta_i = \alpha^2 > 0$. Note that $\beta = \beta(c)$. After I steps, a.s. we will have $\beta_I n + o(n)$ unset variables where β_I can be made arbitrarily close to β by taking I sufficiently large. For each $d \geq 0$, let $\gamma_d n$ be the number of literals of degree i. Since each variable has 2 literals, a.s. $\gamma_0 \leq 2(\beta_I - \beta_{I+1}) + o(1)$ which can be made arbitrarily small by taking I sufficiently large. Similarly, for each $d > 0$, γ_d can be made arbitrarily close to

$$\gamma'_d = \mathbf{Pr}(Z(y) = d) \times \mathbf{Pr}(Z(y) \geq 1) = \frac{\mathrm{e}^{-y} y^d}{d!}(1 - \mathrm{e}^{-y}).$$

(Roughly, the product is the product of the probabilities that a literal x has degree d and that its negation has degree at least 1.)

We now consider CORE3, the slower version of CORE. At each iteration, CORE3 selects a single pure literal uniformly at random from amongst all such literals and sets it. We consider switching from CORE to CORE3 after iteration I of CORE.

For each $j \geq 0$, we let X_j denote the number of pure literals remaining after j iterations of CORE3. We say that a literal is *near-pure* if its negation appears in exactly one clause. Thus, the number of near-pure literals is $\gamma_1 n$. To set a pure literal x, we first expose its degree d, where by our Claim every degree in the literal degree sequence is equally likely, and then we expose the $d(r-1)$ literals that lie in clauses with x. Setting x causes a near-pure literal y to become pure iff its negation is one of those $d(r-1)$ literals. For each of those $d(r-1)$ literals, the probability that it has degree i is roughly $i\gamma_i / \sum_{d\geq 0} d\gamma_d$. Therefore, the expected number of near-pure literals which become pure when x is set is at most:

$$\frac{\sum_{d\geq 0} d(r-1)\gamma_d n}{\sum_{d\geq 0} \gamma_d n} \times \frac{\gamma_1 n}{\sum_{d\geq 0} d\gamma_d n} + o(1).$$

By taking I sufficiently large, this can be made arbitrarily close to

$$\frac{\sum_{d\geq 1} d(r-1)\gamma'_d n}{\sum_{d\geq 1} \gamma'_d n} \times \frac{\gamma'_1 n}{\sum_{d\geq 1} d\gamma'_d n} = \frac{y\mathrm{e}^{-y}(r-1)}{1 - \mathrm{e}^{-y}}. \tag{2}$$

At $c = c^+$, we have $\frac{\partial}{\partial y} y/(1 - \mathrm{e}^{-y})^{r-1} = 0$ and so

$$\begin{aligned}(1 - \mathrm{e}^{-y})^{r-1} &= y(r-1)(1 - \mathrm{e}^{-y})^{r-2}\mathrm{e}^{-y} \\ 1 - \mathrm{e}^{-y} &= y(r-1)\mathrm{e}^{-y}.\end{aligned}$$

By differentiating both sides, we see that for any $c > c^+$, we have $1 - \mathrm{e}^{-y} > y(r-1)\mathrm{e}^{-y}$. Thus the expression in (2) is less than 1. It is straightforward to show that the expected number of pure literals formed that were not near-pure at the beginning of the iteration is $o(1)$. It now follows as in the proof of Lemma 4 that by choosing I sufficiently large we have

$$\mathbf{Exp}(X_1 - X_0) \leq -\kappa,$$

for some $\kappa = \kappa(c) > 0$, and furthermore that a.s. we will reach $X_j = 0$ very quickly and end up with a pure-core with approximately $\beta n + o(n)$ variables. $\square$

References

[1] D. Achlioptas. *Lower bounds for random 3-SAT via differential equations.* Theoretical Computer Science, 265 (1-2), (2001), p.159-185.

[2] D. Achlioptas, P. Beame and M. Molloy, *A sharp threshold in proof complexity.* Proceedings of STOC 2001, 337 - 346.

[3] D. Achlioptas and M. Molloy. *Analysis of a list-colouring algorithm on a random graph.* Proceedings of the 38th Annual Symposium on Foundations of Computer Science (FOCS), 204 - 212 (1997).

[4] K. Azuma. *Weighted Sums of Certain Dependent Random Variables.* Tokuku Math. Journal **19** (1967), 357 - 367.

[5] E. A. Bender and E. R. Canfield, *The asymptotic number of labelled graphs with given degree sequences.* Journal of Combinatorial Theory (A) **24** (1978), 296-307.

[6] B. Bollobás. *The evolution of sparse graphs.* In “Graph Theory and Combinatorics”, editor: B. Bollobás, 35 - 57 (1984).

[7] B. Bollobás, Martingales, Isoperimetric Inequalities and Random Graphs. Colloq. Math. Soc. Janós Bolyai 52 (1987), 113 - 139.

[8] B. Bollobás and G. Brightwell. *The width of random graph orders.* The Mathematical Scientist **20**, 69 - 90 (1995).

[9] A. Broder, A. Frieze and E. Upfal. *On the satisfiability and maximum satisfiability of random 3-CNF formulas.* Proceedings of the 41st Annual ACM-SIAM Symposium on Discrete Algorithms (SODA), 322-330 (1993).

[10] V. Chvátal. *Almost All Graphs With 1.44n Edges are 3-Colourable.* Random Structures and Algorithms **2** (1991), 11 - 28.

[11] V. Chvatal and E. Szemeredi. *Many hard examples for resolution.* Journal of the ACM **35** (1988) 759 - 768.

[12] C. Cooper, A. Frieze, M. Molloy and B. Reed. *Perfect matchings in random r-regular, s-uniform hypergraphs.* Combinatorics, Probability and Computing **5** 1-14 (1996).

[13] P. Erdős and A. Rényi. *On the Evolution of Random Graphs.* Magayr Tud. Akad. Mat. Kutato Int. Kozl. **5** (1960), 17 - 61.

[14] A. Goerdt and M. Molloy. *Analysis of edge deletion processes on random regular graphs* Proceedings of Latin American Theoretical Informatics, 38-47 (2000). To appear in Theoretical Computer Science.

[15] B. Majewski, N. Wormald, G. Havas, Z. Czech. *A Family of Perfect Hashing Methods.* The Computer Journal **39(6)**, 547-554 (1996).

[16] S. Janson, T. Łuczak and A. Ruciński. Random Graphs. Wiley, New York (2000).

[17] J. Kim, in preparation.

[18] M. Luby, M. Mitzenmacher, and A. Shokrollahi. *Analysis of random processes via and-or tree evaluation.* Proceedings of SODA 1998.

[19] T. Łuczak. *Size and connectivity of the k-core of a random graph.* Discrete Math. **91**, 61 - 68 (1991).

[20] M. Mitzenmacher. *Tight thresholds for the pure literal rule.* Technical Note 1997-011, Digital Systems Research Center, Palo Alto, CA (1997).

[21] M. Molloy. *A gap between the appearance of a k-core and a k-chromatic graph.* Random Structures & Algorithms **8**, 159 - 160 (1996).

[22] B. Pittel, J. Spencer and N. Wormald. *Sudden emergence of a giant k-core in a random graph.* J. Comb. Th. B **67**, 111 - 151 (1996).

[23] N. C. Wormald. *Some Problems in the Enumeration of Labelled Graphs.* Doctoral thesis, Newcastle University (1978).