Encores on cores

Julie Cain

Department of Mathematics and Statistics
The University of Melbourne
VIC 3010, Australia

J.Cain@ms.unimelb.edu.au

Nicholas Wormald*

Department of Combinatorics and Optimization
University of Waterloo
Waterloo ON, Canada

nwormald@uwaterloo.ca

Submitted: May 12, 2005; Accepted: Sep 1, 2006; Published: Sep 22, 2006
Mathematics Subject Classification: 05C80

Abstract

We give a new derivation of the threshold of appearance of the k-core of a random graph. Our method uses a hybrid model obtained from a simple model of random graphs based on random functions, and the pairing or configuration model for random graphs with given degree sequence. Our approach also gives a simple derivation of properties of the degree sequence of the k-core of a random graph, in particular its relation to multinomial and hence independent Poisson variables. The method is also applied to d-uniform hypergraphs.

1 Introduction

We introduce a model of random graphs which provides advantages when analysing algorithms that recursively delete vertices of low degree. It is obtained by combining a model of random graphs via pseudographs (graphs with loops and multiple edges) with the standard model of random graphs with given degrees. We use it to re-derive the threshold of emergence of the k-core (defined below) of a random graph, which was first obtained

*Research supported by the Australian Research Council, the Canada Research Chairs Program and NSERC. Research partly carried out while the author was in the Department of Mathematics and Statistics, University of Melbourne.

in [13]. It also enables us easily to give sharp estimates on the distribution of degrees in the k-core of a random graph. Our method will also have advantages in forthcoming work.

An approach similar to the one of this paper was also used by Aronson et al. [1] to find maximum matchings in random graphs. Study of the k-core appears to be important, and different approaches have, simultaneously with the present work, been developed by Kim [8] (which is not very unlike our approach, but uses a different model), Molloy [11] and Cooper [5] for graphs with given degree sequences.[1] One feature emphasised by our approach is the very strong relation between the degrees of vertices in the k-core, and the multinomial distribution (Corollary 1).

The *k-core* of a graph or hypergraph is the largest subgraph of minimum degree at least k. Łuczak [9] proved that for every fixed $k \geq 3$, the k-core of the random graph $\mathcal{G}(n, m)$ a.a.s. (asymptotically almost surely, as $n \to \infty$) either is empty or has at least $0.0002n$ vertices (regardless of m). Work after that on the k-core of $\mathcal{G}(n, m)$, such as in [13], essentially focussed on the following *vertex deletion algorithm*. Throughout this paper, for given k we call a vertex *light* if its degree is less than k, and *heavy* otherwise. The algorithm repeatedly deletes light vertices until no light vertices remain. This algorithm always terminates with the k-core of the graph, which is possibly empty. Note that we may equivalently use an *edge deletion algorithm*: repeatedly delete edges incident with light vertices; then the final graph consists of the k-core of G together with some isolated vertices.

For $k \geq 0$ integer and λ a positive real, define

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

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

$$c_k = \inf\{h_k(\mu) : \mu > 0\}. \tag{1}$$

Take $k \geq 3$. Then c_k is a positive real because $h_k(\mu)$ tends to ∞ if μ tends to 0 or ∞. It is easily checked that for $c > c_k$ the equation $h_k(\mu) = c$ has two positive roots (and just one for $c = c_k$). Define $\mu_{k,c}$ to be the larger one. Also define $c_2 = 1$. The following was proved for $k \geq 3$ in Pittel et al. [13], although in that paper the result about edges requires some searching through the proof. Sharper error terms are also obtained in [13], and also an "explosion" result that we do not obtain here. The case $k = 2$ was also not explicitly stated in [13], but the proof also works for $c > c_2$, and in this case the result on number of vertices was first found by Pittel [12].

Theorem 1 *Let $c > 0$ and integer $k \geq 2$ be fixed. Suppose that $m \sim cn/2$, and $G \in \mathcal{G}(n, m)$. For $c < c_k$ and $k \geq 3$, G has empty k-core a.a.s. For $c > c_k$, the k-core of G a.a.s. has $e^{-\mu_{k,c}} f_k(\mu_{k,c}) n(1 + o(1))$ vertices and $\frac{1}{2}\mu_{k,c} e^{-\mu_{k,c}} f_{k-1}(\mu_{k,c}) n(1 + o(1))$ edges.*

[1] Since submission of this article, two more approaches emerged: Janson and Łuczak [7], and Riordan [15].

Here, and for similar statements, we say a function $h(G) = o(1)$ a.a.s. if there is a property H that is a.a.s. true, such that $\max_{G \in H} h(G) = o(1)$ (as $n \to \infty$).

We give a simpler proof of Theorem 1 in Section 4. Partly, our proof is simpler because we aim for a less sharp result. But the use of the new model in the present case provides a distinct simplification that with a bit more effort would (we claim) give the full results of [13]. It avoids the use of some asymptotic enumeration formulae and some other arguments.

We obtain as an offshoot of our method several results concerning the k-core of a random graph in $\mathcal{G}(n,m)$, and in particular its degree sequence. The k-cores with this distribution form a probability space of graphs, naturally with all vertex degrees at least k, which we denote by $\mathcal{K}(n,m,k)$. The set of vertices occurring is a random subset of $[n]$ of random cardinality $\hat{n}$ (where $[x]$ denotes $\{1, 2, \ldots, x\}$ throughout this paper). When considering the degree sequence of a graph in such a space, we simply take the degrees of the vertices in order of increasing label (i.e. compress the $\hat{n}$ vertices to the interval $[\hat{n}]$ without disturbing their order).

Let $\mathbf{Multi}(n,s)$ be the probability space of nonnegative integer vectors $(X_1, \ldots, X_n)$ summing to s such that for any vector $(d_1, \ldots, d_n)$ of the same type,

$$\mathbf{P}(X_i = d_i \text{ for } 1 \leq i \leq n) = \frac{s!}{n^s \prod_{i=1}^n d_i!}.$$

We will be interested in particular in heavy vertices, and accordingly define

$$H_{n,s,k} := \{(h_1, \ldots, h_n) : \sum h_i = s \text{ and } h_i \geq k \text{ for all } i\}, \tag{2}$$

and let $\mathbf{Multi}(n,s)|_{\geq k}$ be the probability space obtained by restricting $\mathbf{Multi}(n,s)$ to elements of $H_{n,s,k}$. We will find the k-core G of a certain random pseudograph, and set $I = 1$ if G is simple, to obtain the pairs (G, I) in the following theorem. In particular, G will have minimum degree at least k.

Theorem 2 *Fix $k \geq 1$. For all positive n and $m \geq kn/2$, there is a probability space consisting of ordered pairs (G, I) where G is a random pseudograph and $I \in \{0, 1\}$, with, marginally, $\mathbf{P}(I = 1) = \Omega(1)$ if $m = O(n)$, such that*

(i) the distribution of G conditional upon $I = 1$ is identical to that of the random core $\mathcal{K}(n,m,k)$ (so in particular, G is simple if $I = 1$);

(ii) in the marginal distribution of G conditioned on $|V(G)| = \hat{n}$ and $|E(G)| = \hat{m}$, the degree sequence of G is distributed precisely as $\mathbf{Multi}(\hat{n}, 2\hat{m})|_{\geq k}$.

This theorem, proved in Section 3, has consequences when combined with distributional results on the numbers of vertices and edges in the random k-core, such as the main theorem in [13].

Corollary 1 *Let $k \geq 2$ and $c > c_k$ be fixed, $m \sim cn/2$, and consider the the probability space defined by the random vector distributed as the degree sequence of $\mathcal{K}(n,m,k)$. Let H_n be any event in this space, and write $\mathbf{P}_{\mathbf{Multi}(\hat{n},2\hat{m})|_{\geq k}}(H_n)$ for the probability that a random vector in $\mathbf{Multi}(\hat{n},2\hat{m})|_{\geq k}$ lies in H_n. If $\hat{n} \sim e^{-\mu_{k,c}} f_k(\mu_{k,c})n$ and $2\hat{m} \sim \mu_{k,c}e^{-\mu_{k,c}} f_{k-1}(\mu_{k,c})n$ imply that*

$$\mathbf{P}_{\mathbf{Multi}(\hat{n},2\hat{m})|_{\geq k}}(H_n) < P_n$$

then $\mathbf{P}(H_n) = O(P_n) + o(1)$.

Proof. This is immediate from Theorems 1 and 2, with the $o(1)$ error appearing because the conclusion of Theorem 1 includes the "a.a.s." qualification. ∎

One can state sharper but more complicated results immediately by restricting the range of $\hat{n}$ and $\hat{m}$ to the narrower range that was shown in [13] to contain them a.a.s., and using the sharper bounds on probabilities proved there. We consider in detail the number of vertices of a given degree. For $k \geq 1$ denote by $Z(k,\lambda)$ a random variable which has a k-truncated Poisson distribution with a parameter λ, that is

$$\mathbf{P}\big(Z(k,\lambda) = j\big) = \begin{cases} \dfrac{\lambda^j}{j! f_k(\lambda)}, & j \geq k \\ 0, & j < k \end{cases} \tag{3}$$

where f is as defined as above. Let λ_b denote the positive root of the equation

$$\mathbf{E}\, Z(k,\lambda) = \frac{\lambda f_{k-1}(\lambda)}{f_k(\lambda)} = b. \tag{4}$$

It is easily seen that λ_b exists provided $b > k$.

Statements similar to Theorem 2 and Corollary 1 hold for n independent copies of $Z(k,\lambda)$ for appropriate λ, in place of $\mathbf{Multi}(n,s)|_{\geq k}$, and $\Omega(1)$ replaced by $\Omega(1/\sqrt{n})$, due to a connection between independent Poisson variables and multinomials. Exploiting this connection, we will obtain the following result in Section 2, assuming the truth of Theorems 1 and 2. (Direct computations with multinomials can give even more precise results.)

Corollary 2 *Fix $j \geq k \geq 2$, and let $m = m(n) \sim cn/2$ where $c > c_k$ is fixed. Let $\hat{N}$ and $\hat{M}$ be the (random) numbers of vertices and edges of $\mathcal{K}(n,m,k)$, and let Y_j be the number of vertices having degree j. Then for sufficiently small $\epsilon > 0$, conditional upon $\hat{M} - k\hat{N} > \epsilon n$ and $\hat{N} > \epsilon n$, we have*

$$\mathbf{P}\left(|Y_j - np_{2\hat{M}/\hat{N},k,j}| > a\sqrt{\hat{N}}\right) = O(\sqrt{n})e^{-2a^2}$$

for all $a > 0$, where, with λ_b as in (4),

$$p_{b,k,j} = \frac{\lambda_b^j}{j! f_k(\lambda_b)}. \tag{5}$$

This corollary combined with Theorem 1 immediately implies distributional results on the number of vertices of given degree in a large core of a random graph. A (weak) example is the following.

Corollary 3 *Fix $j \geq k \geq 2$, and let $m = m(n) \sim cn/2$ where $c > c_k$. The number of vertices of degree j in $\mathcal{K}(n, m, k)$ is a.a.s. $np_{\tilde{c},k,j} + o(n)$, where $\tilde{c} = \mu_{k,c} f_{k-1}(\mu_{k,c})/f_k(\mu_{k,c})$.*

Note Similar results can also be obtained fairly easily from the results in [14] giving the relation between independent truncated Poisson variables and enumeration of graphs with minimum degree at least k. Another approach, mentioned in [13], would be to model the random k-core using the random pseudograph model described in the first paragraph of Section 3 conditioned on minimum degree at least k, and again use asymptotic enumeration computations to eliminate loops and multiple edges. However, these results would require some work to obtain a result as sharp as our Corollary 1. Fountoulakis [6] also obtained similar results by following the argument of the main results in [13], keeping track of degree information.

We finish by pointing out that our results generalise easily to hypergraphs. For $d \geq 3$ let $\mathcal{G}(d, n, m)$ be the uniform probability space of d-uniform n-vertex m-edge simple hypergraphs and let $\mathcal{K}(d, n, m, k)$ be the probability space of the k-cores of $\mathcal{G}(d, n, m)$. For $k \geq 2$, let $h_{d,k}(\mu) = \frac{\mu}{(e^{-\mu} f_{k-1}(\mu))^{d-1}}$ and define

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

As for $h_k(\mu)$, $h_{d,k}(\mu)$ tends to ∞ if μ tends to 0 or ∞, so $c_{d,k}$ is a positive real (and this applies even when $k = 2$). Define $\mu_{d,k,c}$ to be the larger solution of $h_{d,k}(\mu) = c$ for $c > c_{d,k}$. The result on number of vertices in the following theorem was also independently derived in [5], [11] and first stated for $k = 2$ by Majewski et al. [10] with an application to constructing minimal perfect hash functions.

Theorem 3 *Let $c > 0$ and integers $d \geq 3$, $k \geq 2$ be fixed. Suppose that $m \sim cn/d$, and $G \in \mathcal{G}(d, n, m)$. For $c < c_{d,k}$, G has empty k-core a.a.s. For $c > c_{d,k}$, the k-core of G a.a.s. has $e^{-\mu_{d,k,c}} f_k(\mu_{d,k,c}) n(1+o(1))$ vertices and $\frac{1}{d}\mu_{d,k,c} e^{-\mu_{d,k,c}} f_{k-1}(\mu_{d,k,c}) n(1+o(1))$ hyperedges. Moreover, let $j \geq k$ be fixed, and assume $c > c_{d,k}$. Then the number of vertices of degree j in $\mathcal{K}(d, n, m, k)$ is a.a.s. $np_{\tilde{c},k,j} + o(n)$, where $\tilde{c} = \mu_{d,k,c} f_{k-1}(\mu_{d,k,c})/f_k(\mu_{d,k,c})$.*

The proof is in Section 4.

We note that by incorporating some of the techniques in [13, Section 6], we can obtain much sharper estimates on the numbers of vertices and edges in the k-core than in Theorems 1 and 3 — with errors of size $O(n^{1/2+\epsilon})$ — but we do not pursue this; Kim [8] claims a slightly sharper result.

2 Multinomial and Poisson

Lemma 1 *Let $k \geq 0$ be fixed, and $s - nk = \Omega(n)$. For $(X_1, \ldots, X_n) \in \mathbf{Multi}(n, s)|_{\geq k}$ put $Y_j = |\{i : X_i = j\}|$. Then for all $a > 0$*

$$\mathbf{P}\big(|Y_j - np_{c,k,j}| > a\sqrt{n}\big) = O(\sqrt{n})e^{-2a^2}.$$

and
$$\mathbf{P}(X_n = j) \sim p_{c,k,j}, \tag{6}$$
where $c = s/n$ and $p_{c,k,j}$ is defined in (5).

Proof. Let $\lambda = \lambda_c$ in the following. Let $Z_1, Z_2, \ldots$ be independent copies of $Z(k, \lambda)$, and write $\mathbf{Z}^{(k,t)}$ for $(Z_1, \ldots, Z_t)$. Then
$$\mathbf{P}(Z_i = d_i \text{ for } i = 1, \ldots, n) = \prod_{i=1}^{n} \frac{\lambda^{d_i}}{f_k(\lambda) d_i!} = \frac{\lambda^s}{f_k(\lambda)^n} \prod_{i=1}^{n} \frac{1}{d_i!}.$$
It follows that the restriction of $\mathbf{Z}^{(k,n)}$ to $H_{n,s,k}$ (defined in (2)) has the distribution of $\mathbf{Multi}(n, s)|_{\geq k}$. We have $\mathbf{P}(Z_i = j) = p_{c,k,j}$, and so the variable Y_j defined for $\mathbf{Z}^{(k,n)}$ has the distribution of $\mathrm{Bin}(n, p_{c,k,j})$. So by Chernoff's bound, for $\mathbf{Z}^{(k,n)}$ we have $\mathbf{P}(|Y_j - np_{c,k,j}| > a\sqrt{n}) < 2e^{-2a^2}$. The first statement now follows because $\mathbf{P}(\mathbf{Z}^{(k,n)} \in H_{n,s,k}) = \Omega(1/\sqrt{n})$; see [14, Theorem 4(a)] for a short proof. The first claim of the lemma follows.

To obtain (6) from the first part of the lemma, first put $a = \log n$ for example, showing $Y_j \sim np_{c,k,j}$ with probability at least $1 - o(1/n)$. Thus $\mathbf{E}Y_j \sim np_{c,k,j}$. By linearity of expectation, $\mathbf{E}Y_j = \sum_i \mathbf{P}(X_i = j) = n\mathbf{P}(X_n = j)$ by symmetry, and this yields (6). ∎

Proof of Corollary 2 For m as in the statement of the corollary, we have by Theorem 1 that the k-core is a.a.s. nonempty and $\hat{N}$ and $\hat{M}$ are large: a.a.s. at least ϵn for sufficiently small $\epsilon > 0$. For the claim about edges, this relies on the fact that $\mu_{k,c} f_{k-1}(\mu_{k,c})/f_k(\mu_{k,c}) > k$, which is clear because $\mathbf{E}Z(k, \mu) > k$, c.f. (4). Hence, the event conditioned on in the corollary holds a.a.s. The corollary now follows from Theorem 2 and the first part of Lemma 1. ∎

3 The pairing-allocation model and proof of Theorem 2

The model of random pseudographs used by Bollobás and Frieze [3] and Chvátal [4] can be described as follows. Given positive integers n and m, define the probability space $\mathcal{A}(n, m)$ of functions $a : [2m] \to [n]$, all functions equiprobable. We call such functions *allocations*. A random pseudograph (i.e. graph with loops and multiple edges permitted), $G(a)$, on vertex set $[n]$ is obtained from an allocation $a \in \mathcal{A}(n, m)$ by including an edge between vertices $a(2i - 1)$ and $a(2i)$ for $1 \leq i \leq m$. Clearly $G(a)$ has precisely m edges. Let $\mathcal{M}(n, m)$ denote the probability space of pseudographs obtained in this way.

Not all pseudographs are equiprobable in $\mathcal{M}(n, m)$. However, by considering all the permutations of $[2m]$, we see that there are $m!2^m$ allocations which give rise to each simple graph, which yields the following lemma immediately. Here $N = \binom{n}{2}$.

Lemma 2 (Chvátal [4]) *For $a \in \mathcal{M}(n, m)$, all simple n-vertex, m-edge graphs are equiprobable as $G(a)$. Furthermore, for $c = 2m/n$,*
$$\mathbf{P}(G(a) \textit{ is simple}) = \frac{\binom{N}{m} m! 2^m}{n^{2m}} = \exp(-c/2 - c^2/4) + o(1). \quad ∎$$

We next define another model that is a sort of hybrid between the allocation model and the standard model for random graphs with given degree sequence introduced by Bollobás (see [2]). The parameters of this model are k, m, ℓ, and $t \geq 0$, such that

$$2m \geq kt + \ell. \tag{7}$$

The model contains a set M of *points* with even cardinality $2m$ arranged into m disjoint unordered pairs which we technically call *edges of* M, and disjoint sets L and V of cardinality ℓ and t respectively. The *pairing-allocation* model $\mathcal{P}(M, L, V, k)$ is the uniform probability space of all functions $h : M \to L \cup V$ (called pairing-allocations) such that $|h^{-1}(v)| = 1$ if $v \in L$, and $|h^{-1}(v)| \geq k$ if $v \in V$. Provided (7) holds, such functions exist.

Associated with each pairing-allocation h, there is a pseudograph $G(h)$ with vertex set V, and with an edge joining $h(a)$ and $h(b)$ for each edge $\{a, b\}$ of M with $h(a), h(b) \in V$. We define

$$s = 2m - \ell,$$

the total number of points allocated to vertices in V. From (7), $s \geq kt$.

The connection between the pairing-allocation model and $\mathcal{M}(n, m)$ is given next. First, given a pseudograph G, let $H(G)$ denote the sub-pseudograph induced by its heavy vertices.

Lemma 3 *Fix k, m, n, t and s with $s \geq kt$. Let $V \subseteq [n]$ with $|V| = t$, and take any M and L disjoint from $[n]$ with $|M| = 2m$ and $|L| = 2m - s$. The conditional distribution of $H(G(a))$ for $a \in \mathcal{M}(n, m)$, given the set V of heavy vertices of $G(a)$ and the total degree s of those vertices, is identical to the distribution of $G(h)$ for $h \in \mathcal{P}(M, L, V, k)$.*

Note here that if $2m - s > (k - 1)(n - t)$ the event conditioned on cannot occur, in which case the lemma is intended to say nothing.

Proof. The choice of a random allocation a conditional upon V and s can be done in three steps. First choose which s elements of $[2m]$ are mapped by a to V. Then allocate those to V such that each $v \in V$ receives at least k of them. Then allocate the remaining $2m - s$ elements to $[n] \setminus V$ such that none receives more than $k - 1$ elements. Each step is done u.a.r. (uniformly at random), and $H(G(a))$ is determined by the first two steps.

Without loss of generality, take $M = [2m]$, and $\{\{2i - 1, 2i\} : i \in [m]\}$ for the edges of M. The choice of $h \in \mathcal{P}(M, L, V, k)$ can then be done in three steps, with the first two steps identical to the ones for a. (Thirdly, allocate the remaining $2m - s$ elements bijectively to L.) Since $G(h)$ is also determined by the first two steps, the lemma follows. ∎

Since the k-core of $G(a)$ must equal the k-core of $H(G(a))$, Lemmas 2 and 3 tell us that we may concentrate on the (random) k-core of $G(h)$ for $h \in \mathcal{P}(M, L, V, k)$, in a way to be made precise shortly. The second part of the proof of Lemma 3 also gives the following, where $(d_{G(h)}(v))_{v \in V}$ denotes the degree sequence of $G(h)$.

Lemma 4 *For $h \in \mathcal{P}(M, L, V, k)$ the random vector $(d_{G(h)}(v))_{v \in V}$ is distributed precisely as a vector in* $\mathbf{Multi}(t, s)|_{\geq k}$. ∎

For any $h \in \mathcal{P}(M, L, V, k)$ with $M = [2m]$, and $j \in M$, define $g(j)$ to be the other point in the same edge as t, so g is an involution. We will see that the following algorithm is equivalent to the edge deletion algorithm given in Section 1. In each iteration of the repeat loop, which we call a *step* of the algorithm, this procedure effectively deletes an edge $\{j, g(j)\}$ from M that cannot be contained in the k-core of $G(h)$. The u.a.r. choice of a light vertex in the procedure ensures that its choice does not depend on h.

Light(k):
```
repeat
   choose j ∈ h^{-1}(L) u.a.r.
   delete j and g(j) from M and delete h(j) from L
   set v := h(g(j))
   if v ∈ L, then delete v from L
   if v ∈ V and in addition |h^{-1}(v)| = k, then delete v from V,
         insert k − 1 new elements into L, and redefine the action
         of h on h^{-1}(v) \ {g(j)} as a bijection to the new elements
   if v ∈ V and |h^{-1}(v)| > k, do nothing
until h^{-1}(L) = ∅.
```

Let h_i, M_i, L_i, and V_i be the values of h, M, L, and V after i steps of **Light**(k). Also let i_{fin} denote the number of steps before **Light**(k) terminates. This is clearly finite.

Lemma 5 *$G(h_{i_{\text{fin}}})$ is the k-core of the initial pseudograph $G(h_0)$.*

Proof. No edge of M with an endpoint mapped by h_i to L_i can be mapped to an edge of the k-core of $G(h_i)$. When the other end of such an edge is mapped to $v \in V$ with $h^{-1}(v) = k$, it is impossible for v to be in the k-core; that is, the k-core of $G(h_i)$ is equal to the k-core of $G(h_i) - v$. It is then clear that the k-core of $G(h_{i+1})$ is the same as the k-core of $G(h_i)$. The lemma follows by induction on i, since $G(h_{i_{\text{fin}}})$ is its own k-core. ∎

Lemma 6 *Beginning with $h \in \mathcal{P}(M, L, B, k)$, conditional upon the values of M_i, L_i and V_i, the pairing-allocation h_i is distributed precisely as in $\mathcal{P}(M_i, L_i, V_i, k)$.*

Proof. This is immediate for $i = 0$. For $i \geq 1$ it is enough to show the statement holds conditional upon the random sets M_i, V_i, L_i, M_{i-1}, V_{i-1} and L_{i-1}. By the inductive hypothesis, conditioning on the latter three alone, h_{i-1} is distributed u.a.r. as in $\mathcal{P}(M_{i-1}, L_{i-1}, V_{i-1}, k)$. Further conditioning on M_i, V_i, L_i is equivalent to specifying whether h_{i-1} maps the edge $\{j, g(j)\}$ of M to two elements of L (in which case, which pair is mapped to which pair is also determined), and, if not, specifying whether $h_{i-1}(g(j))$ is heavy of degree k and becomes light, in which case the values of $h_{i-1}(g(j))$ and its inverse image under h_{i-1} are also determined. It is easy to see that, given this information, h_i is distributed u.a.r. subject to the conditions required for $\mathcal{P}(M_i, L_i, V_i, k)$, as required. ∎

Proof of Theorem 2 For $a \in \mathcal{M}(n, m)$, let G denote the k-core of $G(a)$ and set the indicator variable $I = 1$ iff $G(a)$ is simple. This defines a distribution on the pairs (G, I).

Lemma 2 implies that $\mathbf{P}(I=1)=\Omega(1)$ if $m=O(n)$, and also gives part (i) of the theorem. From Lemmas 3 and 5, the distribution of G, given the set V of heavy vertices of $G(a)$ and the total degree s of those vertices, is identical to the distribution of $G(h_{i_{\mathrm{fin}}})$ for $h_0 \in \mathcal{P}(M,L,V,k)$ (with suitable definition of L and M). Conditioning further on the numbers of vertices $\hat{n}$ and edges $\hat{m}$ of G, Lemmas 6 and 4 imply that its vertex degrees are distributed as $\mathbf{Multi}(\hat{n},2\hat{m})|_{\geq k}$. As this distribution is independent of V and s, the same result holds without conditioning on V and s. This is part (ii) of the theorem. ∎

4 Proof of Theorem 1

Let $y_i(G)$ denote the number of vertices of degree i in a graph G and write

$$t(G)=\sum_{i\geq k} y_i(G), \qquad s(G)=\sum_{i\geq k} i y_i(G). \tag{8}$$

Let a be a random allocation in $\mathcal{A}(n,m)$. We will prove the claims about $G(a)$; the theorem then follows for $G\in\mathcal{G}(n,m)$ by Lemma 2.

First, note that by the second moment method, it is easily seen that the number of vertices of $G(a)$ of degree i (i fixed) is a.a.s.

$$ne^{-c}c^i/i!+o(n)$$

and hence the sum of degrees of vertices of degree less than k is a.a.s. asymptotic to

$$ne^{-c}\sum_{i=1}^{k-1} c^i/(i-1)! = cn(1-e^{-c}f_{k-1}(c))$$

with f as in Section 1. It follows that a.a.s. $t(G(a))=\hat{t}_c n+o(n)$ and $s(G)=\hat{s}_c n+o(n)$ where

$$\hat{t}_c=e^{-c}f_k(c), \quad \hat{s}_c=ce^{-c}f_{k-1}(c).$$

So it suffices to restrict consideration to those a with $t(G(a))=t_0$ and $s(G(a))=s_0$, for some

$$t_0\sim \hat{t}_c n, \quad s_0\sim \hat{s}_c n. \tag{9}$$

By Lemma 3, it suffices to consider $G(h)$ for $h\in\mathcal{P}(M,L,V,k)$, with $|V|=t_0$ and $|L|=2m-s_0$. By Lemma 5, we just have to show that when $\mathbf{Light}(k)$ is applied to such h, for $c<c_k$, $G(h_{\mathrm{fin}})$ is a.a.s. empty, whilst for $c>c_k$, $G(h_{\mathrm{fin}})$ a.a.s. has the number of vertices and edges claimed of the k-core of G in Theorem 1.

We will study the sequence of random values of $(T_i,S_i)=\big(t(G(h_i)),s(G(h_i))\big)$ for $i=0,\ldots,i_{\mathrm{fin}}$. Initially, $(T_0,S_0)=(t_0,s_0)$. On the ith step of the algorithm, the probability that $v:=h_{i-1}(g(j))\in V$ (which, in itself, decreases S by 1 in this step) is by Lemma 6

$$\frac{S_{i-1}}{2m-2i+1} \tag{10}$$

because the cardinality of the domain of h_{i-1} is $2m-2i+2$. Furthermore, conditional upon this event, the probability that v has degree k (which causes S to decrease by a further $k-1$ and T by 1) is k/S_{i-1} times the number of vertices in V of degree k, and hence by Lemmas 6 and 4 and (6) is asymptotic to

$$\frac{k}{S_{i-1}} \times \frac{T_{i-1}\lambda_{i-1}^k}{f_k(\lambda_{i-1})k!} = 1 - \frac{\lambda_{i-1}T_{i-1}}{S_{i-1}}(1+o(1))$$

where $\lambda_{i-1} = \lambda_{S_{i-1}/T_{i-1}}$ as defined in (4). Here, we used $\lambda^{k-1}/(k-1)! = f_{k-1} - f_k$ and (4); we also assume that

$$S_{i-1} - T_{i-1}k \to \infty, \tag{11}$$

which also immediately implies $2m-2i \to \infty$, and we also assume that $|h_{i-1}^{-1}(L)| > 0$, i.e.

$$2m - 2i + 2 - S_{i-1} > 0. \tag{12}$$

It follows, assuming (11), that if $|h_{i-1}^{-1}(L)| > 0$,

$$\begin{aligned}
\mathbf{E}(T_i - T_{i-1} \mid M_{i-1}, L_{i-1}, B_{i-1}, h_{i-1}) &\sim -\frac{S_{i-1}}{2m-2i}\left(1 - \frac{\lambda_{i-1}T_{i-1}}{S_{i-1}}\right), \\
\mathbf{E}(S_i - S_{i-1} \mid M_{i-1}, L_{i-1}, B_{i-1}, h_{i-1}) &\sim -\frac{S_{i-1}}{2m-2i}\left(1 + (k-1)\left(1 - \frac{\lambda_{i-1}T_{i-1}}{S_{i-1}}\right)\right).
\end{aligned}$$

We may now apply [16, Theorem 1] to show that the trajectory of S and T throughout the algorithm is a.a.s. close to the solution of the deterministic differential equations suggested by these equations. For $\epsilon > 0$, define the domain $D = D(\epsilon)$ to be

$$\{(x,y,z) : y - kz > \epsilon,\ c - 2x - y > 0,\ z > -\epsilon/2k, -\epsilon < x < c, |y| < c + \epsilon\}. \tag{13}$$

(Here x, y and z represent approximate values of i, S_i and T_i, all divided by n.) The first and second constraints in D arise from (11) and (12) respectively. The third is included so that, combined with the first, we have $y > \epsilon/2$ in D and hence $c - 2x > \epsilon/2$. Inclusion of the last two constraints is just a convenient way to make the domain bounded. The conclusion of [16, Theorem 1] now gives that a.a.s.

$$S_i = ny(i/n) + o(n) \quad \text{and} \quad T_i = nz(i/n) + o(n) \tag{14}$$

uniformly for all $i \le n\sigma$, where

$$z' = -\frac{y}{c-2x}\left(1 - \frac{\mu z}{y}\right), \tag{15}$$

$$y' = -\frac{y}{c-2x}\left(k - (k-1)\frac{\mu z}{y}\right) = (k-1)z' - 1, \tag{16}$$

μ denotes $\lambda_{y/z}$ and σ is the supremum of those x for which the solution of these differential equations, with initial conditions $y(0) = s_0/n$, $z(0) = t_0/n$, remains in D.

To analyse σ, we need to determine which constraint is violated when the solution reaches the boundary of D. It cannot be any of the last three constraints in (13), because (14) must give asymptotically feasible values of S_i and T_i up until the boundary is approached. This eliminates the third and fourth constraints immediately, and the fifth is similar because the maximum possible value of S_i is asymptotically cn. It remains to determine which of the first two constraints is violated when $x = \sigma$.

To proceed, we will assume that $t_0 = \hat{t}_c n$ and $s_0 = \hat{s}_c n$. The result then implies the general case, where these are asymptotically correct given by (9), due to standard stability properties of differential equations with respect to initial conditions.

We want to find integrals of the equations above. It is perhaps easier to change the variable of differentiation from x to $\tilde{x}$ where $\frac{d\tilde{x}}{dx} = y/(c-2x)$. (Incidentally, $\tilde{x}$ represents the number of times that $v \in V$). Then (15) and (16) become

$$\begin{aligned} z' &= -\left(1 - \frac{\mu z}{y}\right), && (17) \\ y' &= (k-1)z' - 1. && (18) \end{aligned}$$

From (4) we have

$$\frac{y}{z} = \mu r \quad \text{where} \quad r = \frac{f_{k-1}(\mu)}{f_k(\mu)}, \tag{19}$$

and we deduce that $\mu(0) = c$. This also gives

$$\left(\frac{y}{z}\right)' = (\mu r)' = \mu'(r + \mu r') = \mu'\big((r-1)(k-1-\mu r) + r\big)$$

after some algebra involving the definition of r and f_k. On the other hand,

$$y\left(\frac{y}{z}\right)' = \frac{y}{z}y' - \frac{y^2}{z^2}z' = \mu r\big((k-1-\mu r)z' - 1\big) = \mu\big((k-1-\mu r)(1-r) - r\big)$$

using (19), (17) and (18). Putting these two equations together, $y\mu' = \mu$ and hence, returning to differentiation with respect to x,

$$\frac{d\mu}{dx} = \frac{-\mu}{c-2x}. \tag{20}$$

From this it follows that the derivative of $\mu^2/(c-2x)$ is 0, i.e. this is a constant, and similarly that $ze^{\mu}/f_k(\mu)$ is constant (c.f. [13, (5.7) and (5.8)]).

Substituting in the initial conditions determines the constants and gives

$$\mu^2 = c(c-2x), \tag{21}$$

$$z = e^{-\mu} f_k(\mu), \tag{22}$$

and using (19),

$$y = \mu e^{-\mu} f_{k-1}(\mu). \tag{23}$$

In the closure of D, $c-2x$ and hence μ are positive. So by (21) and (23) and recalling that $h_k(\mu) = \mu e^\mu / f_{k-1}(\mu)$, we have

$$c - 2x - y = 0 \text{ iff } c = h_k(\mu). \tag{24}$$

So first consider $c < c_k$, where $k \geq 3$. Then by the definition of c_k in (1), there is no solution for μ in (24). So the boundary reached is determined by $y - kz = \epsilon$. Setting $y - kz = \epsilon$ using (22) and (23) forces $\mu = O(\epsilon)$, and so by (22) $z = O(\epsilon)$. The result (14) then implies that for this value of i, the number of vertices of degree at least k is $O(\epsilon n)$. Hence the k-core of $G(f)$ is a.a.s. of size at most $O(\epsilon n)$. Lemma 2 and Łuczak's result stated in the introduction then imply that the k-core of $\mathcal{G}(n,m)$ is a.a.s. empty.

Now take $c > c_k$, where $k \geq 3$. Then by (24), the second constraint is violated when $\mu = \mu_{k,c}$ as defined after (1) (noting that $\mu(0) = c$ and there is no solution to $h_k(\mu) = c$ for $\mu > c$). By (21) this occurs before μ approaches 0, so for sufficiently small ϵ the solution leaves (13) at the boundary $c - 2x - y = 0$. At this point, $c - 2x - y$ goes negative, since there are two distinct solutions to $h_k(\mu) = c$ as mentioned after (1). So we apply instead [17, Theorem 6.1] with $\hat{D}$ for that theorem defined as the domain (13), and the domain D replaced by $\tilde{D}$, which is the same as D except that the constraint $c - 2x - y > 0$ is omitted. Then this theorem gives the convergence in (14) up until the solution leaves $\tilde{D}$ or (12) is violated, whichever occurs first. Since $c - 2x - y$ begins to go negative, from (14) it follows that (12) must be violated a.a.s., and $h^{-1}(L)$ becomes zero at some $i \sim \sigma_c n$ a.a.s. where σ_c is the value of σ with the initial conditions $t_0 = \hat{t}_c n$ and $s_0 = \hat{s}_c n$. Thus the k-core is nonempty a.a.s., and here

$$z = e^{-\mu_{k,c}} f_k(\mu_{k,c}),$$

and

$$y = \mu_{k,c} e^{-\mu_{k,c}} f_{k-1}(\mu_{k,c})$$

as required. Finally, the case $k = 2$ and $c > 1$ is similar. ∎

Proof of Theorem 3 This is almost identical to the proof of Theorem 1 and Corollary 3, so we just sketch the proof, pointing out differences.

We use a modification of $\mathcal{P}(M, L, V, k)$ in which $|M| = dm$ and M is partitioned into sets of cardinality d to define the edges in the corresponding hypergraph. We also use the straightforward generalization of $\mathcal{M}(n,m)$ to a probability space of d-uniform pseudo-hypergraphs. We use a modification of **Light**(k) which treats each of the other points in the same hyperedge as j independently like v in the graph version (a total of $(d-1)$ times). Using $\frac{d\tilde{x}}{dx} = (d-1)y/(c-dx)$, we (of course) again obtain (17) and (18). The formula for $d\mu/dx$ in (20) is multiplied by $(d-1)(c-2x)/(c-dx)$. The equations are just as easily solved to find

$$\mu^d = c(c - dx)^{d-1},$$

and (22) and (23) are the same for z and y. Using (13) for D but with the second constraint changed to $c - dx - y > 0$, and repeating the differential equation argument in the proof of Theorem 1 gives the conclusion on the size of the random core. The modifications required for the proof of the other conclusion, on the number of vertices of given degree, are obvious. ∎

References

[1] J. Aronson, A. Frieze and B.G. Pittel, Maximum matchings in sparse random graphs: Karp-Sipser revisited, *Random Structures and Algorithms* **12** (1998), 111–177.

[2] B. Bollobás, *Random graphs.* Academic, New York, 1985.

[3] B. Bollobás and A. Frieze, On matchings and Hamiltonian cycles in random graphs. In *Random Graphs '83,* M. Karoński and A. Ruciński (Eds.), North-Holland, Amsterdam, pp. 23–46 (1985).

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

[5] C. Cooper, The cores of random hypergraphs with a given degree sequence, *Random Structures and Algorithms* **25** (2004), 353-375.

[6] N. Fountoulakis, *Thresholds and the Structure of Sparse Random Graphs*, D. Phil. Thesis, University of Oxford, 2003.

[7] S. Janson and M. Łuczak, A simple solution to the k-core problem, manuscript.

[8] J.H. Kim, Poisson cloning model for random graph, manuscript.

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

[10] B.S. Majewski, N.C. Wormald, G. Havas and Z.J. Czech, A Family of Perfect Hashing Methods, *Computer Journal* **39** (1996), 547–554.

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

[12] B. Pittel, On trees census and the giant component component in a sparse random graph, *Random Structures and Algorithms* **1** (1990), 311–342.

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

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

[15] O. Riordan, The k-core and branching processes, manuscript.

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

[17] N.C. Wormald, The differential equation method for random graph processes and greedy algorithms. In *Lectures on Approximation and Randomized Algorithms*, M. Karonski and H. J. Promel (eds), pp 73–155. PWN, Warsaw, 1999.

[18] N.C. Wormald, Models of random regular graphs. In J.D. Lamb and D.A. Preece (eds.), *Surveys in Combinatorics 1999*, *LMS Lecture Note Series* **267**, pp. 239–298. Cambridge University Press (1999).

Corrigendum – submitted December 3, 2006

In Corollary 2, $np_{2\hat{M}/\hat{N},k,j}$ should be $\hat{N}p_{2\hat{M}/\hat{N},k,j}$. As a consequence of this adjustment, in Corollary 3 and Theorem 3, $np_{\tilde{c},k,j}$ should be $ne^{-\mu}\mu^j/j!$, where $\mu = \mu_{k,c}$ in Corollary 3, and $\mu = \mu_{d,k,c}$ in Theorem 3. The proofs remain unchanged.

The corrected results thus read as follows.

Corollary 2 *Fix $j \geq k \geq 2$, and let $m = m(n) \sim cn/2$ where $c > c_k$ is fixed. Let $\hat{N}$ and $\hat{M}$ be the (random) numbers of vertices and edges of $\mathcal{K}(n,m,k)$, and let Y_j be the number of vertices having degree j. Then for sufficiently small $\epsilon > 0$, conditional upon $\hat{M} - k\hat{N} > \epsilon n$ and $\hat{N} > \epsilon n$, we have*

$$\mathbf{P}\left(|Y_j - \hat{N}p_{2\hat{M}/\hat{N},k,j}| > a\sqrt{\hat{N}}\right) = O(\sqrt{n})e^{-2a^2}$$

for all $a > 0$, where, with λ_b as in (4),

$$p_{b,k,j} = \frac{\lambda_b^j}{j!f_k(\lambda_b)}. \tag{5}$$

Corollary 3 *Fix $j \geq k \geq 2$, and let $m = m(n) \sim cn/2$ where $c > c_k$. The number of vertices of degree j in $\mathcal{K}(n,m,k)$ is a.a.s. $ne^{-\mu}\mu^j/j! + o(n)$, where $\mu = \mu_{k,c}$.*

Theorem 3 *Let $c > 0$ and integers $d \geq 3$, $k \geq 2$ be fixed. Suppose that $m \sim cn/d$, and $G \in \mathcal{G}(d,n,m)$. For $c < c_{d,k}$, G has empty k-core a.a.s. For $c > c_{d,k}$, the k-core of G a.a.s. has $e^{-\mu_{d,k,c}}f_k(\mu_{d,k,c})n(1+o(1))$ vertices and $\frac{1}{d}\mu_{d,k,c}e^{-\mu_{d,k,c}}f_{k-1}(\mu_{d,k,c})n(1+o(1))$ hyperedges. Moreover, let $j \geq k$ be fixed, and assume $c > c_{d,k}$. Then the number of vertices of degree j in $\mathcal{K}(d,n,m,k)$ is a.a.s. $ne^{-\mu}\mu^j/j! + o(n)$, where $\mu = \mu_{d,k,c}$.*