A SIMPLE SOLUTION TO THE k-CORE PROBLEM

SVANTE JANSON AND MALWINA J. LUCZAK

Abstract. We study the k-core of a random (multi)graph on n vertices with a given degree sequence. We let $n \to \infty$. Then, under some regularity conditions on the degree sequences, we give conditions on the asymptotic shape of the degree sequence that imply that with high probability the k-core is empty, and other conditions that imply that with high probability the k-core is non-empty and the sizes of its vertex and edge sets satisfy a law of large numbers; under suitable assumptions these are the only two possibilities. In particular, we recover the result by Pittel, Spencer and Wormald [19] on the existence and size of a k-core in $G(n,p)$ and $G(n,m)$, see also Molloy [17] and Cooper [3].

Our method is based on the properties of empirical distributions of independent random variables, and leads to simple proofs.

1. Introduction

Let $k \geq 2$ be a fixed integer. The k-core of a graph G is the largest induced subgraph of G with minimum vertex degree at least k. (Note that the k-core may be empty.) The question whether a non-empty k-core exists in a random graph has attracted a lot of attention over the past fifteen years. There have by now been quite a number of studies for the Bernoulli random graph $G(n,p)$ with n vertices and edge probability p, and for the uniformly random graph $G(n,m)$ with n vertices and m edges (see [3, 15, 17, 19] and references therein). Recently, Fernholz and Ramachandran [7, 8] have considered the k-core of a random graph with a specified degree sequence. More generally, Cooper [3] studies cores of random uniform hypergraphs with a given degree sequence. Yet more generally, Molloy [17] considers cores in random structures such as the uniform hypergraph and satisfiability of boolean formulas (see also references therein).

For a constant $\mu > 0$, let $\mathrm{Po}(\mu)$ denote a Poisson random variable with mean μ. Given $\mu > 0$ and $j \in \mathbb{Z}^+$, let $\psi_j(\mu) := \mathbb{P}\big(\mathrm{Po}(\mu) \geq j\big)$. Also, let $\lambda_k := \min_{\mu>0} \mu/\psi_{k-1}(\mu)$; and for $\lambda > \lambda_k$, we use $\mu_k(\lambda) > 0$ to denote the largest solution to $\mu/\psi_{k-1}(\mu) = \lambda$.

In [19], Pittel, Spencer and Wormald discovered that for $k \geq 3$, $\lambda = \lambda_k$ is the threshold for the appearance of a nonempty k-core in the graph

Date: 24 August 2005; revised 10 October 2006.

2000 *Mathematics Subject Classification.* 05C80.

Key words and phrases. cores, random graphs, balls and bins, death process, empirical distributions, law of large numbers.

$G(n, \lambda/n)$ (or, equivalently, $m = n\lambda_k/2$ is the threshold in the graph $G(n, m)$). Their strategy was to analyse an edge deletion algorithm that finds the k-core in a graph, showing that the corresponding random process is well approximated by the solution to a system of differential equations. The proof is rather long and complicated, and involves counting formulae for the number of graphs with a given degree sequence. For an analysis that uses a slightly modified version of their deletion algorithm and differs in some other important technical details too, see [14].

Fernholz and Ramachandran [7, 8] use different techniques to study the existence of a large k-core in a random graph with a given degree sequence. Their core-finding algorithm is basically identical to ours, but they analyse it in quite a different way; they also compare their result to a corresponding result for branching processes.

Cooper [3] has studied the k-core of a uniform multihypergraph with a given degree sequence. His method involves analysing a constructive algorithm generating the multihypergraph and its core, and inductively applying Azuma's inequality over time periods of length $n^{2/3}\Delta^{4/3}\log n$, where Δ is the initial maximum degree.

Molloy [17] gave another proof of the sharp threshold for the k-core, analysing a multi-round vertex and edge deletion algorithm via a branching process type argument.

Kim [13] considers cores in a "Poisson cloning" model of a random graph, which is somewhat different from $G(n, p)$. The slides [13] present a sketch argument, without precise error bounds, showing that the critical threshold for the emergence of a k-core agrees with the threshold in $G(n, p)$.

Darling and Norris [4] analyse cores in a different, weighted, Poisson model of a random hypergraph. Their method involves establishing a differential equation approximation for the Markov chain representing a suitable deletion algorithm. The threshold for $G(n, p)$ follows as a corollary to their main result.

Also see Cain and Wormald [2], who use differential equations to analyse the k-core threshold and the properties of the degree sequence of the giant k-core in a different model of a random graph. They make corresponding statements for $G(n, m)$ as a corollary.

In this paper, we present a simple solution to the k-core problem. Unlike [4] and [19], we do not use differential equations, but rely solely on the convergence of empirical distributions of independent random variables. Apart from $G(n, p)$ and $G(n, m)$, we are also able to handle the uniformly random graph with a given degree sequence under some regularity conditions similar to [3, 7, 8]. In contrast to [4, 19] we do not require counting formulae for graphs but, like [3] and [7, 8], work directly in the configuration model used to construct the random graph, exposing the edges one by one as they are needed.

We shall now state the result concerning the emergence of the k-core in the random graphs $G(n, p)$ and $G(n, m)$. Given a graph G, let $v(G)$ and

$e(G)$ denote the sizes of the vertex and edge sets of G respectively. We consider asymptotics as $n \to \infty$, and say that an event holds **whp** (*with high probability*), if it holds with probability tending to 1 as $n \to \infty$.

We shall use O_{p} and o_{p} in the standard way (see e.g. Janson, Łuczak and Ruciński [11]); for example, if (X_n) is a sequence of random variables, then $X_n = O_{\mathrm{p}}(1)$ means "X_n is bounded in probability" and $X_n = o_{\mathrm{p}}(1)$ means that $X_n \xrightarrow{\mathrm{p}} 0$.

Theorem 1.1 (Pittel, Spencer and Wormald [19]). *Consider the random graph $G(n, \lambda/n)$, where $\lambda > 0$ is fixed. Let $k \geq 2$ be fixed and let* $\mathrm{Core}_k = \mathrm{Core}_k(n, \lambda)$ *be the k-core of $G(n, \lambda/n)$.*

(i) *If $\lambda < \lambda_k$ and $k \geq 3$, then* Core_k *is empty* ***whp***.

(ii) *If $\lambda > \lambda_k$, then* ***whp*** Core_k *is non-empty, and* $v(\mathrm{Core}_k)/n \xrightarrow{\mathrm{p}} \psi_k(\mu_k(\lambda))$, $e(\mathrm{Core}_k)/n \xrightarrow{\mathrm{p}} \mu_k(\lambda)\psi_{k-1}(\mu_k(\lambda))/2 = \mu_k(\lambda)^2/(2\lambda)$.

The same results hold for the random graph $G(n,m)$, for any sequence $m = m(n)$ with $2m/n \to \lambda$.

Part (i) does not hold for $k = 2$. Here $\lambda_2 = 1$ and for $0 < \lambda < 1$ there is a positive limiting probability that there are cycles (as shown already by Erdős and Rényi [6]), and thus a non-empty 2-core. Nevertheless, in this case $e(\mathrm{Core}_k) = O_{\mathrm{p}}(1)$ and $v(\mathrm{Core}_k) = O_{\mathrm{p}}(1)$, so the core is small; cf. Theorem 2.3(i) below.

Acknowledgements. This research was mainly done during a visit by MJL to Uppsala University in April 2005, sponsored by the LSE Nordic Exchange Scheme.

2. Multigraphs

It will be convenient to work with *multigraphs*, that is to allow multiple edges and loops. In particular, we shall use the following type of random multigraph.

Let $n \in \mathbb{N}$ and let $(d_i)_1^n = (d_i^{(n)})_1^n$ be a sequence of non-negative integers such that $\sum_{i=1}^n d_i$ is even. We define a *random multigraph with given degree sequence* $(d_i)_1^n$, denoted by $G^*(n, (d_i)_1^n)$, by the configuration model (see e.g. [1]): take a set of d_i half-edges for each vertex i, and combine the half-edges into pairs by a uniformly random matching of the set of all half-edges. Note that $G^*(n, (d_i)_1^n)$ does not have exactly the uniform distribution over all multigraphs with the given degree sequence; there is a weight with a factor $1/j!$ for every edge of multiplicity j, and a factor $1/2$ for every loop, see [10, §1]. However, conditioned on the multigraph being a (simple) graph, we obtain a uniformly distributed random graph with the given degree sequence, which we denote by $G(n, (d_i)_1^n)$.

Remark 2.1. The distribution of $G^*(n, (d_i)_1^n)$ is the same as the one obtained by sampling the edges as ordered pairs of vertices uniformly with replacement, and then conditioning on the vertex degrees being correct.

Let us write $2m := \sum_{i=1}^n d_i$, so that $m = m(n)$ is the number of edges in the multigraph $G^*(n,(d_i)_1^n)$. We will let $n \to \infty$, and assume that we are given $(d_i)_1^n$ satisfying the following regularity conditions, cf. Molloy and Reed [18].

Condition 2.2. *For each n, $(d_i)_1^n = (d_i^{(n)})_1^n$ is a sequence of non-negative integers such that $\sum_{i=1}^n d_i$ is even and, for some probability distribution $(p_r)_{r=0}^\infty$ independent of n,*

(i) *$\#\{i : d_i = r\}/n \to p_r$ for every $r \geq 0$ as $n \to \infty$;*
(ii) *$\lambda := \sum_r r p_r \in (0,\infty)$;*
(iii) *$2m/n \to \lambda$ as $n \to \infty$.*

We shall consider thinnings of the vertex degrees in $G^*(n,(d_i)_1^n)$. Let W be a random variable with the distribution $\mathbb{P}(W = r) = p_r$. (This is the asymptotic distribution of the vertex degrees in $G^*(n,(d_i)_1^n)$.) For $0 \leq p \leq 1$ we let W_p be the thinning of W obtained by taking W points and then randomly and independently keeping each of them with probability p. For integers $l \geq 0$ and $0 \leq r \leq l$ let π_{lr} denote the binomial probabilities

$$\pi_{lr}(p) := \mathbb{P}\big(\mathrm{Bi}(l,p) = r\big) = \binom{l}{r} p^r (1-p)^{l-r}.$$

(The understanding here is that $\pi_{00}(p) = 1$ for all p.) Thus we have

$$\mathbb{P}(W_p = r) = \sum_{l=r}^\infty p_l \pi_{lr}(p).$$

We further define, for given $(p_r)_{r=0}^\infty$, functions

$$h(p) := \mathbb{E}\big(W_p \mathbf{1}[W_p \geq k]\big) = \sum_{r=k}^\infty \sum_{l=r}^\infty r p_l \pi_{lr}(p),$$

$$h_1(p) := \mathbb{P}(W_p \geq k) = \sum_{r=k}^\infty \sum_{l=r}^\infty p_l \pi_{lr}(p).$$

Note that both h and h_1 are increasing in p, with $h(0) = h_1(0) = 0$. Note further that $h(1) = \sum_{r=k}^\infty r p_r \leq \lambda$ and $h_1(1) = \sum_{r=k}^\infty p_r \leq 1$, with strict inequalities unless $p_r = 0$ for all $r = 1,\dots,k-1$ or $r = 0,1,\dots,k-1$, respectively.

The following theorems are our central results, and are key to proving Theorem 1.1. See Fernholz and Ramachandran [7, 8] and in particular Cooper [3] for similar results.

Theorem 2.3. *Consider the random multigraph $G^*(n,(d_i)_1^n)$ for a sequence $(d_i)_1^n$ satisfying Condition 2.2. Let $k \geq 2$ be fixed, and let Core_k^* be the k-core of $G^*(n,(d_i)_1^n)$. Let $\widehat{p}$ be the largest $p \leq 1$ such that $\lambda p^2 = h(p)$.*

(i) *If $\widehat{p} = 0$, i.e. if $\lambda p^2 > h(p)$ for all $p \in (0,1]$, then Core_k^* has $o_{\mathrm{p}}(n)$ vertices and $o_{\mathrm{p}}(n)$ edges **whp**. Furthermore, if also $k \geq 3$ and $\sum_{i=1}^n e^{\alpha d_i} = O(n)$ for some $\alpha > 0$, then Core_k^* is empty **whp**.*

(ii) *If* $\widehat{p} > 0$, *and further* $\lambda p^2 < h(p)$ *for* p *in some interval* $(\widehat{p} - \varepsilon, \widehat{p})$, *then* ***whp*** Core_k^* *is non-empty, and* $v(\mathrm{Core}_k^*)/n \xrightarrow{\mathrm{p}} h_1(\widehat{p})$, $e(\mathrm{Core}_k^*)/n \xrightarrow{\mathrm{p}} h(\widehat{p})/2 = \lambda\widehat{p}^2/2$.

Theorem 2.4. *Suppose Condition 2.2 holds, and further* $\sum_i d_i^2 = O(n)$. *Then all the conclusions of Theorem 2.3 hold also for the random graph* $G(n, (d_i)_1^n)$.

Naturally, the extra condition $\sum_{i=1}^n e^{\alpha d_i} = O(n)$ in Theorem 2.3(i) implies the extra condition in Theorem 2.4.

3. Finding the core

It is well-known (see for instance [19]) that the k-core of an arbitrary finite graph or multigraph can be found by removing vertices of degree $<k$, in arbitrary order, until no such vertices exist. It is easily seen that we obtain the same result by removing edges where one endpoint has degree $<k$, until no such edges remain, and finally removing all isolated vertices. Again, the order of removal does not matter, and we will use a randomized choice as follows.

Regard each edge as consisting of two *half-edges*, each half-edge having one endpoint. Say that a vertex is *light* if its degree is $< k$, and *heavy* if its degree is $\geq k$. Similarly, say that a half-edge is light or heavy when its endpoint is. As long as there is any light half-edge, choose one such half-edge uniformly at random and remove the edge it belongs to. (Note that this may change the other endpoint from heavy to light, and thus create new light half-edges.) When there are no light half-edges left, we stop. Then all light vertices are isolated; the heavy vertices and the remaining edges form the k-core of the original graph.

We apply this algorithm to a random multigraph with given degree sequence $(d_i)_1^n$. Let us observe only the vertex degrees in the resulting multigraph process, but not the individual edges. In other words, we observe the half-edges, but not how they are connected into edges. At each step, we thus select a light half-edge at random. We then reveal its partner, which is random and uniformly distributed over the set of all other half-edges. We then remove these two half-edges and repeat as long as there is any light half-edge. It is clear, by considering configurations, that this gives a Markov process (the state at any time $t \geq 0$ is the current degree sequence); and that at each step, conditioned on the vertex degrees observed so far, the remaining multigraph is a random multigraph with the given vertex degrees and the distribution specified in Section 2.

We shall analyse this process of half-edges further in Section 5.

4. Some death processes

This section contains some preliminary lemmas that will be used in our proofs. We begin with a classical result, see e.g. Proposition 4.24 in [12].

Lemma 4.1 (The Glivenko–Cantelli theorem)**.** *Let $T_1, \dots, T_n$ be i.i.d. random variables with distribution function $F(t) := \mathbb{P}(T_i \le t)$, and let $X_n(t)$ be their empirical distribution function $\#\{i \le n : T_i \le t\}/n$. Then $\sup_t |X_n(t) - F(t)| \xrightarrow{\text{p}} 0$ as $n \to \infty$.* □

Consider next a pure death process with rate 1; this process starts with some number of balls whose lifetimes are i.i.d. rate 1 exponentials Exp(1).

Lemma 4.2. *Let $N^{(n)}(t)$ be the number of balls alive at time t in a rate 1 death process with $N^{(n)}(0) = n$. Then*

$$\sup_{t \ge 0} \left| N^{(n)}(t)/n - e^{-t} \right| \xrightarrow{\text{p}} 0 \qquad \text{as } n \to \infty.$$

Proof. $1 - N^{(n)}(t)/n$ is the empirical distribution function of the n lifetimes, which are i.i.d. random variables with the distribution function $1 - e^{-t}$, $t \ge 0$. Hence the result is an instance of Lemma 4.1. □

The death process in Lemma 4.2 is a Markov process such that, whenever in state j, the process jumps to $j-1$ with intensity j, that is after a random time with distribution $\mathrm{Exp}(1/j)$. We extend this by allowing the process to take non-integer values as follows.

Lemma 4.3. *Let $\gamma > 0$ and $d > 0$ be fixed. Let $N^{(x)}(t)$ be a Markov process such that $N^{(x)}(0) = x$ a.s. and transitions are made according to the following rule: whenever in state $y > 0$, the process jumps to $y - d$ with intensity γy; in other words, the waiting time until the next event is $\mathrm{Exp}(1/\gamma y)$ and each jump is of size d downwards. Then*

$$\sup_{t \ge 0} \left| N^{(x)}(t)/x - e^{-\gamma d t} \right| \xrightarrow{\text{p}} 0 \qquad \text{as } x \to \infty.$$

Proof. Dividing $N^{(x)}(t)$ by d and t by γd we can rescale the process, and so we may just as well assume that $d = \gamma = 1$. The process is then the same as the one in Lemma 4.2 if $x = n$ is an integer. In general, consider $N^{(\lceil x \rceil)}(t)$, a rate 1 death process satisfying $N^{(\lceil x \rceil)}(0) = \lceil x \rceil$. We can couple $N^{(x)}(t)$ and $N^{(\lceil x \rceil)}(t)$ such that both jump whenever the smaller does, and it is easily seen that under the coupling $|N^{(x)}(t) - N^{(\lceil x \rceil)}(t)| < 1$ for all t. The result thus follows from Lemma 4.2, which yields $\sup_{t \ge 0} \left| N^{(\lceil x \rceil)}(t)/\lceil x \rceil - e^{-\gamma d t} \right| \xrightarrow{\text{p}} 0$. □

Now consider n bins with independent rate 1 death processes. Let $N_j^{(n)}(t)$ denote the number of balls in bin j at time t, where $j = 1, \dots, n$ and $t \ge 0$. Let further $U_r^{(n)}(t) := \#\{j : N_j^{(n)}(t) = r\}$, the number of bins with exactly r balls, for $r = 0, 1, \dots$. In what follows we suppress the superscripts to lighten the notation.

Lemma 4.4. *Consider n independent pure death processes $N_i(t)$ with rate 1 such that $N_i(0) = d_i$, where $(d_i)_1^n$ satisfies Condition 2.2. Then, with the* *above notation, as $n \to \infty$,*

$$\sup_{t\ge 0} \sum_{r=0}^{\infty} r \left| U_r(t)/n - \sum_{l=r}^{\infty} p_l \pi_{lr}(e^{-t}) \right| \xrightarrow{\text{p}} 0.$$

In particular,

$$\sup_{t\ge 0} \left| \sum_{r=k}^{\infty} r U_r(t)/n - h(e^{-t}) \right| \xrightarrow{\text{p}} 0, \tag{4.1}$$

$$\sup_{t\ge 0} \left| \sum_{r=k}^{\infty} U_r(t)/n - h_1(e^{-t}) \right| \xrightarrow{\text{p}} 0. \tag{4.2}$$

Proof. Let $U_{lr}(t)$ be the number of bins that have l balls at time 0 and r balls at time t. We shall actually prove the stronger result

$$\sup_{t\ge 0} \sum_{l=0}^{\infty} \sum_{r=0}^{l} r \left| U_{lr}(t)/n - p_l \pi_{lr}(e^{-t}) \right| \xrightarrow{\text{p}} 0. \tag{4.3}$$

First fix integers l and j, with $1 \le j \le l$. Consider the $u_l := U_l(0)$ bins that start with l balls. For $i = 1, \dots, u_l$ let T_i be the time the j-th ball is removed from the i-th such bin. Then $\#\{i : T_i \le t\} = \sum_{s=0}^{l-j} U_{ls}(t)$. Moreover, the number of balls remaining in one of these bins at time t has the distribution $\mathrm{Bi}(l, e^{-t})$, and thus $\mathbb{P}(T_i \le t) = \sum_{s=0}^{l-j} \pi_{ls}(e^{-t})$. Multiplying by u_l/n and using Lemma 4.1, we obtain that

$$\sup_{t\ge 0} \left| \frac{1}{n} \sum_{s=0}^{l-j} U_{ls}(t) - \frac{u_l}{n} \sum_{s=0}^{l-j} \pi_{ls}(e^{-t}) \right| \xrightarrow{\text{p}} 0.$$

Further, this convergence trivially holds when $l = 0$ or $j = 0$. But $u_l/n \to p_l$ by Condition 2.2(i), and so in fact, for all $j, l \ge 0$,

$$\sup_{t\ge 0} \left| \frac{1}{n} \sum_{s=0}^{l-j} U_{ls}(t) - p_l \sum_{s=0}^{l-j} \pi_{ls}(e^{-t}) \right| \xrightarrow{\text{p}} 0.$$

Take $j = l - r$ and $j = l - r + 1$ and subtract the corresponding quantities under the absolute value sign to deduce that each term in (4.3) tends to 0 in probability. Hence the same holds for any finite partial sum.

Finally, let $\varepsilon > 0$ and let L be such that $\sum_L^\infty l p_l < \varepsilon$. By Condition 2.2(iii), $\sum_l l u_l/n \to \lambda = \sum_l l p_l$. Hence also $\sum_{l\ge L} l u_l/n \to \sum_{l \ge L} l p_l < \varepsilon$. Consequently, if n is large enough, $\sum_{l\ge L} l u_l/n < \varepsilon$, and

$$\begin{aligned}\sup_{t\ge 0} \sum_{l=L}^{\infty} \sum_{r=0}^{l} r \left| U_{lr}(t)/n - p_l \pi_{lr}(e^{-t}) \right| &\le \sup_{t\ge 0} \sum_{l=L}^{\infty} \sum_{r=0}^{l} r \Big(U_{lr}(t)/n + p_l \pi_{lr}(e^{-t}) \Big) \\ &\le \sum_{l=L}^{\infty} l \big(u_l/n + p_l \big) < 2\varepsilon.\end{aligned}$$

We conclude that (4.3) holds. □

5. Proof of Theorem 2.3

We continue to analyse the process of vertex degrees in the core-finding algorithm of Section 3 applied to a random multigraph with given degree sequence $(d_i)_1^n$. We regard vertices as bins and half-edges as balls. The description in Section 3 thus says that at each step we remove first one random ball from the set of balls in light bins (i.e. bins with $<k$ balls) and then a random ball without restriction. We stop when there are no non-empty light bins, and the k-core consists precisely of the heavy bins at the time we stop.

We thus alternately remove a random light ball and a random ball. We may just as well say that we first remove a random light ball. We then remove balls in pairs, first a random ball and then a random light ball, and stop with the random ball leaving no light ball to remove.

We change the description a little by introducing colours. Initially all balls are white, and we begin again by removing one random light ball. Subsequently, in each deletion step we first remove a random white ball and then recolour a random light white ball red; this is repeated until no more white light balls remain. If we consider only the white balls, this is evidently the same process as before.

We now run this deletion process in continuous time such that, if there are j white balls remaining, then we wait an exponential time with mean $1/j$ until the next pair of deletions. In other words, we make deletions at rate j. This means that each white ball is deleted with rate 1 and that, when we delete a white ball, we also colour a random light white ball red. Let $L(t)$ and $H(t)$ denote the numbers of light and heavy white balls at time t respectively; further, let $H_1(t)$ be the number of heavy bins.

Since red balls are ignored, we may make a final change of rules, and say that all balls are removed at rate 1 and that, when a white ball is removed, a random white light ball is coloured red; we stop when we should recolour a white light ball but there is no such ball. Note that all heavy balls are white, and that white balls yield our core-finding process.

Let τ be the stopping time of this process. First consider the white balls only. There are no white light balls left at τ, so $L(\tau)$ has reached zero. However, let us consider the last deletion & recolouring step as completed by redefining $L(\tau) := -1$; we then see that τ is characterized by $L(\tau) = -1$ and $L(t) \geq 0$ for $0 \leq t < \tau$. Moreover, the heavy balls left at τ (which are all white) are exactly the half-edges in the k-core. Hence the number of edges in the k-core is $\frac{1}{2}H(\tau)$, while the number of vertices is $H_1(\tau)$.

Moreover, if we consider only the total number $L(t) + H(t)$ of white balls in the bins, ignoring the positions, the process (up to time τ) is as follows: each ball dies at rate 1 and upon its death another ball is also sacrificed. The process $L(t) + H(t)$ thus is the death process studied in Lemma 4.3, with $\gamma = 1$ and $d = 2$. We start with an odd number $2m - 1$ of white balls, since we began by removing one light ball. Consequently, Lemma 4.3 yields

$$\sup_{t\le\tau}\bigl|L(t)+H(t)-2me^{-2t}\bigr| = o_{\mathrm{p}}\bigl(2m\bigr) = o_{\mathrm{p}}(n). \tag{5.1}$$

Next let us ignore the colours. Our final version of the process then becomes exactly the process studied in Lemma 4.4, apart from the initial removal of a light ball which does not affect the conclusions because, for each t, at most two $U_r(t)$ (in the notation of Section 4) are changed by ± 1.

Since all heavy balls are white, we have $H(t) = \sum_{r=k}^{\infty} rU_r(t)$ and $H_1(t) = \sum_{r=k}^{\infty} U_r(t)$. Hence, by (4.1) and (4.2),

$$\sup_{t\le\tau} |H(t)/n - h(e^{-t})| \xrightarrow{\mathrm{p}} 0, \tag{5.2}$$

$$\sup_{t\le\tau} |H_1(t)/n - h_1(e^{-t})| \xrightarrow{\mathrm{p}} 0. \tag{5.3}$$

In particular,

$$H(\tau)/n - h(e^{-\tau}) \xrightarrow{\mathrm{p}} 0, \quad \text{and} \quad H_1(\tau)/n - h_1(e^{-\tau}) \xrightarrow{\mathrm{p}} 0. \tag{5.4}$$

We deduce from (5.1), (5.2) and $2m/n \to \lambda$ that

$$\sup_{t\le\tau}\bigl|L(t)/n + h(e^{-t}) - \lambda e^{-2t}\bigr| \xrightarrow{\mathrm{p}} 0. \tag{5.5}$$

Assume now that $t_1 > 0$ is a constant independent of n with $t_1 < -\ln\widehat{p}$. Then $\widehat{p} < 1$ and thus $h(1) < \lambda$. Hence, by continuity, $h(p) - \lambda p^2 < 0$ on $(\widehat{p}, 1]$, and thus $h(e^{-t}) - \lambda e^{-2t} < 0$ for $t \le t_1$. By compactness, $h(e^{-t}) - \lambda e^{-2t} \le -c$ for $t \le t_1$ and some $c > 0$. But $L(\tau) = -1$, so if $\tau \le t_1$ then $L(\tau)/n + h(e^{-\tau}) - \lambda e^{-2\tau} < -c$ and from (5.5)

$$\mathbb{P}(\tau \le t_1) \to 0. \tag{5.6}$$

In case (i) we may take any finite t_1 here, and hence find $\tau \xrightarrow{\mathrm{p}} \infty$. As $h(0) = h_1(0) = 0$, (5.4) yields that

$$H(\tau)/n \xrightarrow{\mathrm{p}} 0, \qquad\qquad H_1(\tau) \xrightarrow{\mathrm{p}} 0.$$

The first claim now follows, since $v(\mathrm{Core}_k) = H_1(\tau)$ and $e(\mathrm{Core}_k) = H(\tau)/2$. The second claim will follow from Lemma 5.1 below.

In case (ii) we similarly let $t_2 \in (-\ln\widehat{p}, -\ln(\widehat{p}-\varepsilon))$. Then by the hypothesis $h(e^{-t_2}) - \lambda e^{-2t_2} = c > 0$. If $\tau > t_2$ then $L(t_2) \ge 0$, and thus $L(t_2)/n + h(e^{-t_2}) - \lambda e^{-2t_2} \ge c$. Consequently (5.5) implies that

$$\mathbb{P}(\tau \ge t_2) \to 0.$$

Since we can choose t_1 and t_2 arbitrarily close to $-\ln\widehat{p}$, together with (5.6) this shows that

$$\tau \xrightarrow{\mathrm{p}} -\ln\widehat{p}.$$

Combined with (5.4), this yields $H(\tau)/n \xrightarrow{\mathrm{p}} h(\widehat{p})$ and $H_1(\tau)/n \xrightarrow{\mathrm{p}} h_1(\widehat{p})$, which proves (ii). $\square$

It remains to prove the following lemma extending a result by Łuczak [15].

Lemma 5.1. *If $k \ge 3$ and $\sum_i e^{\alpha d_i} = O(n)$, then there exists $\delta > 0$ such that **whp** $G^*(n,(d_i)_1^n)$ has no non-empty k-core with fewer than δn vertices.*

Remark 5.2. The proof below shows the stronger statement that **whp** $G^*(n,(d_i)_1^n)$ has no non-empty subgraph with fewer than δn vertices and average degree at least k.

We begin the proof of Lemma 5.1 with a sublemma.

Lemma 5.3. *Consider a set X of $2m$ points and a subset $Y \subseteq X$ with y elements. Let M be a random perfect matching of X and let Z be the number of pairs in M where both members belong to Y. Then for every real $u \ge 0$*

$$\mathbb{P}(Z \ge u) \le \Bigl(\frac{y^2}{mu}\Bigr)^u. \tag{5.7}$$

Proof. Denote the right hand side of (5.7) by $f(u)$. Then either $f(u) \ge 1$ or $f(u) \ge f(\lceil u \rceil)$. Hence it suffices to prove (5.7) when u is an integer. In that case

$$\begin{aligned}\mathbb{P}(Z \ge u) \le \mathbb{E}\binom{Z}{u} &= \binom{y}{2u}\frac{(2u)!}{2^u u!}\frac{1}{(2m-1)\cdots(2m-2u+1)}\\ &= \binom{y}{2u}\frac{\binom{m}{u}}{\binom{2m}{2u}} \le \Bigl(\frac{y}{2m}\Bigr)^{2u}\binom{m}{u} \le \Bigl(\frac{y}{2m}\Bigr)^{2u}\Bigl(\frac{em}{u}\Bigr)^u = \Bigl(\frac{ey^2}{4mu}\Bigr)^u.\end{aligned}$$

□

Proof of Lemma 5.1. Let C be such that $\sum_i e^{\alpha d_i} \le Cn$.

Consider a set A of s vertices $i_1,\dots,i_s$, and let $D_A := \sum_{j=1}^s d_{i_j}$. If A is the vertex set of the k-core, it must contain at least $ks/2$ edges. By Lemma 5.3, using the inequality $x \le e^x$, the probability of this event is at most

$$\Bigl(\frac{2D_A^2}{mks}\Bigr)^{ks/2} = \Bigl(\frac{2ks}{m\alpha^2}\Bigr)^{ks/2}\Bigl(\frac{\alpha D_A}{ks}\Bigr)^{ks} \le \Bigl(\frac{2ks}{m\alpha^2}\Bigr)^{ks/2} e^{ks\alpha D_A/(ks)}.$$

Summing over all sets A with s vertices, we obtain

$$\begin{aligned}\mathbb{P}(v(\mathrm{Core}_k^*) = s) &\le \Bigl(\frac{2ks}{m\alpha^2}\Bigr)^{ks/2}\sum_{|A|=s}\prod_{i\in A} e^{\alpha d_i}\\ &= \Bigl(\frac{2ks}{m\alpha^2}\Bigr)^{ks/2}\Bigl(\frac{n}{s}\Bigr)^s\sum_{|A|=s}\prod_{i\in A}\frac{s}{n}e^{\alpha d_i}\\ &\le \Bigl(\frac{2ks}{m\alpha^2}\Bigr)^{ks/2}\Bigl(\frac{n}{s}\Bigr)^s\prod_{i=1}^n\Bigl(1+\frac{s}{n}e^{\alpha d_i}\Bigr)\\ &\le \Bigl(\frac{2ks}{m\alpha^2}\Bigr)^{ks/2}\Bigl(\frac{n}{s}\Bigr)^s\exp\Bigl(\sum_{i=1}^n\frac{s}{n}e^{\alpha d_i}\Bigr)\\ &\le \Bigl(\frac{2ks}{m\alpha^2}\Bigr)^{ks/2}\Bigl(\frac{n}{s}\Bigr)^s\exp(Cs)\end{aligned}$$

Since $2m/n \to \lambda$, $m > \lambda n/3$ for large n, so that

$$\mathbb{P}(v(\mathrm{Core}_k^*) = s) \le \Big(\Big(\frac{6k}{\lambda\alpha^2}\Big)^{k/2} \Big(\frac{s}{n}\Big)^{k/2-1} e^C \Big)^s. \tag{5.8}$$

Choosing δ such that

$$\Big(\frac{6k}{\lambda\alpha^2}\Big)^{k/2} \delta^{k/2-1} e^C = \frac{1}{2},$$

and considering the cases $s < \ln n$ and $s \ge \ln n$ separately, it is easily seen that the sum of the right hand side of (5.8) over $s \in [1, \delta n]$ is $o(1)$. $\square$

6. Proofs of Theorems 2.4 and 1.1

Proof of Theorem 2.4. It is shown in [9] that the condition $\sum_i d_i^2 = O(n)$ implies $\liminf \mathbb{P}(G^*(n,(d_i)_1^n) \text{ is simple}) > 0$. (This is well-known, see for instance [1] and [16], under some extra condition on $\max d_i$.) Hence the result follows from Theorem 2.3 by conditioning on $G^*(n,(d_i)_1^n)$ being simple. $\square$

Proof of Theorem 1.1. The degree sequence $(d_i)_1^n$ is now random, but Condition 2.2 holds for convergence in probability with $p_r = \mathbb{P}(\mathrm{Po}(\lambda) = r)$, see for example [1, Chapter III]. Choosing a suitable coupling of the random graphs $G(n, \lambda/n)$ for different n, we may thus assume that Condition 2.2 holds a.s.

Further, the vertex degrees d_i all have the same distribution, binomial $\mathrm{Bi}(n-1, \lambda/n)$ for $G(n, \lambda/n)$ and hypergeometric for $G(n,m)$, and it follows easily that $\mathbb{E}\sum_i e^{d_i} = n\,\mathbb{E}\,e^{d_1} = O(n)$. This implies that $\sum_i e^{d_i} = O_{\mathrm{p}}(n)$; by suitable conditioning we may thus assume $\sum_i e^{d_i} = O(n)$. Then Theorem 2.4 applies a.s. to $G(n, \lambda/n)$ or $G(n,m)$ conditioned on the degree sequence, with $(p_r)_r = \mathrm{Po}(\lambda)$. In the notation of Section 2, $W \sim \mathrm{Po}(\lambda)$ and so $W_p \sim \mathrm{Po}(\lambda p)$; hence $h_1(p) = \psi_k(\lambda p)$ and

$$h(p) = \sum_{j=k}^{\infty} j \frac{(\lambda p)^j}{j!} e^{-\lambda p} = \lambda p \psi_{k-1}(\lambda p).$$

Consequently,

$$\lambda p^2 > h(p) \iff p > \psi_{k-1}(\lambda p) \iff \frac{\lambda p}{\psi_{k-1}(\lambda p)} > \lambda.$$

It then follows that $\widehat{p} = 0 \iff \lambda < \mu/\psi_{k-1}(\mu)$ for all $\mu \le \lambda$. Since this inequality holds trivially for $\mu > \lambda$, we deduce that $\widehat{p} = 0 \iff \lambda < \lambda_k = \min_{\mu > 0} \mu/\psi_{k-1}(\mu)$, and so part (i) follows.

Similarly, if $\lambda > \lambda_k$, $\lambda\widehat{p} = \mu_k(\lambda)$, and (ii) follows, provided we show that $\mu/\psi_{k-1}(\mu) < \lambda$ for μ slightly less than $\mu_k(\lambda)$. This is done in Section 7 below. $\square$

7. A fixed point equation

To complete the proof of Theorem 1.1 we show the following lemma. Recall that $\lambda_k := \min_{\mu>0} \mu/\psi_{k-1}(\mu)$.

Lemma 7.1. (i) *Assume $k \ge 3$. If $\lambda > \lambda_k$, then the equation $\mu/\psi_{k-1}(\mu) = \lambda$ has exactly two positive solutions, $\mu^-(\lambda)$ and $\mu^+(\lambda)$, with $0 < \mu^-(\lambda) < \mu^+(\lambda)$; thus $\mu_k(\lambda) = \mu^+(\lambda)$. Moreover, $\mu/\psi_{k-1}(\mu) < \lambda$ for $\mu^-(\lambda) < \mu < \mu^+(\lambda) = \mu_k(\lambda)$.*

(ii) *Assume $k = 2$. If $\lambda > \lambda_k$, then the equation $\mu/\psi_{k-1}(\mu) = \lambda$ has exactly one positive solution, $\mu_k(\lambda)$, and $\mu/\psi_{k-1}(\mu) < \lambda$ for $0 < \mu < \mu_k(\lambda)$.*

Proof. Define $\varphi(\mu) := \psi_{k-1}(\mu)/\mu$.

For $k = 2$, $\psi_{k-1}(\mu) = 1 - e^{-\mu}$ so $\varphi(\mu) = \bigl(1 - e^{-\mu}\bigr)/\mu$. Hence φ is (strictly) decreasing on $(0, \infty)$ and $\mu/\psi_{k-1}(\mu)$ is increasing from $\lambda_2 = 1$ to ∞ for $\mu \in (0, \infty)$; the result follows.

For $k \ge 3$, the result follows immediately from the lemma below; note that

$$\lambda_k := \inf_{\mu>0} \frac{1}{\varphi(\mu)} = \frac{1}{\sup_{\mu>0} \varphi(\mu)}.$$

□

Lemma 7.2. *If $k \ge 3$, then $\varphi(x) := \psi_{k-1}(x)/x$ is unimodal: there is a unique maximum point $x_0 > 0$, $\varphi'(x) > 0$ for $0 < x < x_0$ and $\varphi'(x) < 0$ for $x > x_0$. Further, $\varphi(x) \to 0$ as $x \to 0$ or $x \to \infty$.*

Proof. Note first that φ is continuously differentiable on $(0, \infty)$ with $\varphi(x) > 0$, and that $\varphi(x) = O\bigl(x^{k-2}\bigr)$ as $x \to 0$, and $\varphi(x) \le 1/x$; hence $\varphi(x) \to 0$ as $x \to 0$ or $x \to \infty$. It follows that $\varphi(x)$ attains its maximum at some $x_0 > 0$.

Also $\psi'_{k-1}(x) = x^{k-2}e^{-x}/(k-2)!$, and thus $\psi_{k-1}(x)/\bigl(x\psi'_{k-1}(x)\bigr)$ is increasing. Hence

$$x\frac{d}{dx} \ln \varphi(x) = x\frac{\psi'_{k-1}(x)}{\psi_{k-1}(x)} - 1 \tag{7.1}$$

is decreasing. Since $\ln \varphi$ attains its maximum at x_0, $\frac{d}{dx} \ln \varphi(x_0) = 0$, and it follows from (7.1) that $\frac{d}{dx} \ln \varphi(x) > 0$ for $x < x_0$ and $\frac{d}{dx} \ln \varphi(x) < 0$ for $x > x_0$. □

Remarks 7.3. The proof shows that $y \mapsto \ln \varphi(e^y)$ is strictly concave.

In the language of discrete dynamical systems, see for instance [5], for $k \ge 3$, $\mu^{\pm}(\lambda)$ are the fixed points of $f_\lambda(x) := \lambda\psi_{k-1}(x)$, and f_λ undergoes a saddle-node bifurcation at $\lambda = \lambda_k$.

8. Further results

We have studied the k-core of a random multigraph with a given degree sequence. We have determined sufficient conditions on the asymptotic behaviour of the degree sequence for the k-core to be empty, or at least very small, with high probability. We have also given sufficient conditions for the multigraph to have a giant k-core such that the sizes of its vertex and edge sets obey a law of large numbers.

We have further given a new proof that the random graph $G(n, \lambda/n)$ (and hence also the random graph $G(n, m)$) exhibits threshold behaviour. That is, for each integer $k \geq 3$, there is a value λ_k such that, if $\lambda < \lambda_k$ then the k-core is empty **whp**; and if $\lambda > \lambda_k$ then the number of vertices and number of edges in the k-core are almost deterministic, and are very large.

We have not discussed the next level of detail. It is possible to obtain quantitative versions of our results, such as large deviation estimates and a central limit theorem for the size of the k-core. Also, one can use our method to study the transition window: how far above the threshold the edge probability λ/n must be to ensure that $G(n, \lambda/n)$ has a non-empty k-core **whp**. (Some such results were already given by Pittel et al. [19].) These and other issues will be considered in a forthcoming paper.

Furthermore, it seems possible to adapt the methods of this paper to random hypergraphs, but we leave this to the reader.

References

[1] B. Bollobás, *Random Graphs*, 2nd ed., Cambridge Univ. Press, Cambridge, 2001.

[2] J. Cain & N. Wormald, Encore on cores. *The Electronic Journal of Combinatorics*, **13** (2006), no. 1, R81.

[3] C. Cooper, The cores of random hypergraphs with a given degree sequence, *Random Struct. Alg.* **25** (2004), no. 4, 353–375.

[4] R. Darling & J.R. Norris, Cores and cycles in random graphs, in preparation, 2005.

[5] R.L. Devaney, *An Introduction to Chaotic Dynamical Systems*, 2nd ed., Westview Press, Boulder, CO, 2003.

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

[7] D. Fernholz & V. Ramachandran, The giant k-core of a random graph with a specified degree sequence, manuscript, 2003.

[8] D. Fernholz & V. Ramachandran, Cores and connectivity in sparse random graphs, Technical Report TR-04-13, The University of Texas at Austin, Department of Computer Sciences, 2004.

[9] S. Janson, The probability that a random multigraph is simple. `http://arxiv.org/math.CO/0609802`

[10] S. Janson, D.E. Knuth, T. Łuczak & B. Pittel, The birth of the giant component, *Random Struct. Alg.* **3** (1993), 233–358.

[11] S. Janson, T. Łuczak & A. Ruciński, *Random Graphs*, Wiley, New York, 2000.

[12] O. Kallenberg, *Foundations of Modern Probability*, 2nd ed., Springer, New York, 2002.

[13] J. Kim, The Poisson cloning model for random graphs, in preparation; presentation slides at `http://research.microsoft.com/theory/jehkim/`

[14] M.J. Luczak, A differential equation approximation for the k-core process, manuscript, 2005.

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

[16] B.D. McKay, Asymptotics for symmetric 0-1 matrices with prescribed row sums, *Ars Combin.* **19** A (1985), 15–25.

[17] M. Molloy, Cores in random hypergraphs and Boolean formulas, *Random Struct. Alg.* **27** (2005), no. 1, 124–135.

[18] M. Molloy & B. Reed, A critical point for random graphs with a given degree sequence, *Random Struct. Alg.* **6** (1995), no. 2–3, 161–179.

[19] B. Pittel, J. Spencer & N. Wormald, Sudden emergence of a giant k-core in a random graph, *J. Combin. Theor.* **6**7 (1996), 111–151.

Department of Mathematics, Uppsala University, PO Box 480, SE-751 06 Uppsala, Sweden

E-mail address: `svante.janson@math.uu.se`

URL: `http://www.math.uu.se/~svante/`

Department of Mathematics, London School of Economics, Houghton Street, London WC2A 2AE, United Kingdom

E-mail address: `malwina@planck.lse.ac.uk`

URL: `http://www.lse.ac.uk/people/m.j.luczak@lse.ac.uk/`