Available online at www.sciencedirect.com

ScienceDirect

Discrete Mathematics 309 (2009) 1476–1490

DISCRETE MATHEMATICS

www.elsevier.com/locate/disc

On the k-orientability of random graphs

Luc Devroye[a], Ebrahim Malalla[b,*]

[a] School of Computer Science, McGill University, Montreal, Canada H3A 2K6

[b] Information and Computer Sciences Department, King Fahd University of Petroleum & Minerals, Dhahran 31261, Saudi Arabia

Received 3 September 2006; received in revised form 6 February 2008; accepted 8 February 2008
Available online 8 April 2008

Abstract

Let $\mathbb{G}(n, m)$ be an undirected random graph with n vertices and m multiedges that may include loops, where each edge is realized by choosing its two vertices independently and uniformly at random with replacement from the set of all n vertices. The random graph $\mathbb{G}(n, m)$ is said to be k-orientable, where $k \geq 2$ is an integer, if there exists an orientation of the edges such that the maximum out-degree is at most k. Let $c_k = \sup\{c : \mathbb{G}(n, cn) \text{ is } k\text{-orientable w.h.p.}\}$. We prove that for k large enough, $1 - 2^k \exp\left(-k + 1 + \mathrm{e}^{-k/4}\right) < c_k/k < 1 - \exp\left(-2k\left(1 - \mathrm{e}^{-2k}\right)\right)$, and the time $c_k n$ is a threshold for the emergence of a giant subgraph of size $\Theta(n)$ whose edges are more than k times its vertices. Other results are presented.

Keywords: Random graphs; k-orientability; k-core; Maximum density; Static two-way chaining; Off-line load balancing; Competitive analysis; Probabilistic analysis of algorithms

1. Introduction

Consider an undirected random graph $\mathbb{G}(n, m)$ with n vertices and m edges where each edge connects two vertices chosen independently and uniformly at random, one after another, with replacement from the set of all n vertices. That is, the graph may contain multiedges or loops. An orientation of any graph is called a *k-orientation*, for any fixed integer $k \geq 1$, if and only if the maximum out-degree of the graph is at most k. If a k-orientation exists, we say that the graph is *k-orientable*. We remark that the k-orientability is a decreasing property. This means that if $\mathbb{G}(n, m)$ is k-orientable w.h.p., then $\mathbb{G}(n, m')$ is also k-orientable w.h.p., for all $m' \leq m$, where w.h.p. (with high probability) means with probability tending to one as $n \to \infty$.

Throughout, let $c_k = \sup\{c : \mathbb{G}(n, cn) \text{ is } k\text{-orientable w.h.p.}\}$. The purpose of this paper is to estimate c_k, the threshold of k-orientability. Clearly, $c_k \leq k$, because the random graph $\mathbb{G}(n, kn + 1)$ is not k-orientable, as each vertex can orient outward at most k edges. On the other hand, the best known [10] lower bound is $c_k \geq a_{k+1}/2$, where a_k is the threshold for the existence of the k-core (the unique maximal subgraph with minimum degree at least k). It is shown in [33] that $a_k = k + \sqrt{k \log k} + O(\log k)$, for k large enough. In this paper, we show that $c_k \sim k$. Indeed, we prove the following theorem.

* Corresponding author.
E-mail addresses: luc@cs.mcgill.ca (L. Devroye), malalla@kfupm.edu.sa (E. Malalla).

0012-365X/$ - see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.disc.2008.02.023

Theorem 1. *For k sufficiently large, we have*

$$\rho_k \stackrel{\text{def}}{=} 1 - 2^k \exp\left(-k + 1 + \mathrm{e}^{-k/4}\right) < c_k/k < 1 - \exp\left(-2k\left(1 - \mathrm{e}^{-2k}\right)\right).$$

Furthermore, c_k is the threshold for the emergence of a subgraph whose edges are more than k times its vertices. That is, if $X(n, c)$ is the indicator that the random graph $\mathbb{G}(n, cn)$ contains a subgraph whose edges are more than k times its vertices, then $c_k = \sup\{c : X(n, c) = 0$, w.h.p.$\}$. The size of the newborn subgraph that has such property and suddenly emerges around the time $m = c_k n$ is at least $\rho_k n$, w.h.p.

For small k, a lower bound on c_k can be computed by solving simultaneously two equations related to the estimation of the upper tail of the binomial distribution. See Table 3 for these bounds.

1.1. History and motivation

Many orientation processes of the random graph $\mathbb{G}(n, m)$ have been introduced and studied in the context of the historical problem of allocating balls into bins [22,23]. The edges and vertices of the random graph $\mathbb{G}(n, m)$ are viewed as balls and bins, respectively. Thus, each ball has two bins chosen independently and uniformly at random—one after another—with replacement. If a ball chooses two bins, say u and v, then inserting the ball into the bin u means orienting the corresponding edge (u, v) toward the vertex v. This means that the out-degree of any vertex u is equivalent to the load of the bin u which is defined to be the number of balls it contains. The maximum out-degree of the random graph, then, represents the maximum bin load. Thus, the goal is to design an efficient strategy for inserting the balls that minimizes the maximum bin load.

Researchers have studied many allocation processes that insert the balls *sequentially and on-line* into the bins. For example, *the classical allocation process* totally ignores the second choice of each ball and inserts it into the bin chosen first by the ball. In other words, each edge (u, v) is oriented on-line, i.e., upon realization, towards the vertex v. Gonnet [19] showed that if $m = \Theta(n)$, then, upon termination of this classical orientation process, the maximum out-degree of $\mathbb{G}(n, m)$ is asymptotic to $\log n/\log\log n$, in probability. See also [34].

On the other hand, *the greedy multiple-choice allocation process* inserts each ball upon arrival into the less loaded bin among its two bins, breaking ties randomly. This yields an on-line greedy orientation process that orients each edge upon realization toward the vertex with the minimum out-degree, where ties are broken randomly. Surprisingly, the maximum out-degree in this case improves dramatically. Azar et al. [3] proved that if $m = \Theta(n)$, then the maximum out-degree of $\mathbb{G}(n, m)$ is $\log_2 \log n + O(1)$, w.h.p. Other proofs are also presented in [5,29,35]. Another variant of the greedy orientation process [36,37] improves the maximum out-degree by a constant factor. The greedy multiple-choice paradigm is shown to be very fruitful in many applications in computer science such as load balancing and dynamic resource allocation [3,10,28], routing and interconnection networks [8,24,27], and hashing [6,7,11,13,25].

1.2. The uniform vertex model

The random graph $\mathbb{G}(n, m)$, however, is designed here mainly to study the k-orientability property, the subject of this paper, which is strongly related to the *off-line* version of the greedy multiple-choice allocation process. In the off-line process—unlike the on-line version—all the choices available for the balls are known in advance. That is, we assume that the whole random graph $\mathbb{G}(n, m)$ is known in advance before any orientation. The goal then is to solve the following problem: given a constant $c > 0$, find the smallest integer $k(c)$ such that the random graph $\mathbb{G}(n, cn)$ is k-orientable, w.h.p. Equivalently, for any fixed $k \in \mathbb{N}$, we would like to estimate the threshold $c_k = \sup\{c : \mathbb{G}(n, cn)$ is k-orientable w.h.p.$\}$. The off-line analysis (or the k-orientability) can be used to measure how good the on-line allocation process is. This is known as the *competitive analysis*, which has deep roots in load balancing [2–4,10]. It also provides a useful means for designing efficient static hashing schemes that achieve constant worst-case search time. Such schemes have been widely studied in the literature, see e.g., [31,32]. Other applications of the k-orientability like storing graphs and edge membership queries are explained in [1].

1.3. Known results

It is known that for any constant $c \in (0, 1/2)$, the uniform random graph $G(n, cn)$ of Erdös and Rényi [14], which has no loops or multiedges, consists of unicyclic connected components, and isolated trees, and when $c > 1/2$, there is also a unique component of size $\Theta(n)$ with more than one cycle. This classical result is also true for our random graph $\mathbb{G}(n, m)$, see, e.g., [21]. Clearly, any tree or unicyclic component can be oriented easily such that the maximum out-degree is at most one, see Fig. 1. A component that has more than one cycle is not 1-orientable. This means that $c_1 = 1/2$.

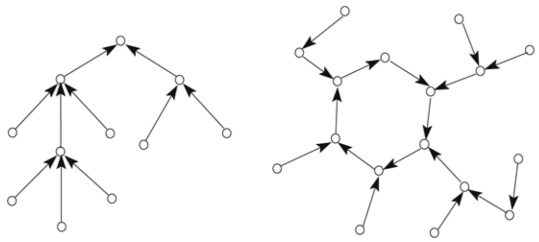

Fig. 1. Orienting the edges in tree and unicycle components such that the maximum out-degree is at most one. In a tree, a root is fixed first and then all the edges are oriented (in a BFS order) towards the root. In a unicycle, the edges of the cycle are oriented in any direction, and all other edges are oriented towards the cycle.

Azar et al. [3] proved that the random graph $\mathbb{G}(n, n)$ is 10-orientable, w.h.p. However, Czumaj and Stemann [9] tightened the result and proved that $\mathbb{G}(n, n)$ is indeed 2-orientable, w.h.p. More strongly, the authors showed in [10] (the final version of [9]) that for any positive constant $\alpha < 1.67545943\ldots$, the random graph $\mathbb{G}(n, \alpha n)$ is 2-orientable, w.h.p. The proof uses the threshold for the existence of the 3-core in random graphs [33], where the k-core is the unique maximal subgraph with minimum degree at least k. The result in its general form says that any undirected graph that does not contain a $(k+1)$-core is k-orientable. So what is the time of the emergence of the $(k+1)$-core? For $k \geq 3$, Pittel, Spencer, and Wormald [33] proved that the random birth time of the k-core in the random graph $G(n, m)$ is sharply concentrated around $m \approx a_k n/2$, where

$$a_k = \min_{\lambda > 0} \frac{\lambda}{\pi_{k-1}(\lambda)}, \quad \text{and} \quad \pi_k(\lambda) = \mathbb{P}\{\text{Poisson}(\lambda) \geq k\} = \sum_{i=k}^{\infty} \frac{e^{-\lambda}\lambda^i}{i!}.$$

Indeed, they showed that for any $\delta \in (0, 1/2)$, if $m \leq a_k n/2 - n^{1-\delta}$, then w.h.p., $G(n, m)$ does not contain any k-core; and if $m \geq a_k n/2 + n^{1-\delta}$, then w.h.p., there is a k-core that is connected and of size $p_k n + o(n)$, where $p_k = \pi_k(\lambda_k)$, and λ_k is the point at which the function $\lambda/\pi_{k-1}(\lambda)$ attains its minimum value. For large k, it is shown that $a_k = k + \sqrt{k \log k} + O(\log k)$. This result can be generalized to our model of random graph $\mathbb{G}(n, m)$, where loops and multiedges are allowed. All this shows that the random graph is k-orientable if it does not contain the $(k+1)$-core. However, the converse is not true, i.e., there are graphs that contain the $(k+1)$-core, yet they are still k-orientable, see Fig. 2. The above analysis only implies the inequality $c_k \geq a_{k+1}/2$, for $k \geq 2$. This means, for instance, that, $c_2 \geq 1.67545943\ldots$, $c_3 \geq 2.57470137\ldots$, and so on (see Table 1).

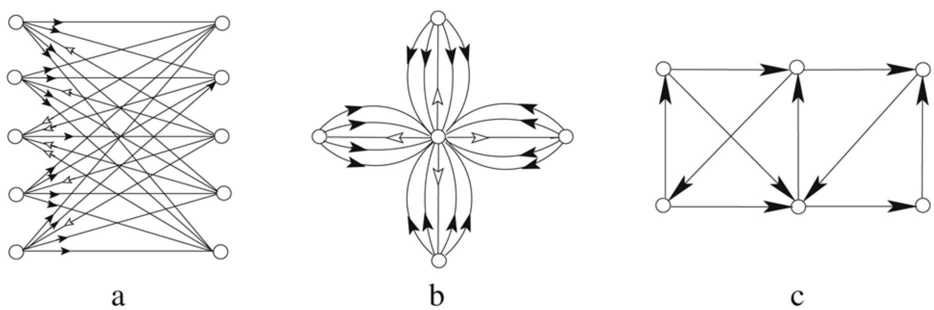

Fig. 2. The graphs in (a) and (b) are 5-cores, but, clearly, they are 4-orientable. The graph in (c) contains a 3-core, but it is still 2-orientable.

In Section 2, we recall a useful characterization of the k-orientability proved by Frank and Gyárfás [18], and we use it to prove Theorem 1 in Section 3 and 4. We also show that for small k, a lower bound on c_k can be computed

Table 1
Numerical computations showing the thresholds of the newborn $(k+1)$-core and the ratio of its (giant) size

$k+1$	a_{k+1}	$a_{k+1}/2$	p_{k+1}
3	3.35091887...	1.67545943...	0.267580655...
4	5.14940274...	2.57470137...	0.438061712...
5	6.79927548...	3.39963774...	0.538433561...
6	8.36534077...	4.18267038...	0.604638183...
7	9.87529072...	4.93764536...	0.651844404...
8	11.3441289...	5.67206445...	0.687379687...
9	12.7810996...	6.39054984...	0.715208554...
10	14.1923894...	7.09619474...	0.737666503...

The threshold $c_k \geq a_{k+1}/2$.

by solving simultaneously two equations related to the estimation of the upper tail of the binomial distribution. See Table 3 for these computed bounds which beat the $(k+1)$-core thresholds except for $k = 2, 3$.

2. Useful characterization

Throughout, we use the following notations and definitions. For any graph G, we write $\mathcal{V}(G)$ to denote the set of its vertices. For any set of vertices $S \subseteq \mathcal{V}(G)$, we write $\mathcal{E}(S)$ to denote the multiset of all edges whose endpoints belong to S. The *density* of any set of vertices S is the ratio $|\mathcal{E}(S)| / |S|$. If the density of a set S is strictly greater than k, for a positive integer k, we say that S is a *k-overloaded set*. The *maximum density* $\Psi(G)$ of any graph G is defined by

$$\Psi(G) = \max_{S \subseteq \mathcal{V}(G)} \lceil |\mathcal{E}(S)| / |S| \rceil .$$

That is, $\Psi(G)$ is the smallest integer such that $|\mathcal{E}(S)| \leq \Psi(G)\,|S|$, for all $S \subseteq \mathcal{V}(G)$.

We begin by restating the k-orientability property in terms of the edge distribution in the graph. Obviously, if a vertex has more than k loops, or if $|\mathcal{E}(G)| > k\,|\mathcal{V}(G)|$, then the graph G is not k-orientable, as each vertex can orient outward at most k edges. The following lemma generalizes this idea; see Fig. 3.

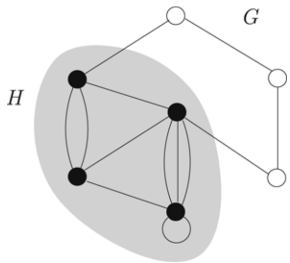

Fig. 3. The graph G is not 2-orientable, because H is a 2-overloaded set.

Lemma 1 (*Frank and Gyárfás [18]*). *Any graph G, possibly containing loops and multiedges, is k-orientable, where $k \in \mathbb{N}$, if and only if its maximum density $\Psi(G) \leq k$, that is, $|\mathcal{E}(S)| \leq k\,|S|$, for all $S \subseteq \mathcal{V}(G)$.*

This means that finding the maximum density of any graph is equivalent to finding the smallest integer k such that the graph is k-orientable. A general form of Lemma 1 was proved in [18], see also [16,17]. A simpler proof that uses the König–Hall theorem [12, Theorem 2.1.2] appeared in [1]. Malalla [25] gave a new constructive proof based on an algorithm which for any given graph G, and $k \in \mathbb{N}$, finds either a k-orientation or a k-overloaded set. The worst-case running time of the algorithm is $O(n^2)$, if $n = \mathcal{V}(G)$ and the density of G is constant. However, it is worth mentioning that Aichholzer, Aurenhammer, and Rote [1] gave an $O(n^{3/2})$ worst-case running time algorithm that is based on Hopcroft's and Karp's algorithm [20] for computing a maximum matching in a bipartite graph. The authors also presented a linear time heuristic for finding a $2k$-orientation.

We use Lemma 1 to prove the upper and lower bounds on the threshold c_k, in the following sections. Recall that $c_k = \sup\{c : \mathbb{G}(n, cn) \text{ is } k\text{-orientable w.h.p.}\}$. Notice that the existence of a k-overloaded set is an increasing property, i.e., if $\mathbb{G}(n, m)$ contains a k-overloaded set, w.h.p., then for all $m' > m$, the random graph $\mathbb{G}(n, m')$ also contains a k-overloaded set, w.h.p.

3. Upper bound

We recall the following inequality.

Lemma 2 (*McDiarmid [26]*). *Let $X_1, \ldots, X_n$ be independent random variables taking values in a set A, and let f be any real-valued measurable function defined on the set A^n. Suppose that for each $i \in [n]$, there exists $c_i > 0$ such that*

$$\sup_{x_1,\ldots,x_n,\hat{x}_i \in A} \left| f(x_1, \ldots, x_n) - f(x_1, \ldots, x_{i-1}, \hat{x}_i, x_{i+1}, \ldots, x_n) \right| \le c_i,$$

i.e., the function f has bounded differences. Then for any $t < 0$, we have

$$\mathbb{P}\{f(X_1, \ldots, X_n) - \mathbf{E}[f(X_1, \ldots, X_n)] \le t\} \le \exp\left(\frac{-2t^2}{\sum_{i=1}^{n} c_i^2}\right).$$

Notice that the random graph $\mathbb{G}(n, m)$ is constructed by choosing $2m$ vertices independently and uniformly at random, with replacement, where each two consecutive vertices represent an undirected edge. This means that each loop is chosen with probability $1/n^2$, and each undirected non-loop edge is chosen with probability $2/n^2$. Throughout, we define the *degree of a vertex* in the random graph $\mathbb{G}(n, m)$ to be the number of its non-loop incident edges plus twice the number of its loops, i.e., it is the number of times the vertex is chosen during the $2m$ trials of drawing the vertices. Clearly, the degree of any vertex is distributed as $\text{Bin}(2m, 1/n)$, that is, a binomial random variable with parameters $2m$ and $1/n$. The next theorem bounds c_k from above.

Theorem 2. *For any constant integer $k \ge 2$, let γ_k be the unique positive solution of $1 - \gamma - \mathrm{e}^{-2\gamma k} = 0$ on $(0, 1)$. Then $c_k \le \gamma_k k < \left(1 - \mathrm{e}^{-2k(1-\mathrm{e}^{-2k})}\right) k$.*

Proof. Suppose that $m = \gamma kn$, for some constant $\gamma \in (0, 1)$. To prove that $c_k < \gamma k$, it suffices to show that the random graph $\mathbb{G}(n, m)$ contains a k-overloaded set, w.h.p. Let S be the set of all non-isolated vertices, i.e., with degree of at least one, in the random graph $\mathbb{G}(n, m)$. Let X be the number of isolated vertices in the random graph $\mathbb{G}(n, m)$, and observe that

$$\begin{aligned}
\mathbf{E}[X] &= n\,\mathbb{P}\{\text{Bin}(2m, 1/n) = 0\} = n\left(1 - \frac{1}{n}\right)^{2m} \\
&= n\left(1 - \frac{1}{n}\right)^{2\gamma k(n-1)} \left(1 - \frac{1}{n}\right)^{2\gamma k} \\
&\ge n\mathrm{e}^{-2\gamma k}\left(1 - \frac{2k}{n}\right) \ge n\mathrm{e}^{-2\gamma k} - 2k,
\end{aligned}$$

where we have used the fact that $(1 - 1/n)^{n-1} \ge e^{-1}$. Notice that $|\mathcal{E}(S)| = m$, and $|S| = n - X$. Moreover, X can be expressed as a function of the $2m$ chosen vertices which are independent, and if one of the vertices is changed, X may increase or decrease by at most one. Therefore, by McDiarmid's inequality, we see that S is a k-overloaded set when γ is large enough. Indeed, for sufficiently large n, we have

$$\mathbb{P}\{|\mathcal{E}(S)| \le k\,|S|\} = \mathbb{P}\{X \le (1 - \gamma)n\}$$

$$\le \mathbb{P}\left\{X - \mathbf{E}[X] \le \left(1-\gamma-\mathrm{e}^{-2\gamma k}\right)n + 2k\right\}$$
$$\le \exp\left(-\left(1-\gamma-\mathrm{e}^{-2\gamma k}\right)^2 n/(\gamma k)+1\right) = o(1),$$

which is true whenever $f_k(\gamma) \stackrel{\text{def}}{=} 1-\gamma-\mathrm{e}^{-2\gamma k} < 0$. In particular, if $\gamma = 1-\mathrm{e}^{-2k(1-\mathrm{e}^{-2k})}$, then

$$f_k(\gamma) < \mathrm{e}^{-2k}\left(\mathrm{e}^{2k\mathrm{e}^{-2k}} - \mathrm{e}^{2k\mathrm{e}^{-2k}}\right) = 0.$$

This implies that $c_k/k \le \inf\{\gamma \in (0,1) : f_k(\gamma) < 0\} < 1-\mathrm{e}^{-2k(1-\mathrm{e}^{-2k})}$. However, $f_k(0) = 0$, $f_k(1/2) > 0$, $f_k(1) < 0$, and since $f_k''(\gamma) = -4\gamma^2\mathrm{e}^{-2\gamma k} < 0$, then f is concave on $[0,1]$. This means that in fact $\gamma_k = \inf\{\gamma \in (0,1) : f_k(\gamma) < 0\}$. □

Remark. Notice that the upper bound on c_k is obtained by estimating the random time at which the 1-core becomes a k-overloaded subgraph, that is, when the density of the 1-core exceeds k. One can improve this bound by considering instead the density of the $(k+1)$-core. That is, if C_k is the smallest constant c such that, w.h.p., the density of the $(k+1)$-core of the random graph $\mathbb{G}(n, cn)$ is more than k, then $c_k \le C_k$. We know from the work of Pittel, Spencer, and Wormald [33] that for $k \ge 2$, the $(k+1)$-core of the uniform random graph $G(n,m)$, where no loops or multiedges are allowed, emerges around the time $m \approx a_{k+1}n/2$, where

$$a_{k+1} = \min_{\lambda>0} \frac{\lambda}{\pi_k(\lambda)}, \quad \text{and} \quad \pi_k(\lambda) = \mathbb{P}\{\text{Poisson}(\lambda) \ge k\}.$$

Moreover, for any given constant $c > a_{k+1}/2$, the number of vertices in the $(k+1)$-core of the random graph $G(n, cn)$ is $\pi_{k+1}(\lambda_k(c))n + o(n)$, w.h.p., where $\lambda_k(c)$ is the largest root of the equation $2c = \lambda/\pi_k(\lambda)$. On the other hand, Fountoulakis [15] proved that the number of edges in the $(k+1)$-core of the random graph $G(n, cn)$ is $\lambda_k^2(c)n/(4c) + o(n)$, w.h.p. These results are also true in the model $\mathbb{G}(n,m)$ which is highly unlikely to have more than a constant number of loops or multiedges. Thus,

$$C_k = \inf\left\{c > \frac{a_{k+1}}{2} \,\middle|\, \frac{\lambda_k^2(c)}{4c\,\pi_{k+1}(\lambda_k(c))} > k\right\}.$$

Table 2 shows some of the computed values of C_k compared to the upper bound of Theorem 2.

Table 2
The threshold $c_k \le C_k \le \gamma_k k$

k	C_k	$\gamma_k k$
2	1.79402374...	1.960345197...
3	2.87746281...	2.992450613...
4	3.92147910...	3.998654534...
5	4.94775681...	4.999772897...
6	5.96443625...	5.999963132...
7	6.97541865...	6.999994180...
8	7.98282627...	7.999999100...

4. Lower bounds

We already know that c_k is at least the threshold of the $(k+1)$-core, which is asymptotic to $k/2$ [33]. In this section, we improve this lower bound, and show that indeed $c_k \sim k$, as $k \to \infty$. Observe that that for any set of vertices $S \subseteq \mathcal{V}(\mathbb{G}(n,m))$ of size $i \in \mathbb{N}$, the probability that we choose an edge whose both endpoints belong to S is i^2/n^2, because each vertex is drawn independently and uniformly at random, with replacement. Therefore, $|\mathcal{E}(S)| \stackrel{\mathcal{L}}{=} \text{Bin}(m, i^2/n^2)$. Thence, by Lemma 1, the probability that the random graph $\mathbb{G}(n,m)$ is not k-orientable is not more than

$$\sum_{i=1}^{\lfloor m/k \rfloor} \sum_{S:|S|=i} \mathbb{P}\{|\mathcal{E}(S)| > ki\} \le \sum_{i=1}^{\lfloor m/k \rfloor} \binom{n}{i} \mathbb{P}\left\{\mathrm{Bin}(m, i^2/n^2) > ki\right\}.$$

We would like to find the maximum constant c such that if $m = cn$, then the above probability tends to zero as n approaches infinity. Notice that for $i = 0, \dots, n$, we have

$$\binom{n}{i} \le \frac{n^i}{i!} \le \left(\frac{en}{i}\right)^i. \tag{1}$$

Recall also the following inequalities.

Lemma 3 (*Okamoto [30]*). *For $p \in (0, 1)$, and $r, t \in \mathbb{N}$, let $\beta := t/r$, and suppose that $r > t > rp > 0$. Then*

$$\mathbb{P}\{\mathrm{Bin}(r, p) \ge t\} \le \Upsilon(\beta, p)^r \stackrel{\text{def}}{=} \left(\left(\frac{1-p}{1-\beta}\right)^{1-\beta} \left(\frac{p}{\beta}\right)^{\beta}\right)^r \tag{2}$$

$$\le \left(\frac{epr}{t}\right)^t \mathrm{e}^{-pr}. \tag{3}$$

The following lemma shows that c_k is at least $k/\sqrt{e}$, asymptotically. However, the approximations used in the proof are not tight enough to prove that $c_k \ge (1 - o(1))k$. Nonetheless, the lemma is an important step towards the main result.

Lemma 4. *Let $k \ge 2$ be any constant integer. The random graph $\mathbb{G}(n, kn)$ does not contain any k-overloaded set of size less than or equal to $n\mathrm{e}^{-(k+1)/(k-1)}$, w.h.p. Furthermore, the threshold c_k is at least $k\mathrm{e}^{-(k+1)/(2k-1)}$.*

Proof. Let $j = \lfloor n\mathrm{e}^{-(k+1)/(k-1)} \rfloor$. Using (1) and (3), we see that for n large enough, the probability of existence of a k-overloaded set of size of at most j in the random graph $\mathbb{G}(n, kn)$ is not more than

$$\begin{aligned}
\sum_{i=1}^{j} \sum_{|S|=i} \mathbb{P}\{|\mathcal{E}(S)| > ki\} &\le \sum_{i=1}^{j} \binom{n}{i} \mathbb{P}\left\{\mathrm{Bin}(kn, i^2/n^2) > ki\right\} \\
&\le \sum_{i=1}^{j} \left(\frac{en}{i}\right)^i \left(\frac{ei}{n}\right)^{ki} \mathrm{e}^{-ki^2/n} \\
&\le \sum_{i=1}^{j} \left(\mathrm{e}^{k+1}(i/n)^{k-1}\right)^i \mathrm{e}^{-ki^2/n} \\
&\le \sum_{i=1}^{\lfloor \log n \rfloor} \frac{\mathrm{e}^{k+1} i}{n} + \sum_{\lceil \log n \rceil}^{\lfloor j/e \rfloor} \left(e^2 (j/n)^{k-1}\right)^i + \sum_{\lceil j/e \rceil}^{j} \mathrm{e}^{-ki^2/n} \\
&\le \frac{\mathrm{e}^{k+1} \log^2 n}{n} + \sum_{\lfloor \log n \rfloor}^{\infty} \mathrm{e}^{-i} + n \exp\left(-\mathrm{e}^{-2} k j^2/n\right) \\
&\le o(1) + \Theta(1/n) + o(1) = o(1).
\end{aligned}$$

Now if $m = \lfloor a_k kn \rfloor$, where $a_k = \mathrm{e}^{-(k+1)/(2k-1)}$, then w.h.p., the random graph $\mathbb{G}(n, m)$ is k-orientable, because the probability that there is a k-overloaded set of size greater than j is less than

$$\begin{aligned}
\sum_{i=j}^{\lceil m/k \rceil - 1} \left(\frac{en}{i}\right)^i \left(\frac{emi}{kn^2}\right)^{ki} \mathrm{e}^{-mi^2/n^2} &\le \sum_{i=j}^{\lfloor a_k n \rfloor} \left(\mathrm{e}^{k+1} (i/n)^{k-1} a_k^k\right)^i \mathrm{e}^{-mi^2/n^2} \\
&\le \sum_{i=j}^{\lfloor a_k n \rfloor} \left(\mathrm{e}^{k+1} a_k^{2k-1}\right)^i \mathrm{e}^{-mi^2/n^2}
\end{aligned}$$

$$= \sum_{i=j}^{\lfloor a_k n \rfloor} e^{-mi^2/n^2}$$
$$\leq n \exp\left(-mj^2/n^2\right) = o(1). \quad \square$$

Lemma 4 clearly improves the lower bound on c_k, for k large enough, but it also establishes a lower bound on the size ratio of any k-overloaded set, where the size ratio of any set is defined to be the size of the set divided by n. Indeed, if $\mathbb{G}(n, m)$, where $m \leq kn$, is not k-orientable, then the size ratio of the smallest k-overloaded set in $\mathbb{G}(n, m)$ is at least $e^{-(k+1)/(k-1)} \geq e^{-3}$, w.h.p. However, we shall further improve the lower bound on c_k, and see that the size ratio of the newborn k-overloaded set grows exponentially to 1, as $k \to \infty$. We do that by tightening the estimation of the upper tail of the binomial distribution: we shall use inequality (2). For that, we need to define the following positive functions. Fix an integer $k \geq 2$. Suppose that $m = \lfloor \alpha n \rfloor$, for some $\alpha \in (0, k]$. For $p \in (0, \alpha/k)$, let

$$f(k, \alpha, p) = \left(\frac{\alpha(1-p^2)}{\alpha - kp}\right)^{\alpha - kp} \left(\frac{\alpha p}{k}\right)^{kp}, \tag{4}$$

and define

$$h(k, \alpha, p) = \begin{cases} 1, & \text{if } p = 0; \\ p^{-p}(1-p)^{p-1} f(k, \alpha, p), & \text{for } p \in (0, \alpha/k); \\ (\alpha/k)^{2\alpha}, & \text{if } p = \alpha/k. \end{cases} \tag{5}$$

The functions f and h, as we are going to see further on, are related to the function Υ defined in Lemma 3. Notice that h is continuous on $[0, \alpha/k]$, and smooth on $(0, \alpha/k)$.

4.1. Tight asymptotic estimations

Our main asymptotic lower bounds are stated in the following theorem. We use the notation $h_p(k, \alpha, q)$ to denote the partial derivative of h with respect to p evaluated at the point (k, α, q).

Theorem 3. *For any fixed integer $k \geq 2$, define*

$$\alpha_k := \sup\left\{\alpha > 0 : \exists \delta \in (0, 1) \text{ such that } h(k, \alpha, p) \leq \delta, \forall p \in (e^{-3}, \alpha/k)\right\},$$

and let $\rho_k := 1 - (2/e)^k e^{1+e^{-k/4}}$. Then the following are true:

1. *The threshold c_k is at least α_k; and for k large enough, $\alpha_k > k\rho_k$.*
2. *If s_k is a point at which the function $h(k, \alpha_k, p)$ attains its maximum on the interval $[e^{-3}, \alpha_k/k]$, then $h(k, \alpha_k, s_k) = 1$, and $h_p(k, \alpha_k, s_k) = 0$.*
3. *For fixed $k \geq 2$, and $\alpha \in (\alpha_k, k]$, the equation $h(k, \alpha, p) = 1$ has two positive solutions. Let $q_1(k, \alpha)$ and $q_2(k, \alpha)$ be the smallest and the largest of these solutions. The size ratio of the newborn k-overloaded set is between $q_1(k, c_k)$ and $q_2(k, c_k)$, w.h.p. Moreover, $q_1(k, c_k) > r_k \stackrel{\text{def}}{=} q_1(k, k) \geq \rho_k$, for large k.*

Theorem 3 also provides a heuristic for computing the exact value of α_k. Solving the two equations $h(k, \alpha, p) = 1$, and $h_p(k, \alpha, p) = 0$, simultaneously, for any given $k \geq 2$, one can obtain the lower bound α_k. Unfortunately, solving these two equations explicitly is somehow impossible. So we used the mathematical software Maple to solve them numerically. The numerical computations of α_k (see Table 3) suggest, indeed, a more tight lower bound on α_k than the one mentioned in the theorem. We conjecture that for all $k \geq 2$,

$$\frac{\alpha_k}{k} \geq 1 - \left(\frac{2}{e}\right)^{k+\sqrt{k}}.$$

Note that this lower bound holds for each computed α_k in Table 3. The reader is invited to verify that.

Table 3
The numerical solutions α_k and s_k of the equations $h(k, \alpha, p) = 1$ and $h_p(k, \alpha, p) = 1$

k	α_k	s_k	k	α_k	s_k
2	1.30343190...	0.323260552...	15	14.9863940...	0.989411495...
3	2.48312473...	0.533227221...	16	15.9899586...	0.992227335...
4	3.61901095...	0.668655045...	17	16.9925912...	0.994293662...
5	4.71902985...	0.761197567...	18	17.9945347...	0.995810163...
6	5.79256286...	0.826480988...	19	18.9959692...	0.996923274...
7	6.84673418...	0.873351248...	20	19.9970278...	0.997740402...
8	7.88671563...	0.907333583...	21	20.9978087...	0.998340329...
9	8.91625922...	0.932106804...	22	21.9983847...	0.998780842...
10	9.93810345...	0.950220868...	23	22.9988094...	0.999104346...
11	10.9542584...	0.963487239...	24	23.9991226...	0.999341946...
12	11.9662054...	0.973211591...	25	24.9993535...	0.999516470...
13	12.9750390...	0.980342855...	26	25.9995236...	0.999644679...
14	13.9815687...	0.985573824...	27	26.9996490...	0.999738877...

The threshold $c_k \geq \alpha_k$ which is strictly greater than the threshold of the $(k+1)$-core (in Table 1), except for α_2 and α_3.

Recall that a k-overloaded set suddenly emerges around time $m = c_k n$. Theorem 3 reveals that one can lower bound the size ratio of the newborn k-overloaded set by computing the smallest positive root of $h(k, k, p) = 1$, which we call r_k. The theorem asserts that $r_k \geq 1 - (2/e)^k e^{1+e^{-k/4}}$, for $k \geq 2$. Obviously, it converges monotonically to one, as k goes to infinity. This is also illustrated in Table 4, and Fig. 4. Notice that the newborn k-overloaded set is giant (i.e., of size $\Theta(n)$). This is expected, because it is unlikely that at the beginning of the evolution, a large number of edges land on a very small set. However, while keeping the number of vertices fixed, and as the number of edges m increases away from $c_k n$ (for fixed $k \geq 2$), the size of the smallest k-overloaded set in $\mathbb{G}(n, m)$ starts to decrease to one.

Table 4
The size ratio of the newborn k-overloaded set is at least r_k, the solution of the equation $h(k, k, p) = 1$

k	r_k	k	r_k	k	r_k
2	0.061389845...	11	0.910842703...	20	0.994142089...
3	0.226773619...	12	0.933714444...	21	0.995686702...
4	0.387019206...	13	0.950841371...	22	0.996824664...
5	0.522609724...	14	0.963613455...	23	0.997662747...
6	0.632575890...	15	0.973107193...	24	0.998279818...
7	0.719774099...	16	0.980146405...	25	0.998734074...
8	0.787849394...	17	0.985355668...	26	0.999068429...
9	0.840355794...	18	0.989205077...	27	0.999314504...
10	0.880458374...	19	0.992046478...	28	0.999495594...

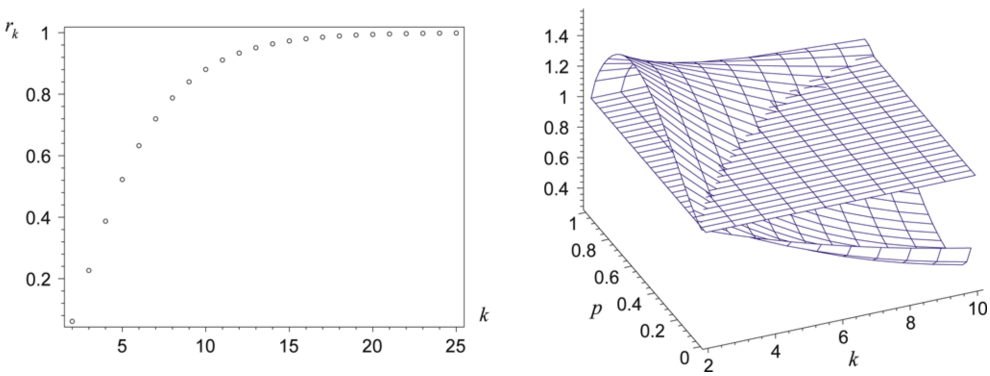

Fig. 4. The lower bound r_k on the size ratio of the newborn k-overloaded set converges exponentially to one. It is where the curve of $h(k, k, p)$ intersects 1.

4.2. Four technical lemmas

Before we start the proof of Theorem 3, we need to establish some lemmas. Recall that for $x \in (0, 1)$, we have

$$(1-x)\log(1-x) > -x, \qquad \text{or} \quad (1-x)^{1-x} > \mathrm{e}^{-x}. \tag{6}$$

The following lemma highlights some of the analytical properties of the function h. Fig. 5 illustrates some of these properties.

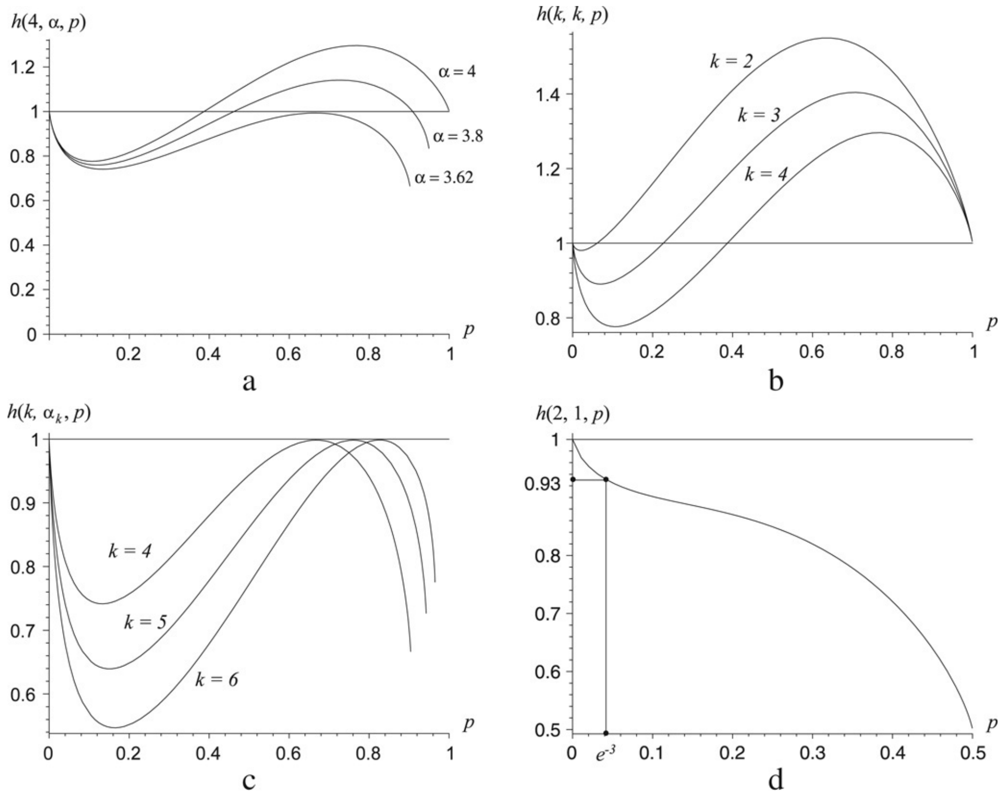

Fig. 5. Figure (a) shows the functions $h(4, 4, p)$, $h(4, 3.8, p)$, and $h(4, 3.62, p)$. It illustrates that the function h is strictly increasing on α, and for $\alpha \in (\alpha_k, k]$, the functions $h(k, \alpha, p)$ intersects the line $y = 1$ at two positive points. Figure (b) shows the functions $h(2, 2, p)$, $h_1(3, 3, p)$, and $h(4, 4, p)$. It illustrates that the function h is strictly decreasing on k and there is an $\epsilon \in (0, 1)$ such that $h(k, k, p) > 1$, for all $p \in (1-\epsilon, 1)$. Figure (c) shows the functions $h(4, \alpha_4, p)$, $h(5, \alpha_5, p)$, and $h(6, \alpha_6, p)$. Clearly, the function $h(k, \alpha_k, p) \leq 1$, on $(0, \alpha_k/k)$, where the equality holds only at one point. Figure (d) shows that the function $h(2, 1, p) < 0.93$ on $[e^{-3}, 1/2)$.

Lemma 5. *Let $\tilde{k} > k \geq 2$ be integers, and $\alpha, \tilde{\alpha} \in (0, k]$ be such that $\alpha < \tilde{\alpha}$. The following are true:*

1. *For all $p \in (0, \alpha/k)$, we have $h(k, \alpha, p) < h(k, \tilde{\alpha}, p)$; and if both $\alpha, \tilde{\alpha} > k/e$, then $h(k, \alpha, \alpha/k) < h(k, \tilde{\alpha}, \tilde{\alpha}/k)$.*
2. *For any constant $a \in (0, 1]$, we have $h(k, ak, p) > h(\tilde{k}, a\tilde{k}, p)$, for all $p \in (0, a)$.*
3. *$h(k, \alpha, p) < 1$, where $0 < p \leq \min(e^{-3}, \alpha/k)$.*
4. *There is an $\epsilon \in (0, 1)$ such that $h(k, k, p) > 1$, for all $p \in (1-\epsilon, 1)$.*

Proof.

1. By the definition of $f(k, \alpha, p)$, we see that for fixed $p \in (0, \alpha/k)$,

$$\begin{aligned}\frac{\partial}{\partial \alpha}(\log f) &= \log(\alpha - \alpha p^2) + 1 - kp/\alpha - \log(\alpha - kp) - 1 + kp/\alpha \\ &= \log\left(\frac{1-p^2}{1-kp/\alpha}\right) > 0,\end{aligned}$$

which is true because $1 - p^2 > 1 - kp/\alpha$. Since f is strictly positive, then

$$\frac{\partial}{\partial \alpha} f(k, \alpha, p) = f(k, \alpha, p) \frac{\partial}{\partial \alpha} \log f(k, \alpha, p) > 0.$$

This means that $f(k, \alpha, p)$, and hence $h(k, \alpha, p)$, is a strictly increasing function of α, where $p \in (0, \alpha/k)$. If $p = \alpha/k$, then

$$h(k, \alpha, \alpha/k) = \left(\frac{\alpha}{k}\right)^{2\alpha} < \left(\frac{\tilde{\alpha}}{k}\right)^{2\tilde{\alpha}} = h(k, \tilde{\alpha}, \tilde{\alpha}/k),$$

which is true because if $t(x) = 2x \log(x/k)$, where $x \in (k/e, k]$, then the derivative $t'(x) = 2 \log(x/k) + 2 > 0$, i.e., $t(x)$ is a strictly increasing function.

2. If $\alpha = ak$, for some constant $a \in (0, 1]$, and $p \in (0, a)$, we have

$$f(k, ak, p) = \left(\frac{a(1 - p^2)}{a - p}\right)^{k(a-p)} (ap)^{kp}.$$

Let $g(a, p) = (a - p) \log(1 - p^2) - (a - p) \log(1 - p/a)$, and notice that

$$\frac{\partial g}{\partial a} = \log(1 - p^2) - \log(1 - p/a) + 1 > 0,$$

because $1 - p^2 > 1 - p/a$. This means that for fixed $p \in (0, a)$, the function $g(a, p)$ strictly increases as a function of a. Thus, using the known inequalities: $\log p < p - 1$ and $\log(1 + p) < p$, we see that

$$\begin{aligned}\frac{\partial}{\partial k}(\log f(k, ak, p)) &= (a - p) \log \frac{1 - p^2}{1 - p/a} + p \log(ap)\\ &\leq (1 - p) \log \frac{1 - p^2}{1 - p} + p \log p\\ &< p(1 - p) + p(p - 1) = 0.\end{aligned}$$

Thus, the function $f(k, ak, p)$, and hence $h(k, ak, p)$, strictly decreases on k.

3. We know thus far that for any integer $k \geq 2$, $\alpha \in (0, k]$, and $p \in (0, \alpha/k)$, we have $h(k, \alpha, p) \leq h(k, k, p) \leq h(2, 2, p)$. However, using (6), we see that for $p \in (0, e^{-3}]$,

$$\begin{aligned}h(2, 2, p) &= (1 - p)^{-(1-p)}(1 + p)^{2(1-p)} p^p\\ &< \exp(p + 2p(1 - p) + p \log e^{-3})\\ &= \exp(-2p^2) \leq 1.\end{aligned}$$

4. When $\alpha = k$,

$$h(k, k, p) = \frac{(1 + p)^{k(1-p)} p^{kp}}{p^p (1 - p)^{1-p}},$$

and hence,

$$\frac{\partial}{\partial p}(\log h) = -k \log(1 + p) + \frac{k(1 - p)}{(1 + p)} + (k - 1) \log p + k + \log(1 - p),$$

which converges to $-\infty$ as p goes to 1. Since the derivatives of h and $\log h$ have the same sign, this means that $h(k, k, p)$ is strictly decreasing on $(1 - \epsilon, 1)$ for some positive ϵ, i.e., $h(k, k, p) > 1$, for all $p \in (1 - \epsilon, 1)$. □

Since h is continuous on its domain, $h(k, k, e^{-3}) < 1$ (Lemma 5), and $h(k, k, q) > 1$ for some $q \in (e^{-3}, 1)$, the equation $h(k, k, p) = 1$ must have a solution in (e^{-3}, q). The following lemma bounds the smallest such solution from below. The lemma helps us later on to establish the lower bound on α_k, and to prove that the smallest k-overloaded set in the random graph $\mathbb{G}(n, kn)$ has size ratio of at least $1 - 2^k \exp\left(-k + 1 + \mathrm{e}^{-k/4}\right)$.

Lemma 6. *For an integer $k \geq 2$, let r_k be the smallest positive root of the equation $h(k, k, r_k) = 1$. Then for k large enough,*

$$r_k > 1 - \left(\frac{2}{e}\right)^k \mathrm{e}^{1+\mathrm{e}^{-k/4}}.$$

Proof. Let $\rho_k = \exp((\log 2 - 1)k + 1 + \mathrm{e}^{-k/4})$. Since h is continuous, and by Lemma 5, we have $h(k, k, p) < 1$, for all $p \in (0, e^{-3}]$, then

$$r_k > 1 - \rho_k \iff h(k, k, p) < 1, \quad \text{for all } p \in (e^{-3}, 1 - \rho_k].$$

We shall show that for k large enough, the function $g(p) := \log h(k, k, p) < 0$, for all $p \in (e^{-3}, 1 - \rho_k]$. First notice that

$$\begin{aligned} g(p) &= k(1-p)\log(1+p) + (k-1)p\log p - (1-p)\log(1-p), \\ g'(p) &= -k\log(1+p) + \frac{k(1-p)}{1+p} + (k-1)\log p + \log(1-p) + k, \end{aligned}$$

and

$$g''(p) = \frac{-k}{1+p} - \frac{2k}{(1+p)^2} + \frac{k-1}{p} - \frac{1}{1-p}.$$

Thus,

$$\begin{aligned} g''(p) = 0 &\iff -kp(1-p^2) - 2kp(1-p) + (k-1)(1+p)^2(1-p) - p(1+p)^2 = 0 \\ &\iff (k-1)p^2 - 2(k+1)p + k - 1 = 0 \\ &\iff p = \frac{k+1-2\sqrt{k}}{k-1} \overset{\text{def}}{=} q_k. \end{aligned}$$

Evidently, $g''(p)$ is strictly positive on $(0, q_k)$, and negative on $(q_k, 1)$. This yields that $g(p)$ is strictly convex on $(0, q_k)$, and $g'(p)$ is decreasing on $[q_k, 1)$. Moreover, using (6), we see that for $p \in [q_k, 1 - \rho_k]$,

$$\begin{aligned} g'(p) &\geq g'(1-\rho_k) \\ &> k - k\log(2-\rho_k) + (k-1)\log(1-\rho_k) + \log\rho_k \\ &> (1-\log 2)k - \frac{(k-1)\rho_k}{1-\rho_k} + (\log 2 - 1)k + 1 \\ &= 1 - \frac{k2^k - 2^k}{\mathrm{e}^{k-2} - 2^k} > 0, \end{aligned}$$

when $k \geq 16$. This means that $g(p)$ is strictly increasing on $[q_k, 1 - \rho_k]$. Consequently, $g(p) \leq \max(g(e^{-3}),\ g(1 - \rho_k))$, for all $p \in [e^{-3}, 1 - \rho_k]$. However, we know that $g(e^{-3}) < 0$, and for k large enough, ($k \geq 100$), we have

$$\begin{aligned} g(1-\rho_k) &= k\rho_k\log(2-\rho_k) + (k-1)(1-\rho_k)\log(1-\rho_k) - \rho_k\log\rho_k \\ &< k\rho_k\log 2 - (k-1)\rho_k(1-\rho_k) - (\log 2 - 1)k\rho_k - \rho_k(1 + \mathrm{e}^{-k/4}) \\ &= \rho_k\left((k-1)\rho_k - \mathrm{e}^{-k/4}\right) < 0, \end{aligned}$$

which completes the proof. □

Next, we turn our attention to the definition of α_k. Let

$$\mathcal{A} := \left\{\alpha > 0 : \exists\delta \in (0,1) \text{ such that } h(k, \alpha, p) \leq \delta, \forall p \in (e^{-3}, \alpha/k)\right\},$$

and recall that $\alpha_k = \sup\mathcal{A}$. Clearly, if $\beta \in \mathcal{A}$, then $(0, \beta) \subseteq \mathcal{A}$, because h is an increasing function of α. Also, if $\gamma \notin \mathcal{A}$, then $\alpha_k \leq \gamma$. This leads to $\alpha_k \leq k$, because $h(k, k, 1) = 1$. The following lemma follows easily.

Lemma 7. *For any fixed integer $k \geq 2$, α_k is well-defined and $\alpha_k \in [k/2, k)$. Moreover, $h(k, \alpha_k, p) \leq 1$, for all $p \in [0, \alpha_k/k]$.*

Proof. First, α_k is well-defined because $\mathcal{A} \neq \varnothing$. Indeed, $k/2 \in \mathcal{A}$ because by Lemma 5-(2), we have $h(k, k/2, p) < h(2, 1, p) < 0.93$, for $p \in (e^{-3}, 1/2)$ (see Fig. 5(d)). Thus, trivially, $\alpha_k \geq k/2 > e^{-3}$. Now notice that

$h(k, \alpha_k, 0) = 1$, and $h(k, \alpha_k, \alpha_k/k) \le 1$. Also, by Lemma 5-(3), $h(k, \alpha_k, p) < 1$, for all $p \in (0, e^{-3}]$. So if possible, assume that there is a point $q \in (e^{-3}, \alpha_k/k)$ such that $h(k, \alpha_k, q) > 1$. By the definition of h, we have $h(k, qk, q) = q^{2qk} < 1$. Therefore, since h is a continuous increasing function of α, then there is $\tilde{\alpha} \in (qk, \alpha_k)$ such that $h(k, \tilde{\alpha}, q) = 1$. That is, $\tilde{\alpha} \notin \mathcal{A}$, because $q \in (e^{-3}, \tilde{\alpha}/k)$, and hence $\alpha_k \le \tilde{\alpha}$ which is a contradiction. Thus, $h(k, \alpha_k, p) \le 1$, for all $p \in (0, \alpha_k/k)$. This also shows that $\alpha_k \ne k$ because, by Lemma 5-(4), there is $q \in (e^{-3}, 1)$ such that $h(k, k, q) > 1$. Consequently, $\alpha_k < k$. □

Finally, we have the following lemma.

Lemma 8. *Let $k \ge 2$ be any fixed integer. The following are true:*

1. *For $\alpha \in (\alpha_k, k]$, the equation $h(k, \alpha, p) = 1$ has at least two positive solutions, and there exists a point $s(k, \alpha) \in (e^{-3}, \alpha/k)$ such that*
$$\max_{0 \le p \le \alpha/k} h(k, \alpha, p) = h(k, \alpha, s) > 1.$$
2. *The equation $h(k, \alpha_k, p) = 1$ has at least one positive solution.*
3. *For $\alpha, \tilde{\alpha} \in [\alpha_k, k]$, if $r(k, \alpha)$ is the smallest positive solution of $h(k, \alpha, p) = 1$, then $r(k, \tilde{\alpha}) > r(k, \alpha)$, whenever $\alpha > \tilde{\alpha}$.*

Proof. First, for $\alpha \in (\alpha_k, k]$, let $s(k, \alpha)$ be any point at which $h(k, \alpha, p)$ attains its maximum on $[e^{-3}, \alpha/k]$, and let $\lambda = h(k, \alpha, s)$, which is positive. If possible, assume that $\lambda \le 1$. Let $\beta = (\alpha + \alpha_k)/2$. Notice that $k/2 \le \alpha_k < \beta < \alpha$. Let q be any point at which $h(k, \beta, p)$ attains its maximum on $[e^{-3}, \beta/k]$. Then by Lemma 5-(1), we see that for all $p \in [e^{-3}, \beta/k]$,

$$h(k, \beta, p) \le \delta \stackrel{\text{def}}{=} h(k, \beta, q) < h(k, \alpha, q) \le \lambda \le 1.$$

Thus, the definition of α_k yields that $\beta \le \alpha_k$ which is a contradiction. Consequently, $\lambda > 1$. Since $h(k, \alpha, \alpha/k) \le 1$, and by Lemma 5-(3), $h(k, \alpha, p) \le 1$, for all $p \in [0, e^{-3}]$, then $s \in (e^{-3}, \alpha/k)$, and

$$\max_{0 \le p \le \alpha/k} h(k, \alpha, p) = h(k, \alpha, s) > 1.$$

Since $h(k, \alpha, s) > 1$, and $h(k, \alpha, e^{-3}) < 1$, then there is a point $q_1(k, \alpha) \in (e^{-3}, s)$ such that $h(k, \alpha, q_1) = 1$, because h is continuous. If $\alpha = k$, we know that $h(k, \alpha, \alpha/k) = 1$; and if $\alpha < k$, we have $h(k, \alpha, \alpha/k) < 1$, and hence—for the same reason again—there is a point $q_2(k, \alpha) \in (s, \alpha/k)$ such that $h(k, \alpha, q_2) = 1$. That is, $h(k, \alpha, p) = 1$ has at least two positive solutions. Next, let

$$\sigma_k := \lim_{\alpha \searrow \alpha_k} s(k, \alpha).$$

and notice that

$$h(k, \alpha_k, \sigma_k) = \lim_{\alpha \searrow \alpha_k} h(k, \alpha, s) \ge 1,$$

because h is continuous on each of its arguments. However, by Lemma 7, we have $h(k, \alpha_k, p) \le 1$, for all $p \in [0, \alpha_k/k]$. Thus, $h(k, \alpha_k, \sigma_k) = 1$. Finally, Lemma 5(3) yields that $h(k, \alpha, p) < h(k, \alpha, r(k, \alpha)) = 1$, for $p \in (0, r(k, \alpha))$. Since $h(k, \alpha, p)$ is an increasing function of α, then for any $\tilde{\alpha} \in [\alpha_k, \alpha)$, we have $h(k, \tilde{\alpha}, p) < h(k, \alpha, p) \le 1$, for all $p \in (0, r(k, \alpha)]$. This means that $r(k, \tilde{\alpha}) > r(k, \alpha)$. □

4.3. Proof of Theorem 3

First, we prove that $c_k \ge \alpha_k$. Let $\epsilon \in (0, 1)$ be any small arbitrary constant. Let $\beta = \alpha_k - \epsilon$. By Lemma 4 and since ϵ is arbitrary, it suffices to show that the random graph $\mathbb{G}(n, \lfloor \beta n \rfloor)$ does not contain any k-overloaded set of size $\ge e^{-3}n$. For $\lfloor e^{-3}n \rfloor \le i \le \lfloor (\beta n - 1)/k \rfloor$, let $p_i := i/n$, and notice that $p_i \in (e^{-3}, \beta/k)$. By the definition of α_k, there exist $\alpha > \beta$, and a constant $\delta \in (0, 1)$ such that $h(k, \alpha, p) \le \delta$, for all $p \in (e^{-3}, \alpha/k)$. Since h is an increasing function of α, then $h(k, \beta, p) \le h(k, \alpha, p) \le \delta$, for all $p \in (e^{-3}, \beta/k)$. Thus, using inequality (2) of Lemma 3, we see that the probability that $\mathbb{G}(n, \lfloor \beta n \rfloor)$ contains a k-overloaded set of size at least $e^{-3}n$ is not more than

$$\sum_{i=\lfloor e^{-3}n \rfloor}^{\lfloor (\beta n-1)/k \rfloor} \binom{n}{i} \mathbb{P}\left\{\mathrm{Bin}(\lfloor \beta n \rfloor, i^2/n^2) > ki\right\} \le \sum_{i=\lfloor e^{-3}n \rfloor}^{\lfloor (\beta n-1)/k \rfloor} \frac{n^n \Upsilon(kp_i/\beta, p_i^2)^{\beta n}}{i^i (n-i)^{(n-i)}}$$
$$= \sum_{i=\lfloor e^{-3}n \rfloor}^{\lfloor (\beta n-1)/k \rfloor} h(k, \beta, p_i)^n$$
$$\le n\,\delta^n = o(1).$$

Secondly, by Lemma 7, $h(k, \alpha_k, p) \le 1$, for all $p \in [0, \alpha_k/k]$. Recall that s_k is a point at which $h(k, \alpha_k, p)$ attains its maximum on $[e^{-3}, \alpha_k/k]$. From Lemma 8 we know that $h(k, \alpha_k, p) = 1$ has a solution, and thus, $h(k, \alpha_k, s_k) = 1$. Since $\alpha_k < k$, then by definition, $h(k, \alpha_k, \alpha_k/k) < 1$. That is, $s_k \in (0, \alpha_k/k)$ which leads to $h_p(k, \alpha_k, s_k) = 0$, because h is smooth on the open interval.

Thirdly, we know, by Lemma 8(1), that for $\alpha \in (\alpha_k, k]$, the equation $h(k, \alpha, p) = 1$ has at least two positive solutions. Notice that if $c_k = \alpha_k$, there is at least one solution for $h(k, c_k, p) = 1$, namely s_k, and we may have $q_1(k, c_k) = s_k = q_2(k, c_k)$. Nevertheless, the following is still true. Since $h(k, c_k, e^{-3}) < 1$, and $h(k, c_k, c_k/k) < 1$, then the definition of the two points $q_1(k, c_k)$ and $q_2(k, c_k)$ implies that $h(k, c_k, p) < 1$, for all $p \in (0, q_1) \cup (q_2, c_k/k)$. This means that for any arbitrary constant $\epsilon \in (0, 1)$ sufficiently small, there exists a constant $\delta \in (0, 1)$ such that $h(k, c_k, p) < \delta$ for all $p \in (e^{-3}, q_1 - \epsilon] \cup [q_2 + \epsilon, \alpha/k)$. Therefore, using the similar argument as above, we conclude that the random graph $\mathbb{G}(n, \lceil c_k n \rceil)$ does not contain any k-overloaded set of size less than $q_1(k, c_k)$ nor greater than $q_2(k, c_k)$, w.h.p. Finally, Lemma 8-(3) and 6 lead to

$$\alpha_k/k > s_k \ge q_1(k, \alpha_k) \ge q_1(k, c_k) > q_1(k, k) = r_k > 1 - \mathrm{e}^{1+\mathrm{e}^{-k/4}}(2/e)^k,$$

for k large enough. $\square$

Remark. Recall that the lower bounds on c_k recorded in Table 3 are better than the ones obtained from the $(k + 1)$-core analysis in Table 1, except for $k = 2, 3$, where $c_2 \ge 1.67545943\ldots > \alpha_2$ and $c_3 \ge 2.57470137\ldots > \alpha_3$. We should point out that, for $k = 3, 4, 5$, the lower bounds on c_k can be improved a little bit as is shown in Table 5. See [25, Sec. 3.5.2] for details.

Table 5
The threshold $c_k \ge \beta_k \ge \alpha_k$

k	β_k	α_k
3	2.61845509...	2.48312473...
4	3.65354252...	3.61901095...
5	4.71959504...	4.71902985...

Acknowledgments

The research of the authors was supported by NSERC Grant A3456. The second author wishes to acknowledge the support provided by KFUPM.

References

[1] O. Aichholzer, F. Aurenhammer, G. Rote, Optimal graph orientation with storage applications, SFB-Report F003-51, SFB ‘Optimierung und Kontrolle’, TU-Graz, Austria, 1995.
[2] Y. Azar, A.Z. Broder, A.R. Karlin, On-line load balancing, Theoretical Computer Science 130 (1994) 73–84.
[3] Y. Azar, A.Z. Broder, A.R. Karlin, E. Upfal, Balanced allocations, SIAM Journal on Computing 29 (1) (2000) 180–200.
[4] Y. Azar, J. Noar, R. Rom, The competitiveness of on-line assignments, Journal of Algorithms 18 (1995) 221–237.
[5] P. Berenbrink, A. Czumaj, A. Steger, B. Vöcking, Balanced allocations: The heavily loaded case, SIAM Journal on Computing 35 (6) (2006) 1350–1385.

[6] A. Broder, M. Mitzenmacher, Using multiple hash functions to improve IP lookups, in: Proceedings of 20th Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM 2001), pp. 1454–1463, 2001. Full version available as Technical Report TR–03–00, Department of Computer Science, Harvard University, Cambridge, MA, 2000.
[7] J. Byers, J. Considine, M. Mitzenmacher, Simple load balancing for distributed hash tables, in: Proceedings of the 2nd International Workshop on Peer-to-Peer Systems, 2003, pp. 80–87.
[8] R. Cole, B.M. Maggs, F. Meyer auf der Heide, M. Mitzenmacher, A.W. Richa, K. Schroeder, R.K. Sitaraman, B. Voecking, Randomized protocols for low-congestion circuit routing in multistage interconnection networks, in: Proceedings of the 29th Annual ACM Symposium on the Theory of Computing, STOC, 1998, pp. 378–388.
[9] A. Czumaj, V. Stemann, Randomized allocation processes, in: Proceedings of the 38th Annual IEEE Symposium on Foundations of Computer Science, FOCS, 1997, pp. 194–203.
[10] A. Czumaj, V. Stemann, Randomized allocation processes, Random Structures and Algorithms 18 (4) (2001) 297–331.
[11] K. Dalal, L. Devroye, E. Malalla, E. McLeish, Two-way chaining with reassignment, SIAM Journal on Computing 35 (2) (2005) 327–340.
[12] R. Diestel, Graph Theory, in: Graduate Texts in Mathematics, vol. 173, Springer-Verlag, New York, 1997.
[13] M. Dietzfelbinger, C. Weidling, Balanced allocation and dictionaries with tightly packed constant size bins, Theoretical Computer Science 380 (2007) 47–68.
[14] P. Erdös, A. Rényi, On the evolution of random graphs, Publication of the Mathematical Institute of the Hungarian Academy of Science 5 (1960) 17–61.
[15] N. Fountoulakis, Thresholds and the structure of sparse random graphs, Ph.D. Thesis, University of Oxford, 2003.
[16] A. Frank, On the orientation of graphs, Journal of Combinatorial Theory, Series B 28 (3) (1980) 251–261.
[17] A. Frank, Orientations of graphs and submodular flows, in: A.J.W. Hilton (Ed.), in: Congressus Numerantium, vol. 113, 1996, pp. 111–142.
[18] A. Frank, A. Gyárfás, How to orient the edges of a graph, in: Combinatorics, in: Colloquia Mathematica Societatis János Bolyai, vol. 18, 1976, pp. 353–364.
[19] G.H. Gonnet, Expected length of the longest probe sequence in hash code searching, Journal of the ACM 28 (1981) 289–304.
[20] J.E. Hopcroft, R. Karp, An $n^{5/2}$ algorithm for maximum matchings in bipartite graphs, SIAM Journal on Computing 2 (1973) 225–231.
[21] S. Janson, D.E. Knuth, T. Łuczak, B. Pittel, The birth of the giant component, Random Structures and Algorithms 4 (3) (1993) 233–358.
[22] N.L. Johnson, A.S. Kotz, Urn Models and Their Application: An Approach to Modern Discrete Probability Theory, John Wiley & Sons, New York, NY, 1977.
[23] V.F. Kolchin, B.A. Sevast'yanov, V.P. Chistyakov, Random Allocations, V. H. Winston & Sons, Washington, DC, 1978.
[24] M.L. Łuczak, E. Upfal, Reducing network congestion and blocking probability through balanced allocation, in: Proceedings of the 40th Annual IEEE Symposium on Foundations of Computer Science, FOCS, 1999, pp. 587–595.
[25] E. Malalla, Two-way hashing with separate chaining and linear probing, Ph.D. Thesis, School of Computer Science, McGill University, 2004.
[26] C. McDiarmid, On the method of bounded differences, in: J. Siemons (Ed.), Surveys in Combinatorics, in: London Mathematical Society Lecture Notes Series, vol. 141, 1989, pp. 148–188.
[27] F. Meyer auf der Heide, K. Schröder, F. Schwarze, Routing on networks of optical crossbars, in: Theoretical Computer Science, vol. 196, 1998, pp. 181–200.
[28] M.D. Mitzenmacher, The power of two choices in randomized load balancing, Ph.D. Thesis, Computer Science Department, University of California at Berkeley, 1996.
[29] M.D. Mitzenmacher, A. Richa, R. Sitaraman, The power of two random choices: A survey of the techniques and results, in: P. Pardalos, S. Rajasekaran, J. Rolim (Eds.), Handbook of Randomized Computing, 2000, pp. 255–305.
[30] M. Okamoto, Some inequalities relating to the partial sum of binomial probabilities, Annals of Mathematical Statistics 10 (1958) 29–35.
[31] R. Pagh, On the cell probe complexity of membership and perfect hashing, in: Proceedings of 33rd Annual ACM Symposium on Theory of Computing, STOC, 2001, pp. 425–432.
[32] R. Pagh, Hashing, randomness and dictionaries, Ph.D. Thesis, University of Aarhus, 2002.
[33] B. Pittel, J. Spencer, N. Wormald, Sudden emergence of a giant k-core in a random graph, Journal of Combinatorial Theory, Series B 67 (1996) 111–151.
[34] M. Raab, A. Steger, Balls and bins — a simple and tight analysis, in: Proceedings of the 2nd Workshop on Randomization and Approximation Techniques in Computer Science, in: LNCS, vol. 1518, Springer-Verlag, 1998, pp. 159–170.
[35] T. Schickinger, A. Steger, Simplified witness tree arguments, in: Proceedings of the 27th Annual Conference on Current Trends in Theory and Practice of Informatics, in: LNCS, vol. 1963, Springer-Verlag, 2000, pp. 71–87.
[36] B. Vöcking, How asymmetry helps load balancing, Journal of the ACM 50 (4) (2003) 568–589.
[37] B. Vöcking, Symmetric vs. asymmetric multiple-choice algorithms, in: Proceedings of the 2nd ARACNE Workshop, Aarhus, 2001, pp. 7–15.