Cores and Connectivity in Sparse Random Graphs

Daniel Fernholz
Department of Computer Sciences
University of Texas at Austin
fernholz@cs.utexas.edu

Vijaya Ramachandran
Department of Computer Sciences
University of Texas at Austin
vlr@cs.utexas.edu

April 30, 2004

UTCS Technical Report TR04-13

Abstract

We consider k-connectivity in a sparse random graph with a specified degree sequence. When dealing with sparse random graphs, properties that require connectivity are most appropriately phrased in terms of a giant subgraph that satisfies that property since a sparse random graph is generally not connected. Here a giant subgraph is one that contains a constant fraction of the vertices in the original graph.

We obtain a tight threshold on the existence of a giant k-vertex or k-edge connected subgraph for $k \geq 3$ in a sparse random graph drawn from $\mathcal{G}_{n,p}$. Although k-connectivity in $\mathcal{G}_{n,p}$ has been widely studied in the literature, results for higher connectivity have applied mainly to the case when $p \geq \ln n/n$, which makes it likely that the graph is connected, and very little is known about higher connectivity in sparse $\mathcal{G}_{n,p}$.

The k-core of a graph is the maximal induced subgraph with minimum degree k. The key tool in the derivation of our connectivity results is our recent theorem and proof strategy for the giant k-core threshold in sparse random graphs with a specified degree sequence.

A degree sequence exhibits a power law if the number of vertices of degree i is proportional to $\frac{1}{i^\beta}$, for a suitable constant β. We establish that for every $k \geq 3$, a random power-law graph has a giant k-core if $2 < \beta < 3$, and it has no giant 3-core if $\beta \geq 3$. Thus for $\beta \geq 3$ our results establish that a random power-law graph has no giant 3-connected subgraph. For $2 < \beta < 3$ we derive a weaker result, that any $(k-1)$-separator in the k-core, must separate a set of size at most n^c where $c < 1$ depends on β.

Finally, the existence of a giant k-core in a random graph is related in an informal way to the probability that the genealogy tree of a certain branching process contains a perfect infinite $(k-1)$-ary tree. We provide a solution to this latter problem in terms of probability generating functions. This result is tangential to the rest of our paper, but may be of independent interest.

This work was supported in part by NSF grant CCR-9988160.

1 Introduction

We consider vertex- and edge-connectivity in graphs that are drawn uniformly at random conditional on a given degree distribution $\mathcal{D}$ that satisfies certain properties. We consider degree distributions that are *1-smooth* (roughly speaking this means that the graph is sparse, i.e., has a number of edges linear in the number of vertices), that have maximum degree $o(n^{1/3})$, and that have the second moment of the degree distribution bounded by either $O(n^{1/2-\epsilon})$ or by a constant (this latter property is called 2-smoothness). This model includes sparse graphs with Poisson degree distributions which are related to the classical random graphs $\mathcal{G}_{n,p}$ and $\mathcal{G}_{n,m}$, as well as random 'power law' graphs, which are governed by parameters α and β: a random power-law graph is chosen uniformly from all graphs with $y = \left\lfloor e^{\alpha}/x^{\beta} \right\rfloor$ vertices of degree x [1].

A sparse random graph is generally not connected. Thus questions relating to any property that requires connectivity need to be phrased in terms of the existence of a large subgraph with that property. In this paper we consider the existence of a *giant* subgraph with a given property, where a giant subgraph is one that contains a constant fraction of the vertices in the original graph.

We give a sharp threshold for the existence of a giant k-edge connected or k-vertex connected subgraph in sparse $\mathcal{G}_{n,p}$ for $k \geq 3$. Although vertex connectivity in $\mathcal{G}_{n,p}$ has been widely studied, most of these results have applied to the case when $p \geq \ln n/n$, which makes it likely that the graph is connected, and very little is known about higher connectivity in sparse $\mathcal{G}_{n,p}$.

The k-core of a graph is the maximal induced subgraph with minimum degree k. In a recent unpublished manuscript [6], we found conditions under which the k-core of a random graph with a 1-smooth degree sequence almost surely contains a constant fraction of the graph's vertices. The results from [6] are key tools used in the derivation of our results on vertex- and edge-connectivity. In particular, for $\mathcal{G}_{n,p}$ and other 2-smooth random graphs we show that the giant k-core of such a graph, if it exists, is almost surely k-connected.

The threshold for giant k-core was solved earlier by Pittel, Spencer, and Wormald [14] for random graphs drawn from $\mathcal{G}_{n,p}$, using a fairly involved proof. Recently the k-core problem has been studied for random hypergraphs in [13]. Very recently we have been informed that results in an unpublished manuscript [4] on random hypergraphs also provide results similar to those in our unpublished k-core paper [6].

We also consider power-law graphs. A degree sequence exhibits a power law if the number of vertices of degree i is proportional to $\frac{1}{i^{\beta}}$, for a suitable constant β. Random graphs with power law degree sequences are of some interest, since graphs that occur in the real world, including the web graph, phone-call graphs, networks of molecules, and networks of social interaction, often exhibit a power law degree sequence. Although it is unlikely that real-world graphs are accurately modeled by random graphs, or for that matter, by graphs which precisely obey a power law, it is nevertheless of interest to know what can be said about random graphs that obey degree sequence similar to several real-world graphs.

For $\beta > 2$ the power-law degree sequence is 1-smooth, and for $\beta > 3$ it is 2-smooth as well. We apply our k-core theorem from [6] to establish that for every constant $k \geq 3$ a random power-law graph almost surely contains a giant k-core if $2 < \beta < 3$, and it almost surely does not contain a giant 3-core if $\beta \geq 3$. We use these results to derive some results on k-connectivity in random power-law graphs.

Finally, the k-core threshold is a key tool in our results, and the existence of a giant k-core in a random graph is related in an informal way to the probability that the genealogy tree of a certain branching process [7] contains a perfect infinite $(k-1)$-ary tree. In this paper we provide a solution to the latter problem in terms of probability generating functions.

The rest of this paper is organized as follows. In section 2 we give some basic definitions and describe the configuration model (CM) for generating a random fixed-degree sequence graph. In section 3 we prove our theorem on the existence of an infinite complete r-ary subtree of a branching tree. We also describe in this section an informal connection between this problem and the k-core threshold, and state our k-core theorem (Theorem 3.2) from [6]. The complete proof of Theorem 3.2 has been included in the appendix for ease of reference since neither [6] nor [4] has been published. In section 4 we apply Theorem 3.2 to obtain the giant k-core thresholds for power-law distributions and for distributions with all convergent moments; this latter result allows us to rederive the giant k-core threshold for $\mathcal{G}_{n,p}$, which was first proved in [14]. Finally in section 5 we present our results on k-connectivity in random fixed-degree sequence graphs.

2 Preliminaries

2.1 Random Graph Definitions

We begin by providing definitions for random graphs with fixed degree sequences (see Molloy and Reed [11, 12]). A sequence $D = \{d_1, d_2, \ldots, d_n\}$ is *graphical* if the set Ω_D of (labelled) graphs with degree sequence D (i.e. such that the degree of the i'th vertex is d_i) is nonempty. If D is a graphical sequence, let $G(D)$ denote a uniformly distributed random element of Ω_D. Thus $G(D)$ is a *random graph with degree sequence* D.

An *asymptotic degree sequence* $\mathcal{D}$ is an infinite sequence $D_1, D_2, \ldots$, where each $D_n = \{d_{n,1}, \ldots, d_{n,n}\}$ is a graphical sequence of length n. A *random graph with asymptotic degree sequence* $\mathcal{D}$, denoted by $G(\mathcal{D})$, is a sequence of random graphs $G(D_n)$. The random graph $G(\mathcal{D})$ has a property P *asymptotically almost surely* (*a.a.s.*) if the probability that $G(D_n)$ has property P converges to 1 as $n \to \infty$; $G(\mathcal{D})$ does not have property P *with exponentially high probability* (*w.e.h.p.*) if the probability that $G(D_n)$ has property P is $c^{-\Omega(n^\epsilon)}$, for some $c > 1$ and $\epsilon > 0$.

For any degree sequence D and any $k > 0$, we define the *kth moment of* D

$$M_k(D) = \frac{1}{n}\sum_{i=1}^{n} d_i^k. \tag{1}$$

An asymptotic degree sequence $\mathcal{D}$ is k*-smooth* if if there exists a sequence of real numbers $\lambda_0, \lambda_1, \ldots$ such that

$$\text{Condition S1:} \qquad \lim_{n\to\infty} \frac{|\{j : d_j = i\}|}{n} = \lambda_i \text{ for all } i, \text{ and} \tag{2}$$

$$\text{Condition S2:} \qquad \lim_{n\to\infty} M_k(D_n) = \sum_{i=0}^{\infty} i^k \lambda_i < \infty. \tag{3}$$

The sequence λ_i is the *limiting degree distribution* of $\mathcal{D}$. Throughout this paper, whenever a property of random graph with degree sequence D is described asymptotically, it is assumed that D is part of a 1-smooth asymptotic degree sequence.

If $\mathcal{D}' = \{D'_1, D'_2, \ldots\}$ is a sequence of random (degree) sequences, we say $\mathcal{D}$ is a.a.s. k-smooth if the convergences described by conditions S1 and S2 occur in probability; that is, if for every $\epsilon > 0$

$$P\left[\left|\frac{|\{j : d_j = i\}|}{|D'_n|} - \lambda_i\right| > \epsilon\right] \to 0 \tag{4}$$

and

$$P\left[\left|M_k(D_n) - \sum_{i=0}^{\infty} i^k \lambda_i\right| > \epsilon\right] \to 0. \tag{5}$$

Similarly, $\mathcal{D}'$ is k-smooth w.e.h.p. if these probabilities are exponentially small.

2.2 The Configuration Model.

It is difficult to directly examine random graphs with given degree sequences, so instead we use the *configuration model (or 'CM')* introduced by Bollobás [2]. For a degree sequence D, consider a set of n vertices and $\sum_i d_i$ *endpoints*, and assign d_i endpoints to the vertex v_i. Now choose a perfect matching of the endpoints uniformly at random, and for each pair of matched endpoints, draw an edge connecting the corresponding vertices.

This procedure generates a graph with degree sequence D; however, the graph may contain loops and/or multiple edges. We shall abuse notation and refer to such a random (multi-)graph as a *random graph with degree sequence D generated by the configuration model.* Definitions for asymptotic degree sequences generalize to the CM in the obvious way.

Under certain circumstances results about random graphs generated by the CM hold in general for random graphs with the same degree sequence [11, 12]. It is easy to see that every simple graph with degree sequence D occurs with the same probability using the CM. A result of McKay and Wormald [10] implies that if the maximum degree of a degree sequence is $o(M_1(D)^{1/3})$, then a random configuration produces a simple graph with probability $e^{-O\left(\frac{M_2(D)^2}{M_1(D)^2}\right)}$. If an asymptotic degree sequence $\mathcal{D}$ is 2-smooth, then this probability is $\Theta(1)$, and a.a.s. and w.e.h.p. results for the CM clearly generalize to random graphs in general. If $\mathcal{D}$ is 1-smooth, then a result which holds with probability $1 - e^{-\omega(M_2(D)^2)}$ using the CM implies an a.a.s. result for general random graphs. Furthermore, if $M_2(D) = O(N^\epsilon)$, then a result which holds with probability $1 - e^{-\Omega(n^{\epsilon+\epsilon'})}$ using the CM implies a w.e.h.p. result for general random graphs.

3 Random Graphs and Branching Processes.

3.1 A Theorem on Branching Processes.

A *branching tree* based on a probability distribution $\{\mu_i\}$ on the non-negative integers is a recursively defined random tree, in which the degree of the root vertex is distributed according to $\{\mu_i\}$, and each child is the root is an independent branching tree based on the same distribution. A *branching process* is a random process $X_0, X_1, X_2, \ldots$, where X_i counts the number of vertices at depth i in a branching tree. A branching tree can also be referred to as the *geneology tree* of the corresponding branching process.

In this section we answer the question of when a branching process generates a infinite complete k-ary tree with positive probability. In the next section, following Pittel, Wormald, and Spencer [14], we shall intuitively argue that a random graph with a fixed degree sequence locally resembles a branching tree, and that the presence of a giant $(k+1)$-core in a random graph is related to the possibility that a branching tree contains an infinite k-ary subtree (note that this result is tangential to results for fixed-degree random graphs, which are obtained through different methods).

Given a probability distribution $\{\mu_i\}$, the *probability generating function (p.g.f.)* [7, 8] for $\{\mu_i\}$ is defined as $g(q) = \sum_{i=0}^{\infty} \mu_i q^i$. The p.g.f. is a central tool in the theory of branching processes [7]. In particular, a classical result states that extinction probability of a branching process (that is, the probability that $X_i = 0$ for all but finitely many i) is given by the smallest fixed point of g in $[0, 1]$.

Now, for each integer $r \geq 0$, define the function

$$f_r(q) = \sum_{i=0}^{r} \frac{(1-q)^i}{i!} g^{(i)}(q), \tag{6}$$

where $g^{(i)}$ is the i'th derivative of the p.g.f. g. Note that $f_r(q)$ is the r'th order Taylor approximation of $g(1)$ about q, so $f_0(q) = g(q)$, $f_1(q) = g(q) + (1-q)g'(q)$, and so on.

The extinction of a branching process is in fact the complement of the event that the corresponding branching tree contains an infinite 1-ary subtree; the classical result cited above states that the probability of this event is determined by the smallest fixed point of f_0. Using similar techniques, we obtain the following generalization.

Theorem 3.1 *Let q_r be the smallest fixed point of the function f_r (as defined in equation 6) in the interval $[0,1]$. Then the probability that a branching tree based on the probability distribution $\{\mu_i\}$ contains an infinite perfect $(r+1)$-ary tree is $1-q_r$.*

Proof. Let X be a random variable with distribution $\{\mu_i\}$. For any $0 \le q \le 1$, let Z_q be a random variable with distribution

$$P(Z_q = i) = \sum_{j=i}^{\infty} \mu_j \cdot \binom{j}{i} q^{j-i}(1-q)^i.$$

Note that

$$\begin{aligned}
P(Z_q \le r) &= \sum_{i=0}^{r}\sum_{j=i}^{\infty} \mu_j \cdot \binom{j}{i} q^{j-i}(1-q)^i \\
&= \sum_{i=0}^{r} E\left[\binom{X}{i} q^{X-i}(1-q)^i\right] \\
&= \sum_{i=0}^{r} \frac{(1-q)^i}{i!} E\left[\frac{X!}{(X-i)!} q^{X-i}\right] \\
&= \sum_{i=0}^{r} \frac{(1-q)^i}{i!} g^{(i)}(q) \\
&= f_r(q).
\end{aligned}$$

Now, consider a branching tree based on distribution $\{\mu_i\}$ First, we calculate the probability that the root tree has at least $r+1$ children. Since Z_0 has the same distribution as X, then $P(X > r) = 1 - f_r(0)$. In order to produce an $(r+1)$-ary tree of depth 2, then the root must have at least $r+1$ children, each of whom produce $r+1$ grandchildren. Each child has probability $1 - f_r(0)$ of producing at least $(r+1)$ grandchildren, thus the number of such children is a random variable with distribution $Z_{f_r(0)}$. Accordingly, the probability of producing an $(r+1)$-ary tree of depth 2 is $1 - f_r(f_r(0))$.

In general, producing an $(r+1)$-ary tree of depth d is equivalent to having at least $(r+1)$ children who produce $(r+1)$-ary trees of depth $d-1$. Thus, we inductively conclude that the probability of an $(r+1)$-ary tree of depth d is $1 - f_r^{[d]}(0)$, where $f_r^{[d]}$ is the d'th iterate of f_r. Since $f_r(q)' \ge 0$ in the interval $[0,1]$ and $f_r(1) = 1$, then $f_r^{[d]}(0)$ approaches the lowest fixed point of f_r as $d \to \infty$. ∎

3.2 Relation to the k-core of a Random Graph.

Let $\mathcal{D}$ be a 1-smooth asymptotic degree sequence with limiting degree distribution $\{\lambda_i\}$. We define the *residual degree distribution* $\{\mu_j\}$ of $\mathcal{D}$ by

$$\mu_j = \frac{(j+1)\lambda_{j+1}}{\sum_{i=1}^{\infty} i\lambda_i}. \tag{7}$$

Since $\mathcal{D}$ is 1-smooth, $\sum_i i\lambda_i$ converges, and the residual degree distribution is well defined.

Pittel, Spencer, and Wormald [14] noted, using the graph model $\mathcal{G}_{n,m}$, that a giant k-core in a random graph relates to the probability of finding an infinite $(k-1)$-ary subtree of a Poisson branching tree. For an arbitrary degree sequence D, the presence of a giant k-core relates the probability of finding an infinite $(k-1)$-ary subtree in a branching tree based on the residual distribution $\{\mu_i\}$. However, in both cases, the argument is incomplete and it is not clear that the link to the branching process can be made sufficiently rigorous to produce a simple proof of the k-core result.

Consider a single endpoint s in the random graph $G(D)$ generated by the CM, and let us examine the structure of $G(D)$ in a neighborhood of s. According to the CM, s is matched to an endpoint t whcih is chosen uniformly from the set of all endpoints (other than s).

If s is matched to an endpoint t which is assigned to a vertex v, we define the *residual degree* of s to be a the number of endpoints assigned to v other then t; hence the residual degree of s is one less than the degree of v. Since, asymptotically, the fraction of vertices in $G(D)$ with degree i is λ_i, and vertex v of degree i has i chances of matching to s, we deduce that the residual degree of s is a random variable distributed according to the residual distribution

$$\left\{\mu_i = \frac{(i+1)\lambda_{i+1}}{\sum_i \lambda_i}\right\}.$$

Next, consider the all of the endpoints assigned to v other than t. By a similar informal argument, the residual degrees of these endpoints will be almost independent, and almost identically distributed. If proceed to examine larger neighborhoods of s, and if we ignore

1. the possibility of small cycles around s, and

2. slight changes in the effective residual distribution caused by the fact that the same endpoint cannot be matched twice (i.e. we are sampling without replacment),

then the graph in a small neighborhood of s will have the structure of a branching tree based on the residual distribution. Eventually, the factors we have ignored will become significant, and even in small neighborhoods, we have not precisely quantified the extent to this analogy is valid. Nevertheless, it is heuristically useful to imagine (or hope) that, the graph $G(D)$ locally resembles a branching tree.

We now consider the question of a giant k-core in the random graph $G(D)$. Pittel, Spencer, and Wormald [14] noted, using the graph model $\mathcal{G}_{n,m}$, that a giant k-core in a random graph relates to the probability of finding an infinite $(k-1)$-ary subtree of a Poisson branching tree. The following informal argument is taken from [14].

Choose any vertex v in $G(D)$, and let us attempt to determine whether or not v is in the k-core of $G(D)$. Clearly, v must have degree at least k to be part of the k-core. Furthermore, v must have at least k endpoints each of whom have residual degree at least $k-1$, and these $k-1$ neighbors must in turn have k endpoints of residual degree at least $k-1$ and so on. If we assume that residual degrees are i.i.d. random variables, then in order for v to be in the k-core of $G(D)$, v must have k endpoints which generate branching trees containing a complete $(k-1)$-ary tree.

Of course, this argument is incomplete. As pointed out in [14], it is not clear that the link to the branching process can be made entirely rigorous; in particular, we have only argued that producing a complete $(k-1)$-ary branching tree with positive probability is necessary for a giant k-core. We are not aware of an equally simple argument that this condition is sufficient. Further, the assumption that the residual degrees are i.i.d. random variables is not accurate. Thus, the branching process argument should be treated as an intuitive explanation or perhaps as a guess at the true solution.

If we now use the μ_i as defined in equation 7, the informal connection between the existence of a k-core in G and the existence of an infinite $(k-1)$-ary subtree in the branching process gives an informal justification of the following Theorem 3.2 from [6].

Theorem 3.2 *[6] Let $\mathcal{D}$ be a 1-smooth asymptotic degree sequence with maximum degree in D_n being $o(n^{1/3})$ and with residual degree distribution $\{\mu_i\}$. Then*

1. *If there exists a value q in the interval $[0,1)$ such that $f_{k-2}(q) < q$ then there exists a constant $C > 0$ such that the k-core of $G(\mathcal{D})$ contains at least Cn vertices w.e.h.p.*

2. *If $f_{k-2}(q) > q$ for all $q \in [0,1)$ then for every $C > 0$, then the k-core of $G(\mathcal{D})$ has less than Cn vertices w.e.h.p.*

The formal proof of Theorem 3.2 in [6] (and in the appendix for reference) uses an algorithm adapted from [14] which, at each time step, removes an edge incident on a vertex of degree less than k, and continues until there are no longer any vertices of nonzero degree less than k. The remaining edges and vertices will be the (possibly empty) k-core of the original graph. This algorithm is incorporated within the CM in a natural way. In particular, the algorithm in [6] chooses the random matching used by the CM while the algorithm executes, exposing edges only as they are needed. When the algorithm terminates, the k-core will remain unexposed, and thus a corollary to Theorem 3.2 is that the k-core of a random graph with asymptotic degree sequence $\mathcal{D}$ is itself a random graph with a different 1-smooth asymptotic degree sequence and a limiting degree distribution which can be calculated from the limiting distribution of $\mathcal{D}$. We state this as a corollary, which we will use in section 5.

Corollary 3.3 *Let $\mathcal{D}$ be a 1-smooth asymptotic degree sequence with maximum degree in D_n being $o(n^{1/3})$. Then the giant k-core of $G(\mathcal{D})$, if it exists, is $G(\mathcal{D}')$, where $\mathcal{D}'$ is a 1-smooth asymptotic degree sequence a.a.s. If $\mathcal{D}$ is 2-smooth, then $\mathcal{D}'$ is a.a.s. 2-smooth as well.*

4 k-cores in $\mathcal{G}_{n,p}$ and Random Power-law Graphs

Using Theorem 3.2 and the results in [11, 12] for the a.a.s. presence of a giant component it is not difficult to determine that a random graph G with a 1-smooth degree sequence has a giant 2-core a.a.s. if and only if it has a giant component a.a.s. (see end of Appendix).

For $k > 2$ the conditions necessary for a giant k-core are less easily verified, since it is not necessarily true that f_{k-2} will have all positive derivatives. In this section, we consider the case where all of the moments of the distribution $\{\mu_i\}$ are convergent, and the case of power-law graphs.

4.1 Distributions with All Convergent Moments.

Let X be a random variable with distribution $\{\mu_i\}$. By assumption, $g^{(i)}(1) = E[X(X-1)\cdots(X-i+1)] = \nu_i$, the i'th factorial moment of the distribution $\{\mu_i\}$, is finite for all i. This allows us to write

$$g(q) = \sum_{i=0}^{\infty} \frac{(q-1)^i}{i!} g^{(i)}(1) = \sum_{i=0}^{\infty} \frac{(q-1)^i}{i!} \nu_i.$$

We can now express f_r as a power series in $(q-1)$

$$\begin{aligned} f_r(q) &= \sum_{i=0}^{r} \frac{(1-q)^i}{i!} g^{(i)}(q) \\ &= \sum_{i=0}^{r} \frac{(1-q)^i}{i!} \sum_{j=i}^{\infty} \frac{j!}{j-i!} \frac{(q-1)^{j-i}}{j!} \nu_j \\ &= \sum_{j=0}^{\infty} \frac{(q-1)^j}{j!} \nu_j \sum_{i=0}^{r} (-1)^i \binom{j}{i} \\ &= 1 + \sum_{j=r+1}^{\infty} \frac{(q-1)^j}{j!} \nu_j (-1)^r \binom{j-1}{r}, \end{aligned}$$

where the last step uses the binomial identity $\sum_{i=0}^{r}(-1)^i\binom{j}{i} = (-1)^r\binom{j-1}{r}$ for $j > 0$.

Now we write $p = 1-q$, and note that and finding a fixed point $f_r(1-p) = 1-p$ is equivalent to solving

$$\begin{aligned} 1-p &= 1 + (-1)^r \sum_{j=r+1}^{\infty} \frac{(-p)^j}{j!} \nu_j \binom{j-1}{r} \\ 0 &= p + (-1)^r \sum_{j=r+1}^{\infty} p^j \frac{(-1)^j}{j!} \nu_j \binom{j-1}{r}. \end{aligned}$$

In order to ascertain the presence of a giant k-core, we must find a point whe re $f_{k-2}(q) < q$, or

$$p + (-1)^k \sum_{j=k-1}^{\infty} p^j \frac{(-1)^j}{j!} \nu_j \binom{j-1}{k-2} < 0. \qquad \text{(equation)}$$

4.1.1 Application to $\mathcal{G}_{n,p}$.

As shown by Molloy and Reed [11], the Erdos-Renyi random graph model $\mathcal{G}_{n,p}$ produces a random graph with a Poisson degree distribution, and thus results derived for random graphs whose limiting degree distribution is a Poisson distribution are valid for $\mathcal{G}_{n,p}$. Since a Poisson distribution has all convergent moments, we can re-derive some of the results of Pittel, Spencer, and Wormald [14] regarding the k-core of $\mathcal{G}_{n,m}$.

Consider a random graph whose limiting degree distribution is a Poisson distribution with expected value r, so $\lambda_i = \frac{r^i e^{-r}}{i!}$. Since

$$i\lambda_i = r\frac{r^{i-1}e^{-r}}{(i-1)!},$$

and $\sum i\lambda_i = r$, then

$$\mu_i = \frac{(i+1)\lambda_{i+1}}{\sum_i \lambda_i} = \lambda_i,$$

the residual degree distribution is identical to the limiting degree distribution.

Now, the factorial moments of a Poisson distribution are $\nu_j = r^j$. Using equation 4.1 from the previous discussion, if

$$p + (-1)^k \sum_{j=k-1}^{\infty} (pr)^j \frac{(-1)^j}{j!} \binom{j-1}{k-2} < 0$$

has a solution, then $\mathcal{G}_{n,p}$ with expected degree r has a giant k-core w.e.h.p. Let

$$C_k(x) = (-1)^{k+1} \sum_{j=k-1}^{\infty} x^j \frac{(-1)^j}{j!} \binom{j-1}{k-2},$$

and let $x = pr$. Then we must solve

$$\begin{aligned} x/r - C_k(x) &< 0 \\ x/C_k(x) &< r. \end{aligned}$$

Thus the giant k-core threshold for $\mathcal{G}_{n,p}$ occurs at

$$\min \frac{x}{C_k(x)}.$$

This is exactly the threshold derived in [14].

4.2 Power Law Graphs

Several massive graphs that occur in the real-world, including the web graph, have degree sequences that obey a power law [9], thus there has been considerable interest in understanding the properties of massive power law graphs. One approach to studying such graphs, introduced by Aiello, Chung, and Lu [1], is to generate random graphs with power law degree sequences.

A degree sequence obeys a power law if the number of vertices of degree i is proportional to $i^{-\beta}$ for some β. If $\beta \leq 2$, this degree sequence is not sparse, but for $\beta > 2$, this power law graph can be characterized by a 1-smooth asymptotic degree sequence with

$$\lambda_i = \frac{1}{\zeta(\beta)} \frac{1}{i^\beta},$$

where $\zeta(\beta) = \sum_{i=1}^{\infty} i^{-\beta}$ is the Riemann Zeta function. The corresponding residual endpoint distribution is

$$\mu_i = \frac{1}{\zeta(\beta-1)} \frac{1}{(i+1)^{\beta-1}}.$$

Since the number of vertices of degree i is approximately $n\lambda_i$, and $\lambda_i = \Theta(i^{-\beta})$, it might be natural to consider the largest degree in a random power law graph to be $\Theta(n^{1/\beta})$. This is the assumption made by [1]. The configuration model and Theorem 3.2 require the maximum degree to be $o(n^{1/3})$, and hence for $\beta < 3$, this power-law graph model would violate the CM maximum degree requirement. However, we may extend our results to the maximum degree bound in the power law model of [1] by the following mechanism. There are $O(n^{1-\delta})$ edges incident on vertices of degree greater than $n^{1/3-\epsilon}$ for some $\delta > 0$ when $\beta > 2$. Consider exposing the edges of such a graph in the CM by first placing the edges of these high degree vertices in any (adversarial) manner while respecting the degree constraints on the other vertices. Then apply the CM mechanism to the graph represented by the residual degrees of the remaining vertices. Since only $O(n^{1-\delta})$ endpoints are affected, the resulting degree distribution is indistinguishable from the starting power-law distribution with respect to smoothness, hence the result in [10] continues to hold for this residual graph.

We must also verify that $M_2(D_n) = O(n^{1/2-\epsilon})$. Using the power law model from [1], we find that

$$M_2(D_n) = \sum_{i=0}^{n} \frac{d_i^2}{n} \simeq \sum_{i=0}^{d_{\max}} \Theta\left(\frac{i^2}{i^\beta}\right) \simeq \Theta(d_{\max}^{3-\beta})$$

If $d_{\max} = n^{1/\beta}$, then $M_2(D_n) = n^{\frac{3}{\beta}-1}$, which is $O(n^{1/2-\epsilon})$ for $\beta > 2$.

By Theorem A.2 in the Appendix, the a.a.s existence of a giant 2-core in a random power-law graph is equivalent to the a.a.s. existence of a giant component, which appears if $\beta < 3.47875\ldots$ [1]. For $k \geq 3$, we have the following theorem.

Theorem 4.1 *Let $k \geq 3$ be an integer constant.*

1. *For $\beta \geq 3$, a random power law graph does not have a giant k-core w.e.h.p.*
2. *For $2 < \beta < 3$, a random power law graph has a giant k-core w.e.h.p.*

Proof. Since $f_r(q) = \sum_{i=0}^{r} \frac{(1-q)^i}{i!} g^{(i)}(q)$, we derive

$$\begin{aligned} {f'}_r(q) &= \sum_{i=0}^{r} \frac{(1-q)^i}{i!} g^{(i+1)}(q) - \sum_{i=1}^{r} \frac{(1-q)^{(i-1)}}{(i-1)!} g^{(i)}(q) \\ &= \frac{(1-q)^r}{r!} g^{(r+1)}(q). \\ &= \frac{1}{r!} \frac{g^{(r+1)}(q)}{\sum_{i=0}^{\infty} \binom{i+r-1}{r-1} q^i}. \end{aligned}$$

For a power law graph, the probability generating function of the endpoint distribution is given by

$$g(q) = \frac{1}{\zeta(\beta-1)} \sum_{i=0}^{\infty} \frac{q^i}{(i+1)^{\beta-1}},$$

and thus

$$g^{(r+1)}(q) = \frac{1}{\zeta(\beta-1)} \sum_{i=0}^{\infty} \frac{i(i-1)(i-2)\cdots(i-r)q^{i-r-1}}{(i+1)^{\beta-1}}.$$

Thus, for power law graphs,

$$\begin{aligned} {f'}_r(q) &= \frac{1}{r!} \frac{g^{(r+1)}(q)}{\sum_{i=0}^{\infty} \binom{i+r-1}{r-1} q^i} \\ &= \frac{1}{r!\zeta(\beta-1)} \frac{\sum_{i=0}^{\infty} \frac{i(i-1)(i-2)\cdots(i-r)q^{i-r-1}}{(i+1)^{\beta-1}}}{\sum_{i=0}^{\infty} \binom{i+r-1}{r-1} q^i} \\ &= \frac{1}{r!\zeta(\beta-1)} \frac{\sum_{i=r+1}^{\infty} \frac{i(i-1)(i-2)\cdots(i-r)q^{i-r-1}}{(i+1)^{\beta-1}}}{\sum_{i=r+1}^{\infty} \binom{i-2}{r-1} q^{i-r-1}}. \end{aligned}$$

Now, let $a_i = \frac{i(i-1)(i-2)\cdots(i-r)}{(i+1)^{\beta-1}}$ and let $b_i = r!\binom{i-2}{r-1}$, so

$${f'}_r(q) = \frac{1}{\zeta(\beta-1)} \frac{\sum_{i=r+1}^{\infty} a_i q^{i-r-1}}{\sum_{i=r+1}^{\infty} b_i q^{i-r-1}},$$

and note that

$$\frac{a_i}{b_i} = \frac{\frac{i(i-1)(i-2)\cdots(i-r)}{(i+1)^{\beta-1}}}{r!\binom{i-2}{r-1}} = \frac{i(i-1)}{r(i+1)^{\beta-1}}.$$

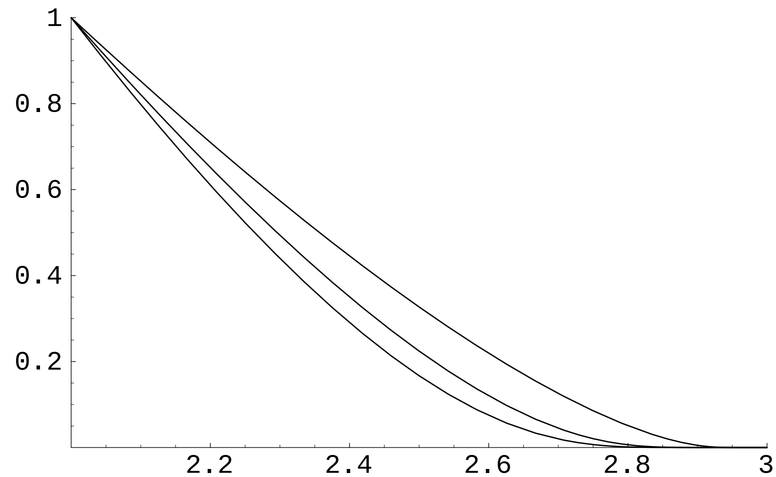

Figure 1: *Plot of the size of k-core as a function of β, for $2 < \beta < 3$ and for $k = 3, 4, 5$:* the y-axis is the fraction of endpoints in a power law graph which belong to the k-core, and the x-axis is β. Since the 3-core is the largest, the top curve corresponds to the 3-core, and the bottom curve corresponds to $k = 5$.

In particular, if $\beta \geq 3$, then $a_i > b_1$ for all i, and if $\beta < 3$, then $\lim_{i \to \infty} \frac{a_i}{b_i} = \infty$.

Hence,

$$\lim_{q \to 1} f_r'(q) = \begin{cases} \infty & \text{if } \beta < 3 \\ \frac{1}{r\zeta(\beta - 1)} & \text{if } \beta = 3 \\ 0 & \text{if } \beta > 3. \end{cases} \quad (8)$$

It follows that if $\beta \geq 3$, then ${f'}_r(q) \leq \frac{1}{r\zeta(\beta-1)} < 1$ for all $0 \leq q \leq 1$, and since $f_r(1) = 1$, then $f_r(q) > q$ for all q in the interval $[0, 1)$. For $\beta < 3$, the limit in equation 8 implies that $\lim_{q \to 1} {f'}_r(q) = \infty$, and thus $f_r(1 - \delta) < 1 - \delta$ for δ sufficiently small.

To determine the presence of a giant k-core, we examine the function f_{k-2}. Hence, for $k > 3$, a power law graph with $2 \leq \beta < 3$ a.a.s. has a giant k-core, and a power law graph with $\beta > 3$ a.a.s. does not have a giant k-core. $\blacksquare$

For $2 < \beta < 3$, we can also calculate the size of the k-core of a random power law graph. Figure 1 plots the asymptotic value of $1 - q_{k-2}$, where q_{k-2} is the lowest fixed point of the function f_{k-2} for the power law distribution, as a function of β, for $k = 3$, 4, and 5. The value of $1 - q_{k-2}$ measures the fraction of endpoints which belong to the k-core of a random power law graph.

5 k-Connectivity in Sparse Random Graphs

If a graph does not contain a giant k-core then it does not contain a giant k-edge connected subgraph or a giant k-vertex connected subgraph since any k-edge connected graph must have minimum degree k. We now show that the giant k-core for $k \geq 3$, when it exists, is almost surely k-vertex connected if $\mathcal{D}$ is 2-smooth.

Theorem 5.1 *Let $\mathcal{D}$ be an asymptotic degree sequence that satisfies the requirements of Theorem 3.2 for the existence of a giant k-core for some $k \geq 3$, and let $\mathcal{D}$ be 2-smooth. Then a.a.s. the k-core of $G(\mathcal{D})$ is k-vertex connected.*

Proof. From the corollary to Theorem 3.2 we know that the giant k-core of $G(\mathcal{D})$ is a random graph on $n' = \Theta(n)$ vertices with an asymptotic degree sequence $\mathcal{D}'$ which is a.a.s. 2-smooth. Thus if the CM is used to generate a random graph from $\mathcal{D}'$, it will will generate a simple graph with probability $\Theta(1)$.

Now, consider the probability of generating a $k-1$ separating set of vertices. Such a separating set S must partition the remaining $n' - k + 1$ vertices into two sets X and Y. Let the smaller set X contain $r \leq (n' + k - 1)/2$ vertices. For each such r, the number of ways of choosing the sets X and S is

$$\binom{n'}{r}\binom{n'-r}{k-1} \leq (en'/r)^r \cdot n'^{k-1} \tag{9}$$

For simplicity of exposition, we shall abuse notation and identify the vertex sets X, Y, and S with the sets of endpoints corresponding to respective vertices; hence an endpoint is "in" S if it corresponds to a vertex in S.

Since the maximum degree of is $d = o(n^{1/3})$, the number of endpoints in S is at most $(k-1)d = o(n^{1/3}) = o(n'^{1/3})$. In order for S to be a $(k-1)$-separating set, all of the endpoints in X must match within $X \cup S$. Let s denote the number of endpoints in X, so each endpoint has probability at most $(s + o(n'^{1/3}))/(n'k)$, of matching into $X \cup S$.

The probability that all endpoints in X match within $X \cup S$ can be bounded by exposing the match of each endpoint in X according to the CM. At each step, we choose an endpoint from X and find its match uniformly at random. Clearly, at least $s/2$ steps must occur before all endpoints in X are matched; also, each newly chosen endpoint has a probability of at most $(s + o(n^{1/3}))/(n'k)$ of matching into $X \cup S$, assuming all of the previous endpoints matched into $X \cup S$. Hence the probability of that all of the endpoints in X match into $X \cup S$ is bounded from above by $\left(\frac{s+(k-1)o(n'^{1/3})}{n'k}\right)^{s/2}$, which is maximized when s achieves its minimum value, which is rk. Hence the probability of finding a $(k-1)$ separator for a set of size r is at most

$$(en'/r)^r \cdot n'^{k-1} \cdot \left(\frac{rk + o(n'^{1/3})}{n'k}\right)^{rk/2}. \tag{10}$$

Note that the expression in equation 10 is exponentially small when $r = \Omega(n'^{\epsilon})$. Hence, we turn our attential to small values of r.

For small r our approach is slightly different. Consider a set X of r vertices, where $r = n'^{o(1)}$, and assume X contains kr endpoints, since adding endpoints will only lower the probability that X is $(k-1)$-separated from the rest of the graph. Now, in order for X to be $(k-1)$-separated, each endpoints in X must match either within X or into a set S of size at most $k-1$. We write $r = (a+b)/k$, and compute the probability that a endpoints match within X and b endpoints match into a set S with size at most $k-1$.

First, the probability that a given endpoint matches within X is $O(r/n')$, so the probability that a given set of $a/2$ endpoints matches within X is $O((r/n)^{a/2})$. Finally, the number ways of choosing a such endpoints is $\binom{r}{a} = O(r^a)$. Hence, the probability of finding $a/2$ internal edges within X is $O(r^a(r/n')^{a/2}) = n'^{-a/2+o(a)}$.

We now compute the probability that the remaining b endpoints match into the same $(k-1)$ vertex set S. First, note that the probability that two uniformly chosen endpoints will belong to the same vertex is

$$\frac{\sum_{i=1}^{n'}\binom{d'_i}{2}}{\binom{m'}{2}} = O\left(\frac{nM_2(D')}{(nM_1(D'))^2}\right)$$

$$= O\left(\frac{M_2(D')}{nM_1(D')^2}\right).$$

Similarly, the probability that any constant number i of uniformly chosen endpoints belong to the same vertex is

$$O\left(\frac{M_i(D')}{n^{i-1}M_1(D')^i}\right),$$

where $M_i(D')$ is the i'th moment of the degree distribution of D'.

The property of 2-smoothness does not guarantee that $M_i(D')$ converges to a constant for arbitrary i. However, the maximum degree satisfies $d_{\max} = o(n^{1/3})$, and hence, for any i, $M_{i+1}(D') = o(n^{1/3}M_i(D'))$, so by 1-smoothness $M_i(D') = o({n'}^{(i-1)/3})$. Hence the probability of choosing i endpoints from the same vertex is $o(n^{-2(i-1)/3})$.

Next, we compute the probability that b endpoints match into at most as set S containing $k-1$ vertices. Note that he probability that b_j endpoints match to the i'th vertex in S (for $j = 1, 2, \ldots, k-1$ and $\sum_j b_j = b$) is

$$\prod_{j=1}^{k-1} o(n^{-2(b_j-1)/3}) = o(n^{-2(b-(k-1)/3}).$$

There are $O(b^{k-1})$ ways of choosing the values b_j, and since $b = O(r) = {n'}^{o(1)}$ and k is constant, the overall probability of matching b endpoints into a set of $k-1$ vertices is

$$o({n'}^{-2(b-k+1)/3+o(1)}).$$

Hence, the total probability of finding a vertex set of size r, with a internally matched endpoints, and b endpoints matched to a set S containing $k-1$ vertices is at most

$$o({n'}^{r+o(a)-a/2-2(b-k+1)/3+o(1)}) = o({n'}^{r+o(a)-a/2-2(b-k+1)/3}). \tag{11}$$

Examining the exponent in equation 11,

$$\begin{aligned} r + o(a) - a/2 - 2(b-k+1)/3 &= o(a) + (a+b)/k - a/2 - 2(b-k+1)/3 \\ &= (o(1) + 1/k - 1/2)a + (1/k - 2/3)b + 2(k-1)/3, \end{aligned}$$

we find that the largest value occurs when a is maximal and b is minimal.

Note that if $b \le k-1$ then this probability is simply

$$O(n^{r+o(a)-a/2}),$$

since in this situation, the endpoints matching outside of X will always match to at most $k-1$ vertices. Thus we consider this case separately. Since $b \le k-1$, then we may assume $b = k-1$, as this minimizes the number of endpoints which must match internally to X. Hence $a = rk - k + 1$ and since $a \le 2\binom{r}{2}$, we have $r(r-1) \ge k(r-1)+1$, and thus $r > k$. Then

$$O(n^{r+o(a)-a/2}) = O(n^{r-((r-1)k-1)(1/2-o(1))}),$$

and since $k \ge 3$ and $r \ge k+1 \ge 4$, this yields $O(n^{o(1)-1})$.

Otherwise, $b \geq k$. As noted above, the probability of finding a $(k-1)$-separating set S is largest when a is maximized. Thus we set $a = r(r-1)$ and $b = rk - r(r-1)$, and we note that

$$\begin{aligned} rk - r(r-1) &= b \geq k \\ k(r-1) &\geq r(r-1) \\ k &\geq r. \end{aligned}$$

In particular, $r = O(1)$, hence $n'/r = \Theta(n')$, and we may drop the $o(a)$ term from the exponent of equation 11, and compute:

$$\begin{aligned} r - a/2 - 2(b-k+1)/3 &= r - \frac{r(r-1)}{2} + \frac{2}{3}(k-1-b) \\ &= -\frac{r^2}{2} + \frac{3r}{2} - \frac{2}{3}(k-1-b). \end{aligned}$$

For $r = 2$, we have $b = kr - 2 = 2k - 2$, so for $k \geq 3$, the probability that X is $(k-1)$ separated is at most

$$O(n^{-2+3-\frac{2}{3}(1-k)}) = O(n^{1-\frac{4}{3}}) = o(1).$$

For $r \geq 3$, this probability is

$$O(n^{-\frac{r(r-3)}{2}-\frac{2}{3}(k-1-b)}) = O(n^{-2/3}) = o(1)$$

as well.

Hence, for $r = n^{\Omega(1)}$, the probability of finding an r vertex set with a $(k-1)$ separator is exponentially small. For $k < r = n^{o(1)}$, the probability is $n^{o(1)-1}$, and using Boole's inequality the probabilty of finding a $(k-1)$ separator for any such r is also $n^{o(1)-1}$. Finally, for $r \leq k$, the probability is $o(1)$. This concludes the proof, as we have shown that the probability of choosing sets X, S, in the k-core, where X is any size and $|S| = k-1$, such that all of the endpoints in X match into $X \cup S$, is $o(1)$. ▮

Corollary 5.2 *For $np = \Theta(1)$, $\mathcal{G}_{n,p}$ has a sharp threshold for a giant k-vertex-connected or giant k-edge-connected subgraph for $k \geq 3$, and the threshold and the giant subgraph are identical to that for k-core.*

For a degree sequence where the second moment is not bounded, our use of the CM requires us to have an exponentially small probability ($e^{-\omega(M_2(D)^2)}$) of a separating set occurring, which we do not obtain with the above method. However, we can adapt the above method to argue that any constant-size separating k-set in $G(\mathcal{D})$, where D is 1-smooth, must separate a component that is not giant. More specifically, Theorem 4.1 and equation 10 yield the following lemma.

Lemma 5.3 *Let $\mathcal{D}$ be the power-law degree distribution with maximum degree $d = o(n^{1/3})$.*

(i) If $\beta \geq 3$ then w.e.h.p. $G(\mathcal{D})$ does not have a giant k-edge connected or k-vertex connected component, for $k \geq 3$.

(ii) If $2 < \beta < 3$, then for any constant $k \geq 3$, any separating $(k-1)$-set in the k-core of $G(\mathcal{D})$ must separate a component with $O(n^{2(3-\beta)/\beta})$ vertices w.e.h.p.

Proof. Clearly, a giant k-connected component cannot be present without a giant k-core. For the second statement, equation 10 with $r = n^\epsilon$ for $\epsilon < 1/3$ yields a bound of

$$\begin{aligned} (en'/r)^r \cdot n'^{k-1} \cdot \left(\frac{rk + o(n'^{1/3})}{n'k}\right)^{rk/2} &= o\left(n^{(1-\epsilon)n^\epsilon} n^{k-1} n^{-kn^\epsilon/3}\right) \\ &= e^{-\Omega(n^\epsilon)}, \end{aligned}$$

on the probability of finding a vertex set of size n^ϵ which is seperated by $k-1$ vertices. For $\epsilon \geq 1/3$, we similarly derive a bound of

$$O\left(n^{(1-\epsilon)n^\epsilon+(3/2)(\epsilon-1)n^\epsilon}\right) \quad = \quad e^{\Omega(n^\epsilon)}.$$

Since $M_2(D) = n^{\Theta(3/\beta-1)} = n^{\Theta(*3-\beta)/\beta-1)}$ for power law graphs with $2 < \beta < 3$. Since a random configuration produces a simple graph with probability $e^{-O\left(\frac{M_2(D)^2}{M_1(D)^2}\right)}$, choosing $\epsilon = \omega(2(3-\beta)/\beta)$ guarantees that a vertex set of size n^ϵ with a $(k-1)$ separator does not occur in a random power law graph w.e.h.p. ∎

Acknowledgement. We thank Joel Spencer for helpful comments on our k-core manuscript.

References

[1] W. Aiello, F. Chung, and L. Lu, A random graph model for massive graphs, *ACM Symposium on Theory of Computing*, 171-180, (2000).

[2] B. Bollobás, A probabilistic proof of an asymptotic formula for the number of labelled regular graphs, *Europ. J. Combinatorics* 1:311-316, (1980).

[3] B. Bollobás, *Random Graphs,* Academic Press, London, (1985).

[4] C. Cooper, The cores of random hypergraphs with a given degree sequence, manuscript, August 2003.

[5] Paul Erdos and Alfred Renyi. On random graphs. *Publ. Math. Debrecen*, 6:290–297, 1959.

[6] D. Fernholz and V. Ramachandran, The giant k-core of a random graph with a specified degree sequence, manuscript, UT-Austin, November 2003.

[7] T. E. Harris, *The Theory of Branching Processes*, Springer-Verlag, Berlin, (1963).

[8] O. Kallenberg, *Foundations of Modern Probability*, Springer-Verlag, New York, (2002).

[9] J. Kleinberg, R. Kumar, P. Raghavan, S. Rajagopalan, and A. Tomkins, The Web as a graph: measurements, models, and methods, *International Conference on Combinatorics and Computing*, 26-28, (1999).

[10] B. D. McKay and N. C. Wormald, Asymptotic Enumeration by Degree Sequence of Graphs with Degrees $o(n^{1/2})$, *Combinatorica*, 11(4):269-382 (1991).

[11] M. Molloy and B, Reed, A critical point for random graphs with given degree sequence, *Random Structures and Algorithms*, 6:161-179, (1995).

[12] M. Molloy and B. Reed, The size of the giant component of a random graph with given degree sequence, *Combin. Probab. Comput.*, 7:295-305, (1998).

[13] M. Molloy, The pure literal rule threshold and cores in random hypergraphs. *ACM-SIAM SODA*, 665-674, 2004.

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

APPENDIX

A The k-core Theorem

In this section we present the result in our unpublished manuscript [6] that provides a proof of Theorem 3.2 for the existence of a giant k-core. Since we shall work with CM, we cannot guarantee that a given configuration will produce a simple graph. Nevertheless, the k-core of a multigraph is well-defined as the maximal induced multigraph with minimum degree k.

Our approach is as follows. First, we describe the *CM k-core algorithm*, which is a variant of an algorithm from [14], and which finds the k-core of a (multi-)graph. Then, we determine the outcome of this algorithm w.e.h.p. under certain conditions. Finally, we conclude that if an asymptotic degree sequence $\mathcal{D}$ is such that the CM produces a simple graph with probability $\Theta(1)$, then our results will be valid for random simple graphs with degree sequence $\mathcal{D}$.

In order to find the k-core of a (multi-)graph, the algorithm in [14] chooses a vertex v of degree less than k, removes all edges incident on v and repeats this procedure until there are no more vertices with degree less than k. The remaining edges and vertices will be the (possibly empty) k-core of the original graph.

We adapt the above algorithm to the CM in a natural way. In particular, we shall choose the random matching used by the CM while the algorithm executes, exposing edges only as they are needed. When the algorithm terminates, the k-core will remain unexposed, and thus a corollary to Theorem 3.2 is that the k-core of a random graph with asymptotic degree sequence $\mathcal{D}$ is itself a random graph with a different asymptotic degree sequence and a limiting degree distribution which can be calculated from the limiting distribution of $\mathcal{D}$.

We first define the following random variables to describe the execution of our algorithm. In the following, t refers to the time step. In each time step, a certain number of endpoints are exposed and matched. The k-core is found at time step t_k. For technical reasons (which will become clear later) we extend the definitions of the random processes past t_k. Note that t_k is a random variable.

- $G(t)$ is the state of the graph at time t. The algorithm begins at time 0, and $G(0)$ is a set of vertices and endpoints according to the degree sequence D.
- An endpoint in $G(t)$ is *exposed* if it is chosen by the algorithm during a step $t' \leq t$. All of the endpoints in initial state $G(0)$ are *unexposed.* (If $t \leq t_k$, an endpoint is exposed if and only if its match has been revealed.)
- The *unexposed degree* of a vertex v in $G(t)$ is the number of unexposed endpoints associated with v. The unexposed degree of an endpoint is the unexposed degree of its associated vertex.
- $X_i(t)$ denotes the number of vertices of unexposed degree i at time t.
- $U(t) = U(0) - 2t$. (If $t \leq t_k$ then $U(t)$ is the total number of unexposed endpoints at time t.)
- A vertex is *k-light* at time t if it has unexposed degree at most $k-1$, otherwise it is *k-heavy.* An endpoint is k-light or k-heavy if its associated vertex is k-light or k-heavy, respectively.
- $H(t)$ denotes the number of unexposed k-heavy endpoints; $H(t) = \sum_{i \geq k} iX_i(t)$ for all t.
- $L(t) = U(t) - H(t)$. (t_k is the first time step in which $L(t_k)$ becomes zero; if $t < t_k$ then $L(t)$ denotes the number of unexposed k-light endpoints.)

We are now ready to describe our k-core algorithm whose analysis will prove Theorem 3.2.

CM k-core Algorithm

(1) *If k-core has not been exposed, i.e., $t < t_k$,* we choose an endpoint $e_1(t)$ uniformly at random from the set of all unexposed k-light endpoints in $G(t)$. We then choose a second endpoint $e_2(t)$ uniformly at random from the set of all unexposed endpoints in $G(t)$, we match $e_1(t)$ and $e_2(t)$, and we designate the endpoints $e_1(t)$ and $e_2(t)$ as *exposed*. If $L(t) = 0$ then the k-core has been found.
(2) *If k-core has been exposed, i.e., $t \geq t_k$,* we choose $\left\lfloor \frac{H(t)}{U(t)-1} \right\rfloor$ unexposed k-heavy endpoints uniformly at random and designate them as *exposed*. Then, with probability $\frac{H(t)}{U(t)-1} - \left\lfloor \frac{H(t)}{U(t)-1} \right\rfloor$, we choose another such endpoint and expose it as well. For simplicity of calculation, these endpoints are chosen *with* replacement (if an endpoint is chosen twice in the same time step, it is only exposed once). Note that endpoints that are exposed in this part of the algorithm are not actually matched, as they are in part one; once the k-core has been found, designating endpoints as exposed is merely an accounting tool.

We now explain the intuition behind the above CM k-core algorithm. The process in part 1 is clear: for each $t < t_k$, it exposes a match for a k-light endpoint until no k-light endpoint remains. While $t < t_k$, the following two properties follow directly from the definitions:

1. $U(t+1) = U(t) - 2$.

2. Any given unexposed k-heavy endpoint remains unexposed with probability $1 - \frac{1}{U(t)-1}$.

We have extended our random processes beyond time t_k in part 2 in such a way that these two properties continue to hold (approximately). To see this, let $t_k \leq t < U(0)/2$ (note that since two endpoints are exposed at each step of part 1, the algorithm can run for at most $U(0)/2$ steps). In order to satisfy property 1, we have defined $U(t) = U(0) - 2t$ for all t.

For property 2, we would like each k-heavy endpoint to have a $\frac{1}{U(t)-1}$ probability of becoming exposed at time t. If $H(t) \leq U_t - 1$ this is easy: with probability $\frac{H(t)}{U(t)-1}$, we choose a single k-heavy endpoint uniformly at random and expose it. However, for $t \geq t_k$ it is possible that $H(t) \geq U(t)$, in which case we obviously cannot perform any action with probability $\frac{H(t)}{U(t)-1} > 1$. However, by first exposing $\left\lfloor \frac{H(t)}{U(t)-1} \right\rfloor$ k-heavy endpoints, and then, with probability $\frac{H(t)}{U(t)-1} - \left\lfloor \frac{H(t)}{U(t)-1} \right\rfloor$ exposing one additional such endpoint, each k-heavy endpoint becomes exposed with probability approximately $\frac{1}{U(t)-1}$.

A.1 Proof of Theorem 3.2.

We prove Theorem 3.2 as follows. First, we calculate the expected values of the $X_i(t)$ and $H(t)$, and then we prove that $H(t)$ concentrates around its expectation w.e.h.p. Then, if there exists a t such that $E[H(t)] - E[U(t)]$ is sufficiently large, $H(t) > U(t)$ must occur w.e.h.p., and the k-core will have been found by time t (also w.e.h.p.). Finally, we show that these conditions are equivalent to the conditions we derived informally using the branching process argument.

First, we define

$$p(t) = \left(\frac{U(t)}{U(0)} \right)^{1/2}.$$

Intuitively, $p(t)$ might be considered the approximate probability that an endpoint remains unexposed at time t. However, this is not quite correct, since once such an endpoint becomes k-light, it is much more likely to become exposed. The meaning or $p(t)$ is more clearly expressed in the following lemma.

Lemma A.1 *For $i \geq k$, and for $t = O(U(t))$,*

$$E\left[\frac{X_i(t)}{n}\right] = \sum_{j=i}^{\infty} \lambda_j \binom{j}{i} p(t)^i (1-p(t))^{j-i} \pm o(1). \qquad (12)$$

Proof. Let $P_{i,j}(t)$ denote the probability that a vertex of original degree j has exactly i unexposed endpoints at time t. Since vertices of original degree j are initially indistinguishable, $P_{i,j}$ is well defined, and

$$E[X_i(t)] = \sum_{j=i}^{\infty} X_j(0) P_{i,j}(t).$$

We will show that

$$P_{i,j}(t) = \binom{j}{i} p(t)^i (1-p(t))^{j-i} \pm o(1) \qquad \text{(equation)}$$

for every j. Then,

$$\begin{aligned} E\left[\frac{X_i(t)}{n}\right] &= \sum_{j=i}^{\infty} \frac{X_j(0)}{n} \left(\binom{j}{i} p(t)^i (1-p(t))^{j-i} \pm o(1)\right) \\ &= \sum_{j=i}^{h} \frac{X_j(0)}{n} \left(\binom{j}{i} p(t)^i (1-p(t))^{j-i} \pm o(1)\right) + \sum_{j=h}^{\infty} \frac{X_j(0)}{n} \left(\binom{j}{i} p(t)^i (1-p(t))^{j-i} \pm o(1)\right) \\ &= \sum_{j=i}^{h} (\lambda_j \pm o(1)) \left(\binom{j}{i} p(t)^i (1-p(t))^{j-i} \pm o(1)\right) + O\left(\sum_{j=h}^{\infty} \frac{X_j(0)}{n}\right), \end{aligned}$$

where the last step follows from the definition of smoothness.

Since h is constant, we have a constant number of $o(1)$ terms in the left summation, which can be consolidated. Then, using the smoothness condition again, we have

$$E\left[\frac{X_i(t)}{n}\right] = \sum_{j=i}^{h} \lambda_j \binom{j}{i} p(t)^i (1-p(t))^{j-i} \pm o(1) + O\left(\sum_{j=h}^{\infty} \lambda_j\right).$$

If we let h grow arbitrarily large, the $O\left(\sum_{j=h}^{\infty} \lambda_j\right)$ term becomes arbitrarily small. Thus, by first choosing h sufficiently large, and then choosing n sufficiently large,

$$\left| E\left[\frac{X_i(t)}{n}\right] - \sum_{j=i}^{\infty} \lambda_j \binom{j}{i} p(t)^i (1-p(t))^{j-i} \right|$$

can be made arbitrarily small, yielding equation 12.

To derive equation A.1, choose a vertex v of original degree j, choose any i of v's endpoints, and assume that these i endpoints are unexposed at time t. We shall calculate the probability that they remain unexposed at time $t+1$.

First, suppose $t < t_k$, so we are in part 1 of the algorithm. Then, at time t, a single light endpoint is picked and matched to an unexposed endpoint chosen uniformly at random from the

$U(t)-1$ remaining unexposed endpoints. Thus the probability that our set of i endpoints remain unexposed is $\frac{U(t)-1-i}{U(t)-1}$.

Next, suppose $t \geq t_k$. Let $r = \left\lfloor \frac{H(t)}{U(t)-1} \right\rfloor$ and $s = \frac{H(t)}{U(t)-1} - r$, and therefore $r + s = \frac{H(t)}{U(t)-1}$. Recall that in part 2 of the algorithm, we pick r unexposed k-heavy endpoints uniformly at random (with replacement) and designate them as exposed, and with probability s we expose an additional such endpoint. Also, note that since $H(t) \leq U(0)$ and $U(t) = \Theta(U(0))$, then $r = O(1)$.

Now, in order for every one of our given set of i endpoints to remain unexposed, none of the r uniformly chosen endpoints, nor the additional endpoint chosen with probability s can be a member of our set. We sample with replacement, so the probability of this occurring is

$$\left(\frac{H(t)-i}{H(t)}\right)^r \left(\frac{H(t)-si}{H(t)}\right).$$

If $H(t) < U(t)$ then $r = 0$, and this is simply

$$\frac{H(t)/s - i}{H(t)/s} = \frac{U(t)-1-i}{U(t)-1}.$$

Otherwise, since r and i are $O(1)$ we calculate

$$\begin{aligned}
\left(1-\frac{i}{H(t)}\right)^r \left(1-\frac{si}{H(t)}\right) &= 1 - \frac{(r+s)i}{H(t)} \pm O\left(\frac{1}{H(t)^2}\right) \\
&= \left(1 - \frac{(r+s)i}{H(t)}\right)\left(1 \pm O\left(\frac{1}{H(t)^2}\right)\right) \\
&= \left(\frac{U(t)-1-i}{U(t)-1}\right)\left(1 \pm O\left(\frac{1}{H(t)^2}\right)\right).
\end{aligned}$$

Since $H(t) \geq U(t)$, then $O(H(t)^{-2}) = O(U(t)^{-2})$, and this probability can be written

$$\left(\frac{U(t)-1-i}{U(t)-1}\right)\left(1 \pm O\left(\frac{1}{U(t)^2}\right)\right).$$

Assuming $t = O(U(t))$, the probability that the i endpoints remain unexposed through t steps is

$$\prod_{j=0}^{t-1}\left(\frac{U(j)-1-i}{U(j)-1}\right)\left(1 \pm O\left(\frac{1}{U(j)^2}\right)\right) = \prod_{j=0}^{t-1}\left(\frac{U(0)-2j-1-i}{U(0)-2j-1}\right)\left(1 \pm O\left(\frac{1}{U(t)^2}\right)\right). \tag{13}$$

Now, for any integers x, y, define the function

$$f_i(x,y) = \prod_{j=0}^{y-1}\left(\frac{x-2j-1-i}{x-2j-1}\right),$$

so the probability in equation 13 can be written

$$f_i(U(0),t)\left(1 \pm O\left(\frac{1}{U(t)^2}\right)\right).$$

Note that

$$\begin{aligned} f_i(x,y)\cdot f_i(x-1,y) &= \frac{(x-1-i)(x-2-i)(x-3-i)\cdots(x-2y-1-i)}{(x-1)(x-2)(x-3)\cdots(x-2y-1)} \\ &= \frac{(x-2y-2)\cdots(x-2y-1-i)}{(x-1)\cdots(x-i)}. \end{aligned}$$

Now, let x grow asymptotically while keeping i constant, and assume $x-2y=\Theta(x)$. Then

$$f_i(x,y)\cdot f_i(x-1,y)=\left(\frac{x-2y}{x}\right)^i(1\pm o_x(1)),$$

where $o_x(1)$ is the asymptotic "little o" with respect to $x\to\infty$. Since $f_i(x,y)=f(x-1,y)(1\pm o_x(1))$, we conclude that

$$f_i(x,y)=\left(\frac{x-2y}{x}\right)^{i/2}(1\pm o_x(1)).$$

Returning to equation 13 , since $U(0)=\Theta(n)$, $U(t)=\Theta(n)$, and i is constant, we deduce and the probability that i endpoints remain unexposed is

$$\begin{aligned} f_i(U(0),t)\left(1\pm O\left(\frac{1}{U(t)^2}\right)\right) &= \left(\frac{U(t)}{U(0)}\right)^{i/2}\pm o(1) \\ &= p(t)^i\pm o(1). \end{aligned}$$

Using inclusion-exclusion,

$$P_{i,j}=\sum_{i\le i'\le j}(-1)^{i'-i}\binom{j}{i'}p(t)^{i'}\pm o(1)=\binom{j}{i}p(t)^i(1-p(t))^{j-i}\pm o(1).$$

This establishes equation A.1 and hence the lemma is proved. ∎

Lemma A.2 *Fix $\epsilon>0$. For $t<\frac{U(0)}{2}(1-\epsilon)$,*

$$P\left(|H(t)-E[H(t)]|>\Theta(n^{2/3})\right)<e^{-\Theta(n^{1/3})}.$$

Proof. For t fixed we define the random process

$$Y(t')=E[H(t)|G(t')]$$

for $0\le t'\le t$. Y is a martingale by definition, and clearly $Y(t)=H(t)$.

We now establish bounded differences for Y. Note that

$$E[H(t)|G(t')]=\sum_{i\ge k}f(i,t')iX_i(t')$$

where $f(i,t')$ is the probability that an endpoint with unexposed degree i at time t' remains both heavy and unexposed at time t. We claim that $|f(i,t')i-f(i-1,t')(i-1)|=O(1)$.

By linearity, $f(i,t')i$ is the expected number of k-heavy endpoints at time t produced by a vertex of unexposed degree i at time t'. There are two ways an endpoint can fail to be heavy and unexposed; accordingly, let $a(i,t')$ denote the expected number of endpoints from a degree i vertex which become exposed before time t, and let $b(i,t')$ denote the probability that a vertex of degree i at time t' becomes k-light before time t. Then

$$f(i,t')i = i - a(i,t') - kb(i,t').$$

Clearly, $|a(i-1,t') - a(i,t')| \leq 1$, so

$$\begin{aligned} |f(i,t')i - f(i-1,t')(i-1)i| &\leq i-(i-1) + |a(i-1,t') - a(i,t')| + k\,|b(i-1,t') - b(i,t')| \\ &= i-(i-1)+1+k, \end{aligned}$$

and since k is constant, this is $O(1)$.

Now, $O(1)$ vertices change unexposed degree at time t', thus

$$Y(t'+1) - Y(t') = \sum_{i\geq k}(f(i,t') - f(i,t'+1))(iX_i(t')) \pm O(1),$$

always. Since $E[Y(t'+1) - Y(t')] = 0$ it follows that

$$\left|\sum_{i\geq k}(f(i,t') - f(i,t'+1))(X_i(t'))\right| = O(1).$$

Hence $|Y(t'+1) - Y(t')| = O(1)$.

By Azuma's inequality,

$$P\left(|Y(t) - E[Y(t)]| > \Theta(n^{1-\delta})\right) < e^{-\Theta(n^{2-2\delta}/t)}.$$

Since $t = O(n)$ and $Y(t) = H(t)$, the proof is complete. ∎

Corollary A.3 *For any $\epsilon > 0$ and $\delta > 0$, with probability $1 - e^{-\Theta(n^{1-\delta})}$, $|H(t) - E[H(t)]| = o(1)$ for all $1 \leq t \leq U(0)/2(1-\epsilon)$.*

Proof. This follows immediately since $U(0) = O(n)$, and $O(n)e^{-\Theta(n^{1-\delta})} = e^{-O(n^{1-\delta'})}$ for any $\delta' < \delta$. ∎

Proof of Theorem 3.2. Note that by Lemma A.1

$$\begin{aligned} E[H(t)] &= \sum_{i=k}^{\infty} E[iX_i(t)] \\ \frac{E[H(t)]}{U(0)} &= \frac{n}{U(0)}\sum_{i=k}^{\infty} i \sum_{j=i}^{\infty} \lambda_j \binom{j}{i} p(t)^i (1-p(t))^{j-i} \pm o(1) \\ &= \frac{n}{U(0)}\sum_{i=k}^{\infty}\sum_{j=i}^{\infty} j\lambda_j \frac{(j-1)!}{(i-1)!(j-i)!} p(t)^i (1-p(t))^{j-i} \pm o(1) \\ &= \frac{n}{U(0)}\sum_{j=k}^{\infty}\sum_{i=k}^{j} j\lambda_j \binom{j-1}{i-1} p(t)^i (1-p(t))^{j-i} \pm o(1). \end{aligned}$$

Letting $B(p,i,j) = \binom{j}{i} p^i (1-p)^{j-i}$, we continue

$$\begin{aligned}
\frac{E[H(t)]}{U(0)} &= \frac{n}{U(0)} \sum_{j=k}^{\infty} j\lambda_j p(t) \sum_{i=k}^{j} B(p(t), i-1, j-1) \pm o(1) \\
&= \frac{n}{U(0)} \sum_{j'=k-1}^{\infty} (j'+1)\lambda_{j'+1} p(t) \sum_{i'=k-1}^{j'} B(p(t), i, j') \pm o(1) \\
&= \frac{n}{U(0)} \sum_{j'=k-1}^{\infty} (j'+1)\lambda_{j'+1} p(t) \left(1 - \sum_{i'=0}^{k-2} B(p(t), i', j')\right) \pm o(1),
\end{aligned}$$

with the last substitution resulting from the fact that $\sum_{i'=0}^{j'} B(p(t), i', j') = 1$.

Now, the smoothness condition S2 implies that $\sum_i i\lambda_i = \frac{U(0)}{n}(1 \pm o(1))$. Thus, we substitute $\mu_j = \frac{(j+1)\lambda_{j+1}}{\sum_i i\lambda_i}$ and proceed

$$\begin{aligned}
\frac{E[H(t)]}{U(0)} &= \sum_{j'=k-1}^{\infty} \mu_{j'} p(t) \left(1 - \sum_{i'=0}^{k-2} B(p(t), i', j')\right) \pm o(1) \\
&= p(t) \left(\sum_{j'=k-1}^{\infty} \mu_{j'} - \sum_{j'=k-1}^{\infty} \sum_{i'=0}^{k-2} \mu_{j'} B(p(t), i', j') \pm o(1) \right) \\
&= p(t) \left(\left(1 - \sum_{j'=0}^{k-2} \mu_{j'}\right) - \sum_{j'=k-1}^{\infty} \sum_{i'=0}^{k-2} \mu_{j'} B(p(t), i', j') \right) \pm o(1) \\
&= p(t) \left(1 - \sum_{j'=0}^{k-2} \sum_{i'=0}^{j'} B(p(t), i', j') \mu_{j'} - \sum_{j'=k-1}^{\infty} \sum_{i'=0}^{k-2} \mu_{j'} B(p(t), i', j') \right) \pm o(1) \\
&= p(t) \left(1 - \sum_{i'=0}^{k-2} \sum_{j'=i'}^{\infty} \mu_{j'} B(p(t), i', j') \right) \pm o(1).
\end{aligned}$$

Since $p(t) = (U(t)/U(O))^{1/2}$, we write $U(t)/U(0) = p(t)^2$. Note that if $H(t) \geq U(t)$, then the k-core of G has been found, and if $H(t) < U(t)$ for all t then there is no k-core. Due to corollary A.3, $H(t)/U(0) = E[H(t)/U(0)] \pm o(1)$ w.e.h.p. Therefore, if there exists a value t such that

$$E[H(t)/U(0)] - p(t)^2 = \Theta(1),$$

then $G(D)$ has a k-core containing at least $U(t)$ endpoints w.e.h.p., and if

$$E[H(t)/U(0)] - p(t)^2 = -\Theta(1)$$

for all $t \leq (1-\epsilon)U(0)/2$, then the k-core of $G(D)$ contains less than $\epsilon U(0)$ endpoints w.e.h.p.

These conditions can be written

$$p(t) \left(1 - \sum_{i'=0}^{k-2} \sum_{j'=i'}^{\infty} \mu_{j'} B(p(t), i', j') \right) - p(t)^2 = \Theta(1) \tag{14}$$

and

$$p(t)\left(1-\sum_{i'=0}^{k-2}\sum_{j'=i'}^{\infty}\mu_{j'}B(p(t),i',j')\right)-p(t)^2=-\Theta(1). \quad (15)$$

Since $t\leq(1-\epsilon)U(0)/2$, then $p(t)=\Theta(1)$, thus we can divide the LHS of equations 14 and 15 by $p(t)$, and the RHS will remain $\pm\Theta(1)$. We set $q(t)=1-p(t)$, and, after dividing by $p(t)$, the LHS becomes

$$\begin{aligned} q(t)-\sum_{i=0}^{k-2}\sum_{j=i}^{\infty}\mu_j B(q(t),j,i) &= q(t)-\sum_{i=0}^{k-2}\sum_{j=i}^{\infty}\mu_j\binom{j}{i}q(t)^{j-i}(1-q(t))^i \\ &= q(t)-f_{k-2}(q(t)), \end{aligned}$$

using the identity

$$f_{k-2}(q(t))=\sum_{i=0}^{k-2}\sum_{j=i}^{\infty}\mu_j\binom{j}{i}q(t)^{j-i}(1-q(t))^i$$

which is derived in the proof of theorem 3.1.

Since ϵ can be made arbitrarily small, the conditions established in equations 14 and 15 are equivalent to the statement of Theorem 3.2. ∎

A.2 The Giant 2-core.

Theorem A.4 *Let $\mathcal{D}$ satisfy the conditions of Theorem 3.2. Then*

1. If $\sum_i i\mu_i>1$ then there exists a constant C such that the 2-core of $G(\mathcal{D})$ contains $\geq Cn$ vertices w.e.h.p.
2. If $\sum_i i\mu_i\leq 1$ then for every $C>0$, the 2-core of $G(\mathcal{D})$ contains less Cn vertices w.e.h.p.

Proof. The conditions necessary for an asymptotic degree sequence to produce a giant 2-core w.e.h.p. according to theorem 3.2 are that $f_0(q)=g(q)$ must have a fixed point in the interval $[0,1)$. This is equivalent to the condition that a $\{\mu_i\}$ branching process survives with positive probability. Since g has all positive derivatives in $[0,1)$, and since $g(1)=1$ and $g(0)\ geq 0$, it is well known that g has such a fixed point if and only if $g'(1)=\sum_i i\mu_i>1$. ∎

From the results in [11, 12] for the a.a.s. presence of a giant component we can obtain the following.

Corollary A.5 *A random graph G with a sparse smooth degree sequence has a giant 2-core a.a.s. if and only if it has a giant component a.a.s.*