Available online at www.sciencedirect.com

ScienceDirect

Electronic Notes in DISCRETE MATHEMATICS

www.elsevier.com/locate/endm

How does the core sit inside the mantle?

Amin Coja-Oghlan [a,1,2], Oliver Cooley [b,3], Mihyun Kang [b,4] and Kathrin Skubch [a,5]

[a] *Mathematics Institute*
Goethe University
60325 Frankfurt, Germany

[b] *Institute of Optimization and Discrete Mathematics (Math B)*
Technische Universität Graz
8010 Graz, Austria

Abstract

The k-core, defined as the largest subgraph of minimum degree k, of the random graph $\boldsymbol{G}(n, p)$ has been studied extensively. In a landmark paper Pittel, Wormald and Spencer [Journal of Combinatorial Theory, Series B **67** (1996) 111–151] determined the threshold d_k for the appearance of an extensive k-core. Here we derive a multi-type Galton-Watson branching process that describes precisely how the k-core is "embedded" into the random graph for any $k \geq 3$ and any fixed average degree $d = np > d_k$. This generalises prior results on, e.g., the internal structure of the k-core.

Mathematics Subject Classification: 05C80 (primary), 05C15 (secondary)

Keywords: Random graphs, k-core, branching process, Warning Propagation

[1] The first author thanks Victor Bapst and Guilhem Semerjian for helpful discussions.

[2] Email: `acoghlan@math.uni-frankfurt.de`
[3] Email: `cooley@math.tugraz.at`
[4] Email: `kang@math.tugraz.at`
[5] Email: `skubch@math.uni-frankfurt.de`

1 Introduction

1.1 Background

For any $k \geq 3$ the k-core $\mathscr{C}_k(G)$ of a graph G is defined as the (unique) maximal subgraph of G of minimum degree k. For fixed $d > 0$ let $\boldsymbol{G} = \boldsymbol{G}(n, d/n)$ denote the random graph on the vertex set $[n] = \{1, \ldots, n\}$ in which any two vertices are connected with probability $p = d/n$ independently. Pittel, Wormald and Spencer [13] were the first to determine the precise threshold d_k beyond which the k-core $\mathscr{C}_k(\boldsymbol{G})$ is non-empty w.h.p. Specifically, for any $k \geq 3$ there is a function $\psi_k : (0, \infty) \to [0, 1]$ such that for any $d \in (0, \infty) \setminus \{d_k\}$ the sequence $(n^{-1}|\mathscr{C}_k(\boldsymbol{G})|)_n$ converges to $\psi_k(d)$ in probability. Furthermore, Pittel, Wormald and Spencer pointed out that a simple "branching process" heuristic predicts the correct threshold and the correct size of the k-core, and this argument has subsequently been turned into an alternative proof of their result [11,14].

The aim of the present paper is to enhance this branching process perspective of the k-core problem. More specifically, we are concerned with the following question. Fix $k \geq 3$, $d > d_k$ and let $s > 0$ be an integer. Generate a random graph $\boldsymbol{G}$ and mark each vertex according to $\sigma_{k,\boldsymbol{G}} : V(\boldsymbol{G}) \to \{0, 1\}$, $v \mapsto \mathbf{1}\{v \in \mathscr{C}_k(\boldsymbol{G})\}$. Further, for a vertex v in $\boldsymbol{G}$ we let $\boldsymbol{G}_v$ denote the component of v. Then $(\boldsymbol{G}_v, v, \sigma_{k,\boldsymbol{G}_v})$ is a rooted $\{0,1\}$-marked graph, whose marks indicate the membership of the k-core of the component $\boldsymbol{G}_v$. Now, pick a vertex $\boldsymbol{v}$ uniformly at random and let $\partial^s[\boldsymbol{G}_{\boldsymbol{v}}, \boldsymbol{v}, \sigma_{k,\boldsymbol{G}_{\boldsymbol{v}}}]$ denote the isomorphism class of the finite rooted $\{0,1\}$-marked graph obtained by deleting all vertices at distance greater than s from $\boldsymbol{v}$ from $\boldsymbol{G}_{\boldsymbol{v}}$. Our aim is to determine the distribution of $\partial^s[\boldsymbol{G}_{\boldsymbol{v}}, \boldsymbol{v}, \sigma_{k,\boldsymbol{G}_{\boldsymbol{v}}}]$. Of course, *without* the marks the standard branching process analogy yields convergence to the "usual" Galton-Watson tree $\boldsymbol{T}(d)$ with $\mathrm{Po}(d)$ offspring. The point of the present paper is to exhibit a multi-type branching process that yields the limiting distribution of the $\{0,1\}$-marked subgraph.

1.2 Results

To accomodate the non-trivial correlations between the k-core and the "mantle" (i.e., the vertices outside the core) we will introduce a Galton-Watson process $\hat{\boldsymbol{T}}(d, k, p)$ that posseses five different vertex types, denoted by 000, 001, 010, 110, 111. Setting

$$q = q(d, k, p) = \mathbb{P}\left[\mathrm{Po}(dp) = k-1 | \mathrm{Po}(dp) \geq k-1\right], \tag{1.1}$$

we define $p_{000} = 1 - p$, $p_{010} = pq$ and $p_{110} = p(1-q)$. The process starts with a single vertex v_0, whose type is chosen from $\{000, 010, 111\}$ according to the

$$
\begin{aligned}
g_{000}(\boldsymbol{x}) &= \exp(d(1-p)x_{000})\frac{\sum_{h=0}^{k-2}(dp)^h(qx_{010}+(1-q)x_{110})^h/h!}{\sum_{h=0}^{k-2}(dp)^h/h!},\\
g_{001}(\boldsymbol{x}) &= \bar{q}\left(\exp(d(1-p)x_{001})\left(qx_{010}+(1-q)x_{110}\right)^{k-2}\right)\\
&\qquad +(1-\bar{q})\left(\exp(d(1-p)x_{000})\frac{\sum_{h=0}^{k-3}(dp)^h(qx_{010}+(1-q)x_{110})^h/h!}{\sum_{h=0}^{k-3}(dp)^h/h!}\right),\\
g_{010}(\boldsymbol{x}) &= \exp(d(1-p)x_{001})\left(qx_{010}+(1-q)x_{110}\right)^{k-1},\\
g_{110}(\boldsymbol{x}) &= \exp(d(1-p)x_{001})\frac{\sum_{h\ge k}(dpx_{111})^h/h!}{\sum_{h\ge k}(dp)^h/h!},\\
g_{111}(\boldsymbol{x}) &= \exp(d(1-p)x_{001})\frac{\sum_{h\ge k-1}(dpx_{111})^h/h!}{\sum_{h\ge k-1}(dp)^h/h!}.
\end{aligned}
$$

Fig. 1. The generating functions $g_{z_1 z_2 z_3}(\boldsymbol{x})$.

distribution $(p_{000}, p_{010}, p_{111})$. Subsequently, each vertex of type $z_1z_2z_3$ spawns a random number of vertices of each type. The offspring distributions are defined by the generating functions $g_{z_1z_2z_2}(\boldsymbol{x})$ detailed in Figure 1, where we denote $\boldsymbol{x} = (x_{000}, x_{001}, x_{010}, x_{110}, x_{111})$ and

$$\bar{q} = \bar{q}(d,k,p) = \mathbb{P}\left[\mathrm{Po}(dp) = k-2|\mathrm{Po}(dp) \le k-2\right].$$

Finally, we turn the resulting Galton-Watson tree into a $\{0,1\}$-marked tree rooted at v_0 by giving mark 0 to all vertices of type 000, 001 or 010, and mark 1 to all others. Let $\boldsymbol{T}(d,k,p)$ signify the resulting (possibly infinite) random rooted $\{0,1\}$-marked tree.

Theorem 1.1 *Assume that $k \ge 3$ and $d > d_k$. Let $s \ge 0$ be an integer and let τ be a rooted $\{0,1\}$-marked tree. Moreover, let p^* be the largest fixed point of*

$$\phi_{d,k} : [0,1] \to [0,1], \qquad p \mapsto \mathbb{P}\left[\mathrm{Po}(dp) \ge k-1\right]. \tag{1.2}$$

Then

$$\frac{1}{n}\sum_{v\in V(\boldsymbol{G})} \mathbf{1}\left\{\partial^s[\boldsymbol{G}, v, \sigma_{k,\boldsymbol{G}_v}] = \partial^s[\tau]\right\}$$

converges to $\mathbb{P}\left[\partial^s[\boldsymbol{T}(d,k,p^*)] = \partial^s[\tau]\right]$ in probability.

Since we will derive our result from the convergence of the empirical distribution of the marked neighbourhoods of vertices v in $V(\boldsymbol{G})$, our proof will involve several statements about the convergence of probability distributions on the set of isomorphism classes of rooted $\{0,1\}$-marked graphs.

2 Related work

Since the work of Pittel, Wormald and Spencer [13] several different arguments for determining the location of the k-core for $k \geq 3$ have been suggested. Some results on the local structure of the core and the mantle follow directly from these analyses. For instance, the Poisson cloning model [9] immediately implies that the *internal* local structure of the k-core can be described by a simple (single-type) Galton-Watson process. Riordan also pointed out that this local description follows from his analysis [14]. Furthermore, Cooper [2] derived the asymptotic distribution of the internal and the external degree sequences of the vertices in the mantle, i.e., of the number of vertices with a given number of neighbours in the core and a given number of neighbours outside.

The contribution of the present work is that we exhibit a branching process that describes the structure of the core together with the mantle. Neither the construction of the core via the "peeling process" nor the branching process analogy from [11,13,14] reveal how the core "embeds" into the mantle. In fact, even though [2] asymptotically determines the degree distribution of the core along with the combined degrees of the vertices in the mantle, the conditional random graph is *not* uniformly random subject to these.

Structures that resemble cores of random (hyper)graphs have come to play an important role in the study of random constraint satisfaction problems. This was first noticed in non-rigorous but analytic work based on ideas from statistical physics (see [10] and the references therein). Indeed, in the physics literature it was suggested to characterise the core by means of a "message passing" algorithm called *Warning Propagation* [10, Chapter 18]. A similar idea is actually implicit in Molloy's paper [12, proof of Lemma 6].

3 Proof outline

There is a very natural formulation of Warning Propagation to identify the k-core of a given graph G. It is based on introducing "messages" on the edges of G and marks on the vertices of G, both with values in $\{0,1\}$. These will be updated iteratively in terms of a "time" parameter $t \geq 0$. At time $t = 0$ we start with the configuration in which all messages are equal to 1, i.e., $\mu_{v\to w}(0|G) = 1$ for all $\{v, w\} \in E(G)$. Inductively, writing ∂v for the neighbourhood of vertex v and abbreviating $\partial v \setminus w = \partial v \setminus \{w\}$, we let

$$\mu_{v\to w}(t+1|G) = \mathbf{1}\left\{\sum_{u\in\partial v\setminus w} \mu_{u\to v}(t|G) \geq k-1\right\}. \tag{3.1}$$

The messages are "directed". That is, at each time $t \geq 0$ there are *two* messages $\mu_{v\to w}(t|G)$, $\mu_{w\to v}(t|G)$ travelling along the edge $\{v,w\}$. Additionally, the mark of $v \in [n]$ at time t is

$$\mu_v(t|G) = \mathbf{1}\left\{\sum_{u\in\partial v} \mu_{u\to v}(t|G) \geq k\right\}. \tag{3.2}$$

It is easy to see that the messages converge to a fixed point for any G, and that the set of vertices marked one in the fixed point coincides with the k-core. Our aim is to show that the multi-type branching process $\boldsymbol{T}(d,k,p^*)$ describes the distribution of the Warning Propagation fixed point on the infinite $\mathrm{Po}(d)$ Galton-Watson tree. The key step of the proof is to turn the problem of tracing how Warning Propagation passes messages from the "bottom" of the Galton-Watson tree up toward the root into a process where messages are passed "top-down", i.e., in the fashion of a branching process.

But first we will reduce the study of the Warning Propagation fixed point on $\boldsymbol{G}$ to the study of Warning Propagation on the (infinite) Galton-Watson tree with $\mathrm{Po}(d)$ offspring. Let v be a vertex in $\boldsymbol{T}(d)$ and let w be its parent. Let $\mu_{v\uparrow}(t|\boldsymbol{T}(d)) = \mu_{v\to w}(t|\boldsymbol{T}(d))$ denote the "bottom-up" message from v to w. Furthermore, define

$$\mu_{v_0\uparrow}(t|\boldsymbol{T}(d)) = \mathbf{1}\left\{\sum_{w\in\partial v_0} \mu_{w\to v_0}(t|\boldsymbol{T}(d)) \geq k-1\right\}.$$

Then the following fixed point problem will help us to make a connection between Warning Propagation on $\boldsymbol{G}$ and on the Galton-Watson tree.

Lemma 3.1 *Suppose $d > d_k$ and let p^* be the largest fixed point of the function $\phi_{d,k}$ from (1.2). Then $\phi_{d,k}$ is contracting on $[p^*,1]$. Moreover, it holds that $\psi_k(d) = \mathbb{P}\left[\mathrm{Po}(dp^*) \geq k\right] = \phi_{d,k+1}(p^*)$.*

The above lemma together with the recursive structure of $\boldsymbol{T}(d)$ imply that the sequence $(\mu_{v_0\uparrow}(t|\boldsymbol{T}(d)))_{t\geq 0}$ converges almost surely to a random variable whose expectation is p^*. By the definition of the Warning Propagation marks $\mu_{v_0}(t|\boldsymbol{T}(d))$ at v_0 it is easily verified that this implies that $(\mu_{v_0}(t|\boldsymbol{T}(d)))_{t\geq 0}$ converges to $\psi_k(d)$ in probability as t tends to infinity. On the other hand, the standard branching process analogy implies that the relative number of vertices marked with $\mu_v(t|\boldsymbol{G}) = 1$ in $\boldsymbol{G}$ converges to the probability that the root of a standard Galton-Watson tree with $\mathrm{Po}(d)$ offspring is marked with 1 after t iterations of Warning Propagation on the tree (as n tends to infinity). Since $\psi_k(d)$ is the asymptotic size of $\mathscr{C}_k(\boldsymbol{G})$, we obtain that in the case of the random graph $\boldsymbol{G}$, a *bounded* number of iterations of the Warning Propagation message passing algorithm on $\boldsymbol{G}$ suffice to obtain an accurate approximation of the k-core w.h.p.

It remains to determine the limit of the distributions of $\{0,1\}$-marked rooted trees obtained by marking $\boldsymbol{T}(d)$ with the Warning Propagation marks $\mu_v(t|\boldsymbol{T}(d))$ at time t (as t tends to infinity). We begin with determining the limit of the distribution of the first s levels of $\boldsymbol{T}(d)$ marked with the messages $\mu_{v\uparrow}(t|\boldsymbol{T}(d))$ from vertices v to their parent. Let,

$$\theta^s_{d,k,t} = \mathscr{L}(\partial^s[\boldsymbol{T}(d), v_0, \mu_{\cdot\uparrow}(t|\boldsymbol{T}(d))])$$

be the distribution of the isomorphism class of the obtained tree. The key feature of the messages $\mu_{v\uparrow}(t|\boldsymbol{T}(d))$ is that they are solely determined by the tree pending on v and are therefore much more convenient to work with. On the other hand they also contain all the information that we need to compute the Warning Propagation marks $\mu_v(t|\boldsymbol{T}(d))$. The recursive structure of $\boldsymbol{T}(d)$ and the results established in the first part of the poof suggest that the "boundary messages" sent out by the vertices at distance precisely s from v_0 converge to a sequence of mutually independent $\mathrm{Be}(p^*)$ variables. Since the messages $\lim_{t\to\infty}\mu_{u\uparrow}(t|\boldsymbol{T}(d))$ for vertices u at distance less than s from the root v_0 are *determined* by the "boundary messages", the expected limiting distribution is the one obtained by creating the first s levels of a random tree $\boldsymbol{T}(d)$, marking each vertex at distance precisely s by an independent $\mathrm{Be}(p^*)$ "message", and passing the messages up to the root (iteratively in terms of t). Let $\mu^*_{v\uparrow}(t|\boldsymbol{T}(d), s)$ denote the corresponding mark of each vertex v in $\boldsymbol{T}(d)$ and define

$$\theta^{s,*}_{d,k} = \mathscr{L}\left(\partial^s[\boldsymbol{T}(d), v_0, \mu^*_{\cdot\uparrow}(s|\boldsymbol{T}(d), s)]\right).$$

Then the following lemma confirms this hypothesis.

Lemma 3.2 *We have* $\lim_{t\to\infty}\theta^s_{d,k,t} = \theta^{s,*}_{d,k}$ *for all* $s \geq 0$.

We proceed by constructing a "top-down" process that produces the limiting distribution of the truncated tree. More precisely, define a random $\{0,1\}$-marked tree $\boldsymbol{T}^*(d,k)$ by means of the following two-type branching process. Initially, there is a root vertex v_0 that has type 1 with probability p^* and type 0 with probability $1-p^*$. The offspring of a type 0 vertex is $\mathrm{Po}(d(1-p^*))$ type 0 vertices and independently $\mathrm{Po}_{<k-1}(dp^*)$ type 1 vertices. Further, a type 1 vertex spawns $\mathrm{Po}(d(1-p^*))$ type 0 offspring and independently $\mathrm{Po}_{\geq k-1}(dp^*)$ type 1 offspring. The mark of each vertex v, denoted by $\mu^*_{v\uparrow}$, is identical to its type. Then the fixed point property of p^* implies the following lemma.

Lemma 3.3 *For any* $s \geq 0$ *we have* $\mathscr{L}(\partial^s[\boldsymbol{T}^*(d,k)]) = \theta^{s,*}_{d,k}$.

Lemmas 3.2 and 3.3 show that the marks $\mu^*_{v\uparrow}$ of $\boldsymbol{T}^*(d,k)$ correspond to the "upward messages" that are sent toward the root in the tree $\boldsymbol{T}(d)$. Our ultimate interest is in the marks $\mu_v(t|\boldsymbol{T}(d))$. To compute these marks on $\boldsymbol{T}(d)$ it remains to get a handle on the Warning Propagation messages from vertices to their children. Of course, in the tree $\boldsymbol{T}(d)$ these "top-down" messages and the Warning Propagation marks can be computed recursively from the messages $\mu_{v\uparrow}(t|\boldsymbol{T}(d))$. In the next part of the proof we are going to mimic this recursive construction of the "top-down" messages and marks on $\boldsymbol{T}(d)$ and add two corresponding bits to the mark of each vertex in $\boldsymbol{T}^*(d,k)$. Let $\hat{\boldsymbol{T}}^*(d,k)$ signify the resulting tree. Since the construction of the additional bits on $\boldsymbol{T}^*(d,k)$ matches with the construction of the Warning Propagation marks and messages from vertices in $\boldsymbol{T}(d)$ to their children, it holds that the distribution of the analogously marked Galton Watson tree converges to $\hat{\boldsymbol{T}}^*(d,k)$. In the last part of the proof we make the connection to the branching process $\hat{\boldsymbol{T}}(d,k,p^*)$. We will show that for every vertex in the supplemented tree $\hat{\boldsymbol{T}}^*(d,k)$ the distribution of the number of children of each type coincides with the offspring distribution determined from the generating functions $g_{z_1 z_2 z_3}$, which implies the following lemma.

Lemma 3.4 *We have* $\mathscr{L}([\hat{\boldsymbol{T}}^*(d,k)]) = \mathscr{L}([\hat{\boldsymbol{T}}(d,k,p^*)])$.

Finally, we obtain our main result from the above statements by reformulating it within the theory of local weak convergence.

References

[1] Cain, J., N. Wormald, *Encores on cores*, The Electronic Journal of Combinatorics **13** (2006) R81.

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

[3] Darling, R., J. Norris, *Differential equation approximations for Markov chains* Probability Surveys **5** (2008), 37–79.

[4] Erdős, P., A. Rényi, *On the evolution of random graphs* Magayar Tud. Akad. Mat. Kutato Int. Kozl. **5** (1960), 17–61.

[5] Fernholz, D., V. Ramachandran, *The giant k-core of a random graph with a specified degree sequence*, Manuscript (2003).

[6] Fernholz, D., V. Ramachandran, *Cores and connectivity in sparse random graphs*, UTCS Technical Report TR04-13 (2004).

[7] Janson, S., M. Luczak, *A simple solution to the k-core problem* Random Structures and Algorithms **30** (2007), 50–62.

[8] Karp, R., *The transitive closure of a random digraph* Random Structures and Algorithms **1** (1990), 73–93.

[9] Kim, J.H., *Poisson cloning model for random graphs* Proceedings of the International Congress of Mathematicians (2006), 873–897.

[10] Mézard, M., A. Montanari, "Information, physics and computation", Oxford University Press (2009).

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

[12] Molloy, M., *The freezing threshold for k-colourings of a random graph*, Proc. 43rd STOC (2012), 921–930.

[13] Pittel, B., 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.

[14] Riordan, O., *The k-core and branching processes*, Combinatorics, Probability and Computing **17** (2008), 111–136.