Analysis of the parallel peeling algorithm: a short proof

Pu Gao
University of Toronto
pu.gao@utoronto.ca

Given a (hyper)graph $H$ and a positive integer $k$, the parallel peeling algorithm repeatedly removes all vertices of degree less than $k$ and their incident edges. When the algorithm terminates, the output is the $k$-core of $H$. Let $s(H)$ denote the number of rounds the algorithm takes. It was first proved by Achlioptas and Molloy [1] that, if $H_r(n, p)$ is a random $r$-uniform hypergraph on $[n]$ with edge density $p = c/n^r$, where $c > 0$ is a constant not equal to $c_{r,k}$, the emergence threshold of a non-empty $k$-core, then $s(H) = O(\log n)$ (here $r, k$ are both at least 2 and are not both equal to 2). Recently, a paper by Jiang, Mitzenmacher and Thaler [2] improved this result by showing that, if $c > c_{r,k}$, then $s(H_r(n, c/n^r - 1)) = \Omega(\log n)$, i.e. the upper bound in [1] is tight; if $c < c_{r,k}$, then $s(H_r(n, c/n^r - 1)) \leq a_{r,k} \log \log n + O(1)$ where $a_{r,k} = 1/\log((r-1)(k-1))$, which significantly improves [1]. The lower bound in the supercritical case is relatively easier whereas most of the technical proof of [2] was for the upper bound in the subcritical case. In this note, I give a very short proof of asymptotically the same upper bound as in [2] (with a slightly larger coefficient than $a_{r,k}$) in the subcritical case. In fact, my proof mainly combines several well-known results in literature. I will prove the following.

**Theorem 1** Assume $k, r \geq 2$, $(k, r) \neq (2, 2)$ and $c < c_{r,k}$. Then a.a.s. $s(H_r(n, c/n^r - 1)) \leq (a_{r,k}^* + o(1)) \log \log n$, where $a_{r,k}^* = 1/\log(k(r-1)/r)$.

Here is the key lemma I use.

**Lemma 2** Assume $k, r \geq 2$, $(k, r) \neq (2, 2)$ and $c = O(1)$. A.a.s. every subgraph of $H_r(n, c/n^r - 1)$ with less than $\log^2 n$ vertices has average degree less than $r/(r-1) + \epsilon$ for every constant $\epsilon > 0$.

**Proof.** Let $X_{s,t}$ denote the number of subgraphs of $H_r(n, c/n^r - 1)$ with $s$ vertices and at least $t$ edges. Then,

$$
\mathbb{E}X_{s,t} \leq \binom{n}{s} \left( \binom{s^r}{t} \left( \frac{c}{n^{r-1}} \right)^t \right).
$$

Fix a constant $0 < \epsilon < 1$; let $t = (1 + \epsilon)s/(r-1)$; then

$$
\mathbb{E}X_{s,t} \leq \left( \frac{en}{s} \left( \frac{e(r-1)s^r c}{(1 + \epsilon)s n^{r-1}} \right)^{(1+\epsilon)/(r-1)} \right)^s \leq \left( C \left( \frac{s}{n} \right)^\epsilon \right)^s,
$$

*Research supported by an NSERC Postdoctoral Fellowship.*
for some constant $C > 0$ depending only on $r$, $k$ and $c$. Now immediately we have $\sum_{1 \leq s \leq \log^2 n} E(X_{s,t}) = o(1)$ and the lemma follows as each edge contributes $r$ to the total degree of a subgraph and $\epsilon > 0$ is arbitrary. $lacksquare$

The following proposition is from [1, Section 8].

**Proposition 3** Assume $k, r \geq 2$, $(k,r) \neq (2,2)$ and $c < c_{r,k}$; let $H = \mathcal{H}_r(n, c/n^{r-1})$. Then a.a.s. there is a constant $I > 0$, such that after $I$ rounds of the parallel peeling algorithm are applied to $H$, every component of the remaining graph, denoted by $H_I$, has size $O(\log n)$.

**Proof of Theorem 1.** Let $I$ be a constant chosen to satisfy Proposition 3 and let $H_I$ be the remaining graph after $I$ rounds of the parallel peeling algorithm. Then, a.a.s. every component of $H_I$ contains $O(\log n)$ vertices. By Lemma 2 we may assume that each component has average degree at most $r/(r - 1) + \epsilon$ for any constant $\epsilon > 0$. Take an arbitrary constant $C$ of $H_I$. Let $C_0, C_1, \ldots.$ denote the process produced by running the parallel peeling algorithm on $C_0 = C$. By Lemma 2 we may assume that each $C_i$ has average degree at most $r/(r - 1) + \epsilon$. Let $\rho_i$ denote the proportion of vertices in $C_i$ with degree at least $k$. Then $k \rho_i \leq r/(r - 1) + \epsilon$ for every $i \geq 0$; i.e. $\rho_i \leq \rho := r/k(r - 1) + \epsilon/k$. By our assumption on $k$ and $r$, we always have $\rho < 1$. Since all vertices with degree less than $k$ are removed in each step of the algorithm, we have $|V(C_{i+1})| \leq \rho |V(C_i)|$ for every $i \geq 0$. This immediately gives $s(C) \leq (\log \log n + O(1))/\log \rho^{-1}$. Since $\epsilon > 0$ can be taken arbitrarily small, we have $s(C) \leq (a^*_{r,k} + o(1)) \log \log n$. This holds a.a.s. for every component of $H_I$. Hence, a.a.s. $s(\mathcal{H}_r(n, c/n^{r-1})) \leq I + (a^*_{r,k} + o(1)) \log \log n = (a^*_{r,k} + o(1)) \log \log n$. $lacksquare$

**References**

[1] D. Achlioptas and M. Molloy. The solution space geometry of random linear equations. Random Structures and Algorithms (to appear).

[2] J. Jiang, M. Mitzenmacher and J. Thaler, Parallel Peeling Algorithms, [arXiv:1302.7014](http://arxiv.org/abs/1302.7014).