

On the Expected Performance of a Parallel Algorithm for Finding Maximal Independent Subsets of a Random Graph

Neil J. Calkin and A. M. Frieze*

*Department of Mathematics, Carnegie Mellon University, Pittsburgh,
PA 15213-3890*

L. Kučera

*Faculty of Mathematics, Dept. of Applied Mathematics, Charles University,
Malostranské nám. 25, 11800 Prague, Czechoslovakia*

ABSTRACT

We consider the parallel greedy algorithm of Coppersmith, Raghavan, and Tompa (*Proc. of 28th Annual IEEE Symp. on Foundations of Computer Science*, pp. 260–269, 1987) for finding the lexicographically first maximal independent set of a graph. We prove an $\Omega(\log n)$ bound on the expected number of iterations for most edge densities. This complements the $O(\log n)$ bound proved in Calkin and Frieze (*Random Structures and Algorithms*, Vol. 1, pp. 39–50, 1990).

1. INTRODUCTION

In this note we consider the problem of finding the lexicographically first maximal independent set (LFMIS) in a random graph. Coppersmith, Raghavan, and Tompa [3] describe a parallel version of the standard greedy algorithm for this problem:

* Research supported in part by NSF Grant CCR-8900112.

Suppose we are given a graph $G = (V, E)$, $V = [n] = \{1, 2, \dots, n\}$. For $Z \subseteq V$ we let

$$\Gamma^+(Z) = \{x \notin Z : xz \in E \text{ for some } z < x, z \in Z\},$$

and

$$\Gamma^-(Z) = \{x \notin Z : xz \in E \text{ for some } z > x, z \in Z\}.$$

Note that we have implicitly oriented the edges from low to high.

algorithm PARALLEL GREEDY (G);
begin
 $\text{GIS} \leftarrow \emptyset$;
 until G has no vertices **do**
 begin
 let $S = \{a : \Gamma^-(a) = \emptyset\}$;
 $\text{GIS} \leftarrow \text{GIS} \cup S$;
 remove $S \cup \Gamma(S)$ from G
 end
 output GIS
end

It is easy to see [3, Lemma 2.1] that GIS is the LFMIS. Cook [2] showed that the problem of computing the LFMIS of a graph is complete for P and so is not in NC unless $\text{NC} = \text{P}$. PARALLEL-GREEDY can be implemented on a CRCW PRAM in $O(1)$ time per iteration if one processor is allocated to each edge of G .

Coppersmith, Raghavan, and Tompa showed that if $T(n, p)$ denotes the *expected* number of iterations $\tau = \tau(G)$ when $G = G_{n,p}$ then $T(n, p) = O\left(\frac{(\log n)^2}{\log \log n}\right)$. ($G_{n,p}$ is the random graph with vertex set $[n]$ where each edge occurs independently with probability $p = p(n)$).

They conjectured that $T(n, p) = O(\log n)$ and this was proved in Calkin and Frieze [1]. They proved:

Theorem 1. (a) $\frac{\alpha \log n}{4 \log \log n} \leq T(n, p)$ for $\frac{1}{n} \leq p \leq \frac{1}{n^\alpha}$ where $0 < \alpha \leq 1$ is constant,
 (b) $T(n, p) = O(\log n)$. The hidden constant in (b) is independent of p .

Note that our inequalities are only claimed for n large.

The upper bounds and lower bounds in Theorem 1 are slightly different. It leaves open the possibility that $T(n, p) = O\left(\frac{\log n}{\log \log n}\right)$ throughout. The aim of this article is to shed more light on this problem, and to prove:

Theorem 2. Assume $0 \leq \alpha < 1$, α constant. (a) $T(n, p) \leq \frac{3 \log n}{(1 - \alpha) \log \log n}$ for $p \leq \frac{(\log n)^\alpha}{n}$, (b) $T(n, p) = \Omega(\log n)$ for $\alpha \geq p \geq \frac{1}{n^\alpha}$, where the hidden constant in (b) depends on α .

Proof. (a) Let $G = G_1 \supseteq G_2 \supseteq G_3 \supseteq \dots$ denote the sequence of graphs produced by each iteration of the algorithm.

For $v \in V(G_t)$ and $t \geq 1$ let $\alpha(t, v)$ = the length of the longest directed path in G_t which ends at v (a path $(v_1, v_2, \dots, v_k)$, is directed if $v_1 < v_2 < \dots < v_k$)).

Clearly, if $v \in V(G_{t+1})$ then $\alpha(t+1, v) \leq \alpha(t, v) - 2$.

Hence

$$\tau(G) \leq \frac{1}{2} \max\{v \in V(G) : \alpha(1, v)\}.$$

Thus

$$\Pr(\tau(G_{n,p}) \geq k) \leq E(\# \text{ of directed paths of length } 2k)$$

$$\begin{aligned} &= \binom{n}{2k} p^{2k-1} \\ &\leq n \left(\frac{nep}{2k} \right)^{2k-1} \\ &\leq n \left(\frac{e(\log n)^\alpha}{2k} \right)^{2k-1}. \end{aligned}$$

Hence, with $k_0 = \left\lceil \frac{2 \log n}{(1-\alpha) \log \log n} \right\rceil$,

$$\begin{aligned} T(n, p) &= \sum_{k=1}^n \Pr(\tau(G_{n,p}) \geq k) \\ &\leq k_0 + n \sum_{k=k_0+1}^n \left(\frac{e(\log n)^\alpha}{2k} \right)^{2k-1} \\ &\leq k_0 + 2n \left(\frac{e(\log n)^\alpha}{2k_0} \right)^{2k_0-1} \\ &\leq k_0 + 2n \left(\frac{A \log \log n}{(\log n)^{1-\alpha}} \right)^{2k_0-1} \end{aligned}$$

where $A = e(1-\alpha)/4$,

$$= k_0 + o(1).$$

This completes the proof of (a). ■

(b) This is somewhat less trivial.

Let

$$\begin{aligned} V_t &= V(G_t) \\ &= \{\text{vertices remaining at the start of round } t\} \\ S_t &= \text{Set } S \text{ found in round } t \\ &= \{\text{sources found in round } t\}, \\ N_t &= \Gamma(S_t) \cap V_t \\ &= \{\text{neighbors of } S_t \text{ deleted in round } t\}. \end{aligned}$$

Suppose $i \geq 2$ and $A_t, B_t, 1 \leq t \leq i-1$ is some disjoint collection of subsets of V . Then we have $S_i = A_i, N_i = B_i$ for $1 \leq t \leq i-1$ if and only if

- (2a) $v \in A_i$ implies $\Gamma^-(v) \subseteq \bigcup_{s=1}^{i-1} B_s$ and $\Gamma^-(v) \cap B_{i-1} \neq \emptyset, 1 \leq t \leq i-1$ (when $t=1$, drop the second condition)
 (2b) $v \in B_i$ implies $\Gamma^-(v) \cap \bigcup_{s=1}^{i-1} A_s = \emptyset$ and $\Gamma^-(v) \cap A_i \neq \emptyset, 1 \leq t \leq i-1$ and

$$v \in C = V - \bigcup_{t=1}^{i-1} (A_t \cup B_t) \text{ implies}$$

- (3a) $\Gamma^-(v) \cap \bigcup_{t=1}^{i-1} A_t = \emptyset,$
 (3b) $\Gamma^-(v) \cap (B_{i-1} \cup C) \neq \emptyset.$

Suppose now that we choose sets $A_t, B_t, 1 \leq t \leq i-1$ satisfying (2) and condition on the event

$$\mathcal{E} = \{S_t = A_t, N_t = B_t, V_i = C : 1 \leq t \leq i-1\}.$$

It is important to establish the conditional distribution of the sets $\Gamma_i^-(v) = \Gamma^-(v) \cap V_i, v \in V_i, i \geq 2$. For $v \in V_i$ let $R_v^i = [v-1] \cap (V_i \cup B_{i-1})$ and $r_v = |R_v^i|$.

Claim 1. (i) The sets $\Gamma_i^-(v), v \in V_i$ are stochastically independent, (ii) $\Gamma_i^-(v)$ is a random subset of R_v^i chosen through r_v Bernoulli trials conditioned on the occurrence of at least one success, i.e.,

$$(4) \Pr(|\Gamma_i^-(v)| = k) = \binom{r_v}{k} p^k (1-p)^{r_v-k} / (1 - (1-p)^{r_v}), 1 \leq k \leq r_v$$

and each k -subset is equally likely.

Proof (of Claim). To prove (i) simply observe that condition (3) on $v \in C$ only involves edges directed into v , and that the conditions in (2) only involve edges directed into $V - C$.

Now consider (ii). $v \in V_2$ if and only if $\Gamma_i^-(v) \neq \emptyset$ and $\Gamma_i^-(v) \cap S_1 = \emptyset$ and these conditions are equivalent to (ii). We can now proceed inductively. Fix $v \in V_i$. If $v \notin S_i \cup N_i$ then we learn (a) $\Gamma_i^-(v) \cap V_i \neq \emptyset$, then (ii) $\Gamma_i^-(v) \cap S_i = \emptyset$ and so finally that

$$\Gamma_i^-(v) \cap (V_i - S_i) = \Gamma_i^-(v) \cap R_v^{i+1} \neq \emptyset.$$

Thus (4) continues to hold.
 End of proof (of claim). ■

We now continue with the proof of our Theorem. Choose $\beta, \alpha < \beta < 1$. Now choose $i \leq \tau = \left\lceil \frac{(1-\alpha)\log n}{10} \right\rceil$ and assume that $V_i = \{x_1 < x_2 < \dots < x_s\}$. Partition V_i into X_1, X_2, Y where $X_1 = \{x_1, x_2, \dots, x_a\}, a = \lceil \log n/p \rceil, X_2 = \{x_{a+1}, x_{a+2}, \dots, x_b\}, b = \lceil (\log n)^2/p \rceil$, and Y is the rest of V_i . We will show that a good proportion of Y is likely to remain in V_{i+1} , when V_i is large enough so that the above partition is actually possible.

Observe first that the proof of Claim 1 implies that if $r = |B_{i-1} \cap [x_j - 1]|$, then

$$(5) \Pr(x = x_j \in S_i) = (1 - (1 - p)^r)(1 - p)^{j-1}/(1 - (1 - p)^{r_x}) \\ \leq (1 - p)^{j-1}$$

(At least one success is required in the r trials corresponding to $B_{i-1} \cap [x_j - 1]$ and no further successes.)

So if $\mathcal{A}_i = \{S_i \cap (X_2 \cup Y) = \emptyset\}$, then

$$(6) \Pr(\bar{\mathcal{A}}_i) \leq \sum_{j>a} (1 - p)^{j-1} = \frac{(1 - p)^a}{p} \leq \frac{1}{np}.$$

Let

$$\mathcal{B}_i = \{\Gamma^-(y) \cap X_2 \neq \emptyset, \forall y \in Y\}$$

It follows from Claim 1(ii) that if $y \in Y$, then

$$\Pr(\Gamma^-(y) \cap X_2 = \emptyset) \leq (1 - p)^{b-a} \\ \leq n^{-(1-o(1))\log n}$$

and so

$$(7) \Pr(\bar{\mathcal{B}}_i) \leq n^{-(1-o(1))\log n}.$$

Note that (6),(7) can be taken as true even if $Y = \emptyset$.

Let us now consider the size of S_i . Let $\delta_j = 1$ if $x_j \in S_i$ and $\delta_j = 0$ otherwise. It follows from Claim 1(i) that $\delta_1, \delta_2, \dots, \delta_s$ are independent random variables. Also

$$E(|S_i|) = \sum_{j=1}^s \Pr(\delta_j = 1) \\ \leq \sum_{j=1}^s (1 - p)^{j-1} \\ \leq \frac{1}{p}.$$

Note that we have $\Pr(\delta_j = 1) \leq (1 - p)^{j-1}$ regardless of the history of the algorithm to this point. It follows that $|S_1| + |S_2| + \dots + |S_i|$ is dominated by the sum of independent random variables each of which is the sum of a large number of independent 0–1 random variables. It follows from Theorem 1 of Hoeffding [4] that if

$$\mathcal{C}_i = \left\{ |S_1| + |S_2| + \dots + |S_i| < \frac{(1 - \alpha)\log n}{2p} \right\}$$

then

$$\Pr(\bar{\mathcal{C}}_i) \leq \left(\frac{2ei}{(1-\alpha)\log n} \right)^{(1-\alpha)\log n/2p}$$

(Hoeffding proves that if $Z_1, Z_2, \dots, Z_m$ are independent random variables with $0 \leq Z_j \leq 1$, $j = 1, 2, \dots, m$ and $E(Z_1 + Z_2 + \dots + Z_m) = m\mu$ then

$$\Pr(Z_1 + Z_2 + \dots + Z_m \geq m(\mu + t)) \leq \left(\left(\frac{\mu}{\mu + t} \right)^{\mu + t} \left(\frac{1 - \mu}{1 - \mu - t} \right)^{1 - \mu - t} \right)^m.$$

So if $t = (\theta - 1)\mu$

$$\Pr(Z_1 + Z_2 + \dots + Z_m \geq \theta m\mu) \leq (\theta^{-\theta} e^{\theta-1})^{m\mu} < \left(\frac{e}{\theta} \right)^{\theta m\mu}.$$

We use this inequality with $m\mu = \frac{i}{p}$ and $\theta m\mu = \frac{(1-\alpha)\log n}{2p}$.

Note that $\mathcal{C}_\tau \subseteq \mathcal{C}_{\tau-1} \subseteq \dots \subseteq \mathcal{C}_1$ and

$$(8) \Pr(\bar{\mathcal{C}}_\tau) \leq n^{-(1-\alpha)\log(5/e)/2\alpha}.$$

Consider the size of $Y \cap V_{i+1}$. Using Claim 1(ii) we see that, given $\mathcal{A}_i \cap \mathcal{B}_i$, the edges joining X_1 to Y are unconditioned. So, by another use of [4],

$$(9) \Pr(|V_{i+1}| \leq \left(1 - \frac{1}{(\log n)^2}\right) |Y| (1-p)^{|S_i|} | \mathcal{A}_i \cap \mathcal{B}_i, |S_i|) \\ \leq \exp \left\{ - \frac{|Y| (1-p)^{|S_i|}}{2(\log n)^4} \right\}$$

since if $y \in Y$ then $\Pr(y \in V_{i+1} | \mathcal{A}_i \cap \mathcal{B}_i, |S_i|) = (1-p)^{|S_i|}$.
Now let

$$\mathcal{D}_i = \left\{ |V_i| > \left(1 - \frac{2}{(\log n)^2}\right)^{i-1} n (1-p)^{|S_1| + |S_2| + \dots + |S_{i-1}|} \right\}.$$

Then we have

$$(10) \Pr(\bar{\mathcal{D}}_{i+1}) \leq \Pr(\bar{\mathcal{A}}_i \cap \bar{\mathcal{B}}_i \cap \bar{\mathcal{C}}_i \cap \bar{\mathcal{D}}_i) + \Pr(\bar{\mathcal{D}}_{i+1} | \mathcal{A}_i \cap \mathcal{B}_i \cap \mathcal{C}_i \cap \mathcal{D}_i).$$

Now if $\mathcal{C}_i \cap \mathcal{D}_i$ occurs, then

$$|V_i| (1-p)^{|S_i|} \geq n \left(1 - \frac{2}{(\log n)^2}\right)^{i-1} (1-p)^{|S_1| + |S_2| + \dots + |S_i|} \\ \geq n \left(1 - \frac{2}{(\log n)^2}\right)^{i-1} (1-p)^{(1-\alpha)\log n/2p} \\ = (1 - o(1)) n^{1 + ((1-\alpha)/2p)\log(1-p)}$$

$$\text{and } |Y| \geq |V_i| - \frac{(\log n)^2}{p} \geq \left(1 - \frac{1}{(\log n)^2}\right) |V_i|.$$

Now, since $\mathcal{C}_i, \mathcal{D}_i$ refer to the history of the algorithm prior to the construction of $Y \cap V_{i+1}$ we may again argue as in (9) that

$$\Pr(\bar{\mathcal{D}}_{i+1} | \mathcal{A}_i \cap \mathcal{B}_i \cap \mathcal{C}_i \cap \mathcal{D}_i) \leq \exp \left\{ - \frac{(1 - o(1))n^{1 + ((1-\alpha)/2p)\log(1-p)}}{2(\log n)^4} \right\}.$$

Thus, from (6), (7), (8), (10) and the above

$$\Pr(\bar{\mathcal{D}}_{i+1}) \leq \Pr(\bar{\mathcal{D}}_i) + o((\log n)^{-1})$$

and so

$$\begin{aligned} \Pr(\bar{\mathcal{D}}_{i+1}) &\leq \Pr(\bar{\mathcal{D}}_1) + o(1) \\ &= o(1), \end{aligned}$$

since $\bar{\mathcal{D}}_1 = \emptyset$.

Thus $\Pr(\bar{\mathcal{D}}_r) = o(1)$. Combining this with $\Pr(\mathcal{C}_r) = 1 - o(1)$ we see that

$$\Pr(V_r = \emptyset) = o(1)$$

and this proves part (b) of the Theorem. ■

REFERENCES

- [1] N. Calkin and A. Frieze, Probabilistic analysis of a parallel algorithm for finding maximal independent sets, *Random Struct. Alg.* **1**, 39–50 (1990).
- [2] S. A. Cook, A taxonomy of problems with fast algorithms, *Inf. Control*, **64**, 2–22 (1985).
- [3] D. Coppersmith, P. Raghavan, and M. Tompa, Parallel graph algorithms that are efficient on average, *Proceedings of 28th Annual IEEE Symposium on Foundations of Computer Science*, 260–269 (1987).
- [4] W. Hoeffding, Probability inequalities for sums of bounded random variables, *J. Am. Stat. Assoc.*, **58**, 18–30 (1963).

Received March 22, 1991