Distributive Graph Algorithms - Global Solutions from Local Data

Nathan Linial

Institute of Mathematics and Computer Science, Hebrew University, Jerusalem,Israel

Abstract

This paper deals with distributed graph algorithms. Processors reside in the vertices of a graph G and communicate only with their neighbors. The system is synchronous and reliable, there is no limit on message lengths and local computation is instantaneous. The results: A maximal independent set in an n-cycle cannot be found faster than $\Omega(\log^* n)$ and this is optimal by [CV]. The d-regular tree of radius r cannot be colored with fewer than $\sqrt{d}$ colors in time $2r / 3$. If Δ is the largest degree in G which has order n, then in time $O(\log^* n)$ it can be colored with $O(\Delta^2)$ colors.

1. Introduction

In distributed processing all computations are made based on local data. The aim of this paper is to bring up limitations that follow from this local nature of the computation. Notice that within the various computational models for parallel computers this difficulty is specific to the distributed model. Shared memory allows fast dissemination of data, but no such means are available when dealing with distributed systems.

In the present paper we are mostly interested in proving lower bounds. We therefore assume a powerful version of the distributed model. There is a graph $G = (V,E)$ each node of which is occupied by a processor. Computation is completely synchronous and reliable. Every time unit each processor may pass messages to each of his neighbors. There is no limit on the size the messages may have. Also computations done by each processor individually take no time at all and are not restricted in any way. Our concern is only with the radius of the neighborhood around each node from which data may be collected. This is indeed all that matters in this model as we later elaborate. We are interested in the time complexity of various "global" functions of G. We deal mostly with coloring and finding maximal independent sets.

Before we proceed we have to discuss symmetry-breaking. As is well known most functions cannot be computed in a distributed fashion by anonymous processors even for very simple graphs G. This impossibility usually results from symmetries of G. The common methods for breaking such symmetries are either randomization or the use of ID's. In this present paper we are only concerned with the latter. Thus we assume that there is a 1:1 mapping ID from the set of vertices V to the positive integers. In most cases the range of ID is $1, \ldots, |V|$. It is assumed that at time zero the processor occupying a node in G knows the ID of that node. Incidently, all our lower bounds hold even if every processor knows in advance what the graph G is, and only the labeling function ID is unknown.

It is clear that our model allows to compute every function of G in time $O(diameter(G))$. Because by such time every processor obtains complete knowledge of both G and ID. By storing at each processor a solution for the problem for every possible labeling we are done. Our concern is therefore only with time complexities below $diam(G)$. The major question we raise is which functions may be computed by a non trivial algorithm in this model, in the sense that they can be computed faster than $diam(G)$.

Here are our main results:

(1) Finding a maximal independent set in a labeled n-cycle, distributively, requires time $\Omega(log^* n)$. This bound is tight in view of the $O(log^* n)$ algorithm by Cole and Vishkin [CV]. Our proof relies on a new interesting construct of neighborhood graphs. A different proof using Ramsey Theorem was found by E. Tardos, L. Lovasz and B. Awerbuch (private communications).

(2) Coloring trees: Let T be the d-ary tree of height r. In time $\frac{2}{3}r$ it is impossible to color T in fewer than $\sqrt{d}$ colors. Note in contrast the algorithm by Goldberg and Plotkin [GP] who show that if every node in T knows who its father is, then a 3-coloring can be found in time $\log^* n$.

(3) In a labeled graph of order n with maximal degree Δ it is possible to find an $O(\Delta^2)$ coloring in time $O(\log^* n)$. This was previously shown in [GP] for the

0272-5428/87/0000/0331$01.00 © 1987 IEEE

case of constant Δ.

Our terminology is standard. Given a graph $G=(V,E)$ a *k-labeling* of G is a 1:1 mapping $f:V \to \{1,\ldots,k\}$. A k-labeling with $k=|V|$ is a *labeling*. A graph with a k-labeling is said to be *k-labeled* etc. Logarithms are to base 2. The k times iterated logarithm is denoted by $\log^{(k)}x$. That is $\log^{(1)}x := \log x$, and $\log^{(k)}x = \log(\log^{(k-1)}x)$. The least integer k for which $\log^{(k)}x < 1$ is denoted by $\log^* x$.

2. Lower bound on finding a maximal independent set in a cycle.

In [CV] a very nice algorithm was presented to find a maximal independent set of vertices (=MIS) in the n-cycle C_n in time $\log^* n$. In this section we show that this is optimal even in our model, where computation is for free. The algorithm presented in section 4 achieves this time bound in an alternative manner.

A basic observation is that in our model there is no loss of generality in assuming that processing proceeds by first collecting all data and then deciding. That is at time t each processor knows the labeling of all nodes which are at distance t or less from him. Also he knows all edges between these nodes, except for edges both endpoints of which are at distance exactly t from him. Note that no further information can reach him by time t.

Let us state our theorem:

Theorem 2.1: A synchronous distributed algorithm which finds a maximal independent set in a labeled n-cycle must take at least $\frac{1}{2}\log^* n - 4$ units of time.

Proof: The proof holds even under the assumption that there is a consistent notion of clockwise orientation common to all processors. Given an algorithm which finds a maximal independent set in the n-cycle where a notion of clockwise direction exists it is easily seen that in one more time step the cycle may be 3-colored. Our lower bounds in fact apply to 3-coloring.

Coming back to the observation we made before, at time t the data known to a processor P is an ordered list of $2t+1$ labels, starting t places before him, through his own and on to the next t labels. Let V be the set of all vectors $(x_1,\ldots,x_{2t+1})$ where the x_i are mutually distinct integers from $\{1,\ldots,n\}$. The algorithm is nothing but a mapping c from V into $\{1,2,3\}$.

Let us denote by $B_{t,n}$ the graph whose set of vertices is V. All edges of $B_{t,n}$ are given by:

$$(x_1,\ldots,x_{2t+1}) \ \ and \ \ (y,x_1,\ldots,x_{2t})$$

are neighbors for all $y \neq x_{2t+1}$. So $B_{t,n}$ has $n(n-1),\ldots,(n-2t)$ vertices and is regular of degree $2(n-2t-1)$.

We next point out that the mapping $c: V \to \{1,2,3\}$ is in fact a proper 3-coloring of $B_{t,n}$. To see this suppose that c maps

$$(x_1,\ldots,x_{2t+1}) \ \ and \ \ (y,x_1,\ldots,x_{2t})$$

to the same color. Then our algorithm for 3-coloring the n-cycle fails if the labeling happens to contain the segment

$$y,x_1,x_2,\ldots,x_{2t+1}.$$

The proof follows now by standard graph-theoretic arguments, since it can be shown that

$$\chi(B_{t,n}) = \Omega(log^{(2t)} n),$$

the $2t$ times iterated logarithm of n. Therefore, for $\chi(B_{t,n})$ to be at most 3 we must have $t = \Omega(log^* n)$.

To prove this lower bound on $\chi(B_{t,n})$ we introduce a very closely related family of digraphs $D_{s,n}$. The nodes of $D_{s,n}$ are all sequences $(a_1,\ldots,a_s)$ with the a_i mutually distinct and in the range $\{1,\ldots,n\}$. The out-neighbors of $(a_1,\ldots,a_s)$ are all vertices of the form $(a_2,\ldots,a_s,b)$ with $b \neq a_1$. Note that $B_{t,n}$ is the underlying graph of $D_{2t+1,n}$.

Given a digraph $H=(V,E)$ its *dilinegraph* $DL(H)$ is a digraph whose vertex set is E with (u,w) an edge if $head_H(u) = tail_H(w)$. The relation between the digraphs $D_{s,n}$ is given by:

Proposition 2.1: $D_{s,n}$ is obtained from $DL(D_{s-1,n})$ by removing a collection of disjoint cycles.

Proof: We identify the edge connecting $(x_1,\ldots,x_s)$ and $(x_2,\ldots,x_s,y)$ in $D_{s,n}$ with the vertex $(x_1,\ldots,x_s,y)$ in $V(D_{s+1,n})$. The adjacency relationship in $D_{s+1,n}$ is almost that in $DL(D_{s,n})$. The only exception is that in $DL(D_{s,n})$ there is an edge from $(x_1,\ldots,x_{s+1})$ to $(x_2,\ldots,x_{s+1},x_1)$ which misses in $D_{s+1,n}$. These exceptional edges are easily seen to form a set of disjoint cycles in $DL(D_{s,n})$. □

Our bound on $\chi(D_{s,n})$ is derived from the following simple propositions:

Proposition 2.2: If G_1 is a subgraph of G which is obtained by removing a collection of disjoint cycles, then

$$\chi(G_1) \geq \frac{1}{3}\chi(G).$$

Proof: A disjoint union of cycles can be 3-colored.

Take the product of a $\chi(G_1)$-coloring and a 3-coloring of the cycles in $E(G) \setminus E(G_1)$. This product is a proper $3 \cdot \chi(G_1)$ coloring of G. □

Proposition 2.3: For a digraph G

$$\chi(DL(G)) \geq \log \chi(G).$$

Proof: A coloring of $DL(G)$ may be thought of as a mapping $\psi: E(G) \to \{1, \ldots, k\}$ such that if $u, w \in E(G)$ and $head(u) = tail(w)$ then $\psi(u) \neq \psi(w)$. Now vertex color G by associating with node x the set

$$c(x) = \{ \psi(u) \mid x = tail(u) \}.$$

This is easily seen to be a 2^k vertex-coloring of G. Indeed if $u = (x,y) \in E(G)$ then $\psi(u) \in c(x)$, but $\psi(u)$ is not in $c(y)$ or else ψ is improper. Therefore $c(x) \neq c(y)$. □

Proposition 2.4: Define

$$\log_8^{(1)} x = \log_8 x \ , \log_8^{(j)} x = \log_8(\log_8^{(j-1)} x)$$

and $\log_8^* x$ to be the least j or which $\log_8^{(j)} x < 1$. Then

$$\log_8^* x > \log^* x - 4.$$

Proof: Set

$$a_0 = b_0 = 1 \ , a_{n+1} = 2^{a_n} \ , b_{n+1} = 8^{b_n}.$$

and show that $a_{n+4} > b_n$. In fact using standard estimates it is easily shown by induction on n that

$$a_{n+4} > 3b_n + \log 3 \sum_{j=0}^{\infty} \frac{(\frac{1}{3}\log e)^j}{b_n b_{n+1} \ldots b_{n+j}}$$

and the proposition follows. □

The theorem can be derived now. If a MIS can be found in time $t-1$ then we must have

$$3 \geq \chi(B_{t,n}).$$

But

$$\chi(B_{t,n}) = \chi(D_{2t+1,n}) \geq \log_8^{(2t)} n$$

So

$$2t \geq \log_8^* n - 2 \geq \log_2^* n - 6$$

$$t \geq \frac{1}{2}\log_2^* n - 3. \quad \square$$

In contrast with the low time complexity of 3-coloring, we can show that for even n, finding a 2-coloring of C_n requires time $\Omega(n)$.

Theorem 2.2: A synchronous distributed algorithm which 2-colors a labeled $2n$ cycle with labels from $\{1, \ldots, 2n\}$ must take at least $n-1$ units of time.

Proof: Now we must find the lowest t for which $B_{t,2n}$ is bipartite. But even for $t = n-2$ the graph $B_{t,2n}$ contains an odd cycle whose vertices are:

$$(i, \ldots, 2t+i) \quad (i = 1, \ldots, 2t+3)$$

(all integers taken modulu $2t+3$). The claim follows. □

Let us point out that for even n we just proved that finding a maximum independent set requires time $[\frac{n}{2}] - 1$. It is easily verified that the same is true for odd n as well.

We want to elaborate on the method developed here and point out its general features. Given a graph $G = (V,E)$ of order n its *t-neighborhood graph* $N_t(G)$ is constructed as follows: For every $x \in V$ let $S_t(x)$ be the subgraph of G spanned by those vertices y whose distance from x is at the most t. For every x consider all the n-labelings of $S_t(x)$.

Every such labeling ψ is a node in $N_t(G)$. Let $\psi_1: V_t(x) \to \{1, \ldots, n\}$ and $\psi_2: V_t(y) \to \{1, \ldots, n\}$ be two of these vertices of $N_t(G)$. They are taken to be neighbors in $N_t(G)$ if $[x,y] \in E(G)$ and there is a labeling $\Phi: V(G) \to \{1, \ldots, n\}$ such that

$$\Phi \mid_{S_t(x)} = \psi_1$$

and

$$\Phi \mid_{S_t(y)} = \psi_2.$$

Some easy observations regarding $N_t(G)$ are:

Proposition 2.5:

1) $\chi(N_t(G))$ is the least number of colors with which G may be colored distributively on time t.

2) $N_1(G) = K_n$ so $\chi(N_1(G)) = n$.

3) $\chi(N_t(G)) = \chi(G)$ for $t \geq diam(G)$.

4) $\chi(N_t(G))$ is nonincreasing with t.

5) For $G = C_n$, the graph $B_{t,n}$ is obtained from $N_t(C_n)$ by identifying vertices in $N_t(C_n)$ with identical sets of neighbors. In particular

$$\chi(B_{t,n}) = \chi(N_t(C_n)). \quad \square$$

As in the case of the cycle, it is helpful to identify nonadjacent vertices with an identical set of neighbors. Such an operation is called a *reduction*. Clearly it does not change the chromatic number while it may significantly simplify the graph. Neighborhood graphs seem to be very interesting and promising constructs. Unfortunately, the only case where we managed to calculate their graphical parameters is that of a cycle. Particularly interesting are situations where in G the graph $S_t(x)$ is independent of x. This is for example the case in vertex transitive graphs.

3. Lower bound on coloring trees.

Theorem 3.1: Let $T = T_{d,r}$ be the rooted d-regular tree of radius r. Any synchronous distributed algorithm which runs for time $\leq \frac{2}{3}r$ cannot color T by fewer than $\frac{1}{2}\sqrt{d}$ colors.

Proof: Lubotzky, Philips and Sarnak [LPS], recently showed how to construct d-regular graphs, $R_{d,n}$ on n vertices with girth $\geq \frac{4}{3}\frac{\log n}{\log(d-1)}$ and chromatic number $\geq \frac{1}{2}\sqrt{d}$. Let us remark that the bound on the chromatic number follows from Hoffman's estimate on the chromatic number in terms of the eigenvalues [H]. The LPS graphs are Ramanujan graphs, which means that all their eigenvalues but for the largest one are at most $2\sqrt{d-1}$ in absolute value and so Hoffman's bound applies directly. It follows that for $t < \frac{2}{3}\frac{\log n}{\log(d-1)}$ the graph $N_t(T_{d,r})$ contains a copy of a reduction of $N_t(R_{d,n})$. Therefore

$$\chi(N_t(T_{d,r})) \geq \chi(N_t(R_{d,n})) \geq \chi(R_{d,n}) \geq \frac{1}{2}\sqrt{d}$$

and so $T_{d,r}$ cannot be colored with fewer than $\frac{1}{2}\sqrt{d}$ colors in time t. The conclusion follows now on observing that for $T_{d,r}$

$$\frac{\log n}{\log(d-1)} \geq r. \quad \square$$

Two remarks are in order now: The lower bound of $\frac{1}{2}\sqrt{d}$ may be increased to $\Omega(\frac{d}{\log d})$ by using an appropriate random graph rather than Ramanujan Graphs (e.g. [B section 11.4]). We omit the details of this derivation. As we shall see in the next section, it is always possible to find for a d-regular graph an $O(d^2)$ coloring in time $O(\log^* n)$. The gap between $\frac{d}{\log d}$ and d^2 is quite intriguing and we tend to believe that the larger of the two is closer to the truth. This question is closely related to the complexity of finding a MIS distributively as we explain below.

It is well known now how to find a maximal independent set in a graph by means of an NC algorithm ([KW], [ABI], [L]). However none of these algorithms is both deterministic and distributive. In [GP] it is shown how to find a MIS in time $\log^* n$ for graphs of bounded degree. It is not clear what the situation is for unbounded degrees.

By a standard trick (e.g. [L]) it is known how to transform an efficient MIS algorithm to one for $\Delta+1$ coloring. Take the cartesian product of G with $K_{\Delta+1}$. It is easily verified that $(\Delta+1)$-colorings of G and MIS's of $G \times K_{\Delta+1}$ are in a natural 1:1 correspondence. It is therefore particularly interesting to find out the best time complexity in terms of n for finding a $(\Delta+1)$-coloring.

4. An $O(\log^* n)$ algorithm for $O(\Delta^2)$ coloring.

In this section we give some positive results on what can be done distributively. This can be interpreted either in terms of neighborhood graphs or algorithmically. What we present here is not a deterministic distributed algorithm in the strict sense of the word. We make use of the existence of a family of sets with certain properties. The existence of such families is proved indirectly and it is not known how to construct them. So when we say that a graph may be k-colored in time t deterministically in a distributed fashion this is to be interpreted as follows: The information on the labeling of the t-neighborhoods suffices to give a k-coloring. We suspect that some relevant methods for explicit constructions of similar families (e.g. [A] and [Bs]) are powerful enough to turn our results into deterministic algorithms in the strict sense. We plan to return to this in a later version of this article.

Theorem 4.1: Let G be a graph of order n and largest degree Δ. It is possible to color G with $3\Delta^2 \log n$ colors in one unit of time distributively. Equivalently

$$\chi(N_1(G)) \leq 3\Delta^2 \log n.$$

Proof: We need some combinatorial preparation:

Lemma 4.1: For integers $n > \Delta$ where is a family J of n subsets of $\{1, \ldots, \lceil \Delta^2 \log n \rceil\}$ such that if $F_0, \ldots, F_\Delta \in J$ then

$$F_0 \not\subseteq \bigcup_1^\Delta F_i$$

Proof: The existence of such families was considered in the literature ([KSS] and [EFF]) . But due to the simplicity of this lemma, we prove it here. Alternatively we could quote Theorem 3.1 from [EFF] which suffices for our purposes. To prove the lemma, let $m = \lceil \Delta^2 \log n \rceil$ and consider a random collection J of n subsets of $\{1, \ldots, m\}$. For $1 \leq x \leq m$ and $1 \leq i \leq n$ $Pr(x \in S_i) = \frac{1}{\Delta}$. All the events $x \in S_i$ are independent.

We claim that there is a selection of such a family for which the lemma holds. The probability that for a given $1 \leq x \leq m$ and given $F_0, \ldots, F_\Delta \in J$ there holds

$$x \in F_0 \setminus (\bigcup_1^{\Delta} F_i)$$

is

$$\frac{1}{\Delta}(1 - \frac{1}{\Delta})^{\Delta}$$

that is,

$$\frac{1}{e\Delta}(1 + o(1)).$$

Therefore, the probability that $F_0 \subseteq \bigcup_1^{\Delta} F_i$ is approximately $(1 - \frac{1}{e\Delta})^m$. There are $(\Delta+1)\binom{n}{\Delta+1}$ such choices of $F_0, \ldots, F_\Delta$. Therefore if

$$(1 - \frac{1}{e\Delta})^m (\Delta+1)\binom{n}{\Delta+1} < 1$$

then there is a selection of a family J which satisfies the lemma. Standard estimates show that this holds when

$$m > 3\Delta^2 \log n.$$

To prove our theorem now, we fix a family $J = \{F_1, \ldots, F_n\}$ as in the theorem. Consider a vertex colored i with neighbors colored $j_1, \ldots, j_d$ with $d \le \Delta$. Since

$$F_i \not\subseteq \bigcup_{\nu=1}^{d} F_{j_\nu}$$

there is an $1 \le x \le m$ with $x \in F_i \setminus (\bigcup_{\nu=1}^{d} F_{j_\nu})$. The new color of this vertex is x. It is easily verified that this is a proper m-coloring of G. □

If we iterate the coloring from theorem 4.1 $\log^* n$ times, we get a $7\Delta^2 \log\Delta$ coloring. To eliminate the $\log\Delta$ term we prove a result of the same type as Lemma 4.1 only in this range:

Lemma 4.2: Let q be a prime power. Then, there is a collection J of q^3 subsets of $\{1, \ldots, q^2\}$ such that if $F_0, \ldots, F_{[\frac{q-1}{2}]} \in J$ then $F_0 \not\subseteq \bigcup_1^{[\frac{q-1}{2}]} F_i$.

Proof: Use example 3.2 in [EFF] with $d = 2$.□

We select q to be the smallest prime power with $q \ge 2\Delta+1$. There is certainly one with $4\Delta+1 \ge q \ge 2\Delta+1$. This construction enables us then to transform an $O(\Delta^3)$ coloring to an $O(\Delta^2)$ coloring as in the proof of theorem 4.1 .

We conclude that:

Theorem 4.2: Let G be a labeled graph of order n with largest degree Δ. Then in time $O(\log^* n)$ it is possible to color G with $10\Delta^2$ colors in a distributive synchronous algorithm. □

We remark that proposition 3.4 of [EFF] sets a bound on our present method. They do not enable one to reach below $\binom{\Delta+2}{2}$ coloring. We do not know whether this is really impossible or that new ideas are called for.

References:

[A] N. Alon, Explicit construction of exponential sized families of k-independent sets, Discrete Mathematics 58 (1986) 191-193.

[ABI] N. Alon, L. Babai and A. Itai, A fast and simple randomized algorithm for the maximal independent set problem, J. of Algorithms, to appear.

[B] B. Bollobas, "Random Graphs", Academic Press 1985.

[Bs] N. Bshouti, Vector sets for exhaustive testing of logic circuits, M.Sc. Thesis, The Technion, Israel Inst. Tech. Haifa, Israel 1986.

[CV] R.Cole and U.Vishkin, Deterministic coin tossing and accelerating cascades: micro and macro techniques for designing parallel algorithms, Proc. 18th ACM Symp. on Theory of Computing 206-219, 1986.

[EFF] P.Erdos, P.Frankl and Z.Furedi, Families of finite sets in which no set is covered by the union of r others, Israel Journal of Mathematics, 51 (1985) 79-89.

[GP] A.V. Goldberg and S.A. Plotkin, Efficient parallel algorithms for $(\Delta+1)$ -coloring and maximal independent set problems, to appear in Proc. 19th ACM Symp. on Theory of computing, May 1987.

[H] A. J. Hoffman, Eigenvalues of graphs, in "Studies in Graph Theory" part II (D.R. Fulkerson ed.) MAA Studies in Math. 1975, pp. 225-245.

[KSS] D.J. Kleitman, J. Shearer and D. Sturtevant, Intersections of k-element sets, Combinatorica 1 (1981) 381-384.

[KW] R.M. Karp and A. Wigderson, A fast parallel algorithm for the maximal independent set problem, Proc. 16th ACM Symp. on theory of computing 266-272, 1984.

[L] M. Luby, A simple parallel algorithm for the maximal independent set problem, Proc. 17th ACM Symp. on Theory of Computing 1-10, 1985.

[LPS] A. Lubotzky, R. Phillips and P. Sarnak, Ramanujan graphs, Combinatorica, to appear.