SIAM J. COMPUT
Vol. 18, No. 2, pp. 268-277, April 1989

006

BOUNDS ON UNIVERSAL SEQUENCES*

AMOTZ BAR-NOY†‡, ALLAN BORODIN†§¶, MAURICIO KARCHMER†§, NATHAN LINIAL†, AND MICHAEL WERMAN†

Abstract. Universal sequences for graphs, a concept introduced by Aleliunas [M.Sc. thesis, University of Toronto, Toronto, Ontario, Canada, January 1978] and Aleliunas et al. [*Proc. 20th Annual Symposium on Foundation of Computer Science*, 1979, pp. 218–223] are studied. By letting $U(d, n)$ denote the minimum length of a universal sequence for d-regular undirected graphs with n nodes, the latter paper has proved the upper bound $U(d, n) = O(d^2n^3 \log n)$ using a probabilistic argument. Here a lower bound of $U(2, n) = \Omega(n \log n)$ is proved from which $U(d, n) = \Omega(n \log n)$ for all d is deduced. Also, for complete graphs $U(n-1, n) = \Omega(n \log^2 n/\log \log n)$. An explicit construction of universal sequences for cycles ($d = 2$) of length $n^{O(\log n)}$ is given.

Key words. universal sequences, graph connectivity, complexity theory

AMS(MOS) subject classifications. 05C40, 68R10, 68Q15

1. Introduction. In addition to their obvious computational interest, graph connectivity problems play a central role in complexity theory. Let STCON (respectively, USTCON) denote the problem of determining if a directed (respectively, undirected) graph has a path from a given source node s to a given goal node t. As usual, let NSPACE(S) (respectively, DSPACE (S) and RSPACE (S)) denote those sets accepted in nondeterministic (respectively, deterministic and random) space S. Savitch's [7] fundamental result that NSPACE $(S) \subseteq$ DSPACE (S^2) is based on the fact that STCON is complete for NSPACE ($\log n$) with respect to log-space reducibility. (In fact, it is complete with respect to log-depth $= \mathrm{NC}^1$ reducibility.) Similarly, Lewis and Papadimitriou [6] show that USTCON is complete for symmetric log-space bounded computation.

It is easy to see that STCON $\subseteq \mathrm{NC}^2 \subseteq$ DSPACE ($\log^2 n$). However, despite considerable attention to this problem, there has been no improvement to this upper bound. In one of the few significant attempts to give evidence that STCON is not contained in DSPACE ($\log n$), Cook and Rackoff [4] introduce the JAG (Jumping Automata for Graphs) model and show that within this restricted model STCON requires space $\Omega(\log^2 n/\log \log n)$.

Although USTCON appears to be a computationally easier problem (and indeed Cook and Rackoff [4] cannot prove such a strong result for JAGs operating on undirected graphs), the best known deterministic algorithms for USTCON also apply to STCON. However, when we consider random space bounded computations, the situation seems to be different, since Aleliunas et al. [2] show that USTCON is in RSPACE ($\log n$).

Motivated by the Cook and Rackoff [4] paper, Aleliunas [1] (for the special case of degree two) and then Aleliunas et al. [2] introduce the concept of a "universal sequence" for graphs. Let $G(d, n)$ denote the class of all connected d-regular graphs with n nodes and with labeled edges. Think of every edge as a pair of directed edges.

* Received by the editors July 15, 1986; accepted for publication (in revised form) June 20, 1988.

† Computer Science Department, Hebrew University, Jerusalem 91904, Israel.

‡ Present address, Computer Science Department, Stanford University, Stanford, California 94305-2140.

§ Present address, Department of Computer Science, University of Toronto, Ontario, Canada M5S 1A7.

¶ This research was performed while this author was a Lady Davis Visiting Professor at the Hebrew University, Jerusalem 91904, Israel.

Every directed edge is labeled with a label from $\{0, \cdots, d-1\}$ in such a way that edges going out from the same vertex are labeled differently. (It is easy to verify that for $2 \leqq d \leqq n-1$, $G(d, n)$ is not empty if and only if dn is even, see Lovász [5, Ex. 5.2.] A sequence $s = s_1 s_2 \cdots s_r$ in $\{0, \cdots, d-1\}^*$ is interpreted as a sequence of edge traversal commands. Thus a sequence s and a node u_0 on a graph G in $G(d, n)$ define a unique sequence of nodes $u_0, u_1, \cdots, u_r$ in G with (u_{i-1}, u_i) labeled by s_i for $i = 1, \cdots, r$. We say that s *visits* the set of nodes $\{u_0, \cdots, u_r\}$. A sequence s *covers* a graph G in $G(d, n)$ if the sequence visits every node in G independent of the starting node. A sequence s is *universal* for $G(d, n)$ if s covers every G in $G(d, n)$. Finally, we let $U(d, n)$ denote the minimal length of a universal sequence for $G(d, n)$. We will see in § 5 that the restriction to regular graphs serves some aesthetic purposes.

Aleliunas et al. [2] show that the expected time for a random walk to visit all nodes of $G = (V, E)$ is at most $2|E|\,|V|$. (No such result holds for directed graphs.) Hence the result that USTCON is in RSPACE $(\log n)$. They then use this result to assert the existence of a (nonuniform) universal sequence $s(d, n)$ for $G(d, n)$. The length of $s(d, n)$ is asymptotically bounded by

$$dn^2 \log(|G(d, n)|) = O(d^2 n^3 \log n).$$

In fact, they argue that most sequences of this length must be universal. Clearly, such universal sequences give a nonuniform method to test connectivity (using only two pebbles in the JAG model) within $O(\log n)$ space.

There are a number of reasons to study $U(d, n)$ further. If we could obtain a "sufficiently" explicit construction of polynomial length, then USTCON would be in DSPACE $(\log n)$. (We need to be able to generate any element of the sequence in DSPACE $(\log n)$.) In order to beat the previously mentioned $\log^2 n$ deterministic space bound, it suffices to show, by an explicit construction, that $U(3, n) = n^{o(\log n)}$. In this regard, we should also note that at present there is no deterministic sublinear space algorithm that runs in polynomial time. Second, for the purpose of time-space tradeoffs, it is important to determine the asymptotic behavior of $U(d, n)$ by any type of construction since lower bound techniques tend to apply to nonuniform models. In this regard an $U(d, n) = O(dn)$ or even $O(n^2)$ lower bound would have serious implications for any attempt at time-space lower bounds. In addition to complexity theory, universal sequences may play a role in the study of distributed systems (e.g., anonymous rings). And finally, of course, we think that the study of $U(d, n)$ raises a number of interesting combinatorial problems.

In §§ 2 and 3 we consider the special case of $d = 2$, the subject of Aleliunas [1]. First we give an explicit construction of length $n^{O(\log n)}$. Then we prove a nonlinear lower bound, $U(2, n) = \Omega(n \log n)$. Section 4 considers the other extreme, namely the case of complete graphs $(d = n-1)$. Here we observe that the probabilistic bound yields an upper bound of $n^3 \log^2 n$. We are able to prove that $U(n-1, n) = \Omega(n \log^2 n / \log \log n)$. In order to establish this lower bound, we view the problem as a game consisting of a graph generator (perhaps thought of as a taxi driver) versus a very powerful sequence generator (thought of as a passenger) where the passenger wants to see all n sites in as little time as possible and the driver would like to prolong the tour as long as possible. In § 5 we discuss the implication of the previous results for arbitrary d.

2. An explicit construction for the case of $d = 2$. There is only one regular connected graph of degree two, namely a cycle. In order to study $U(2, n)$, there is an equivalent way to formulate the problem as first discussed by Aleliunas in [1]. Instead of considering different labelings of the n-cycle, we consider the infinite line and label each integer coordinate (= node) with a "0" or "1" with the interpretation that at a node labeled "b" the edge to the right is labeled "b" and the edge to the left is labeled "$1-b$." We now interpret $U(2, n)$ as the smallest length of a sequence that is guaranteed to visit at least n nodes on any labeled line.

From the probabilistic constructions of Aleliunas [1] and Aleliunas et al. [2] we know that $U(2, n) = O(n^3)$. In fact, Aleliunas [1] conjectured $U(2, n)$ to be exactly $\binom{n}{2}$. Exhaustive tests confirm $U(2, n) = \binom{n}{2}$ for $n < 8$ but we are aware of at least one claim (again by testing) that $U(2, n) < \binom{n}{2}$ for $n = 9$.

To gain insight for both a lower bound and an explicit construction, we first consider a special class of labeled lines. Let ODD denote the class of labeled lines of the form $\cdots 0^{i_1}1^{i_2}0^{i_3}1^{i_4}\cdots$, where all i_j are odd. Let $L(n)$ be the class of all sequences that cover at least n nodes on any line in ODD and let $U(n)$ denote the minimal length of any sequence in $L(n)$. Without loss of generality, we shall assume that n is even in order to avoid ceilings and floors. Let $n' = n+1$ so that n' is odd.

LEMMA 1. *The sequence* $w_n = (0^{n'}1^{n'})^{(n/2+1)}$ *has the following properties*:

(A) *If begun on the leftmost node of a block labeled with zeros (respectively, on the rightmost node of a block labeled with ones)* w_n *will move right (respectively, left) until encountering the first block of nodes with an even number of zeros or ones wherein it will terminate on the leftmost zero or rightmost one of this block. If no such block is encountered within the first n nodes visited, the sequence will be exhausted having visited at least n nodes.*

(B) *If not started on a leftmost zero or rightmost one the directional behavior of* w_n *on the line will depend on the parity of the initial location within the block on which the sequence is started. In any case the sequence will either visit n nodes or will terminate on the leftmost zero or rightmost one of some block of even length. In particular,* $w_n \in L(n)$ *and* $U(n) \leqq (n+1)^2$.

LEMMA 2. *The sequence* v_n *defined recursively by* $v_1 = 01$ *and* $v_n = v_{n/2}(0^{n'}1^{n'})v_{n/2}$ *is in* $L(n)$ *so that* $U(n) = O(n \log n)$.

We leave it to the reader to verify both lemmas. Let us remark that Theorem 2 of the next section shows that Lemma 2 is asymptotically optimal. We use the w_n sequences repeatedly to explicitly construct a universal sequence of length $n^{O(\log n)}$. We chose to use the w_n sequences for ODD rather than the shorter $O(n \log n)$ sequences v_n since its properties are easier to state and since the shorter length v_n would not significantly change the length of the universal sequence of Theorem 1.

THEOREM 1. *There is a recursively defined sequence* $s(n)$ *that is universal for* $G(2, n)$ *with length* $|s(n)| = n^{O(\log n)}$. *Furthermore, any bit of* $s(n)$ *can be computed in time bounded by a polynomial in n.*

Proof. By induction on n we construct $s(n)$. The basis of the induction is immediate. Let w_n be as in Lemma 1 and let $s(n/2) = s_1 s_2 \cdots s_t$. Then, $s(n) = w_n s_1 w_n s_2 \cdots w_n s_t w_n$. Consider any labeled line and mark the leftmost zero and rightmost one in every even-length block. Note that in a segment of length n at most $n/2$ nodes have been marked. Now after the first w_n, $s(n)$ has either visited n nodes or has positioned itself on a marked node. Once on a marked node, a sequence symbol s_i of $s(n/2)$ will move either left or right so that the next w_n (by Lemma 1) will continue to move in that direction until it is stopped at the next marked node. In this way we are guaranteed to visit at least n nodes with some w_n or at least $n/2$ marked nodes and all the nodes within the blocks containing those marked nodes. In either case, at least n nodes have been visited.

To bound the length we see that $|s(n)| \leqq (n+1)^2 |s(n/2)|$ from which the length bound easily follows. It is also easy to see how to compute any particular bit of $s(n)$ in polynomial time. □

3. A lower bound for the case of $d=2$. The aim of this section is to prove an $\Omega(n \log n)$ lower bound for $U(2, n)$. We shall pick a small subset of ODD and show that, just to traverse this subset, a sequence must be "long."

We begin by introducing some notation. Let S be the set of infinite lines of the form: $\cdots 0^a 1^a 0^a 1^a \cdots$, a odd. Let S_n be the set of segments of length n of lines in S with starting point a leftmost zero. We show that a sequence that traverses every line in S_n must be $\Omega(n \log n)$ long. We say that a sequence α is *good* for a segment w if, when started at the left of w, α eventually reaches the right of w. For a string $w \in \{0, 1\}^*$ and $x \in \{0, 1\}$ let $\#_x w$ be the number of occurrences of x in w.

We illustrate the idea of our proof by considering some very simple sequences. So let α cover every line in S_n where α is of the form $0^{r_1}1^{r_1} \cdots 0^{r_k}1^{r_k}$ with all r_i odd. Assume, moreover, that α covers at least $n/2$ locations to the right of the starting point.

FACT 1. $0^{r_i}1^{r_i}$ is good for $0^a 1^a$ if and only if $r_i \geqq a$ and r_i is odd.

Define a_j to be the biggest odd number less than or equal to $n/(2j)$. Fact 1 implies that at least j r_i's must be bigger than a_j so that

$$|\alpha| \geqq \sum_{j=1}^{n/2} a_j \leqq \sum_{j=1}^{n/2} \left(\frac{n}{2j} - 2\right) = \Omega(n \log n).$$

We now give a lower bound for arbitrary α's. Consider the runs of a sequence α and the sequence 1α on a line from S. Since we start from a leftmost zero, these runs are symmetric with respect to the starting point. For a sequence β, let R_β (L_β) be the set of indices j such that β covers at least $n/2$ nodes to the right (left) of the starting point when run on the line $\cdots 0^{a_j}1^{a_j}0^{a_j}1^{a_j} \cdots$. Either $\sum_{j \in R_\alpha} a_j = \Omega(n \log n)$ or $\sum_{j \in L_\alpha} a_j = \Omega(n \log n)$. Since $R_\alpha = L_{1\alpha}$ and the lengths of α and 1α differ only by one we assume, without loss of generality, that $\sum_{j \in R_\alpha} a_j = \Omega(n \log n)$. We deal only with the runs of α on lines $\cdots 0^{a_j}1^{a_j}0^{a_j}1^{a_j} \cdots$, where $j \in R_\alpha$.

Fact 1 motivates the following lemma.

LEMMA 3. *Let α' be good for $0^a 1^a$, then $\alpha' = u_1 \beta_a^0 u_2 \beta_a^1 u_3$, where*

(C1) $\#_0 \beta_a^0 - \#_1 \beta_a^0 = a$, *and* $\#_1 \beta_a^1 - \#_0 \beta_a^1 = a$;

(C2) *Every nonempty prefix and suffix of $\beta_a^0(\beta_a^1)$ has more zeros (ones) than ones (zeros).*

Proof. (C1) is necessary in order to pass the block of zeros or ones. (C2) can be obtained by extending u_1 to the right and u_2 to the left so as to make β_a^0 minimal, and extending u_2 to the right and u_3 to the left to make β_a^1 minimal. Note that $\#_0 u_2 = \#_1 u_2$. □

We denote $\beta_a^0 u_2 \beta_a^1$ by β_a and call it an a-block. β_a^0 and β_a^1 are called half blocks.

LEMMA 4. *Let α' be good for $(0^a 1^a)^m$; then α' contains m disjoint a-blocks.* □

We say that two half blocks have a *trivial intersection* if they are either disjoint or one is contained in the other.

LEMMA 5. *Let β_{a_i}, β_{a_j} be an a_i-block and a_j-block respectively; then $\beta_{a_i}^0$ and $\beta_{a_j}^1$ have a trivial intersection.*

Proof. Follows by the prefix and suffix properties of $\beta_{a_i}^0$ and $\beta_{a_j}^1$. □

We say that a sequence $\{\beta_{a_j}^{x_j}\}_{j=1,\cdots,r}$; $x_j \in \{0, 1\}$, of half blocks is *nested* if we have the following.

(i) Every two half blocks have a trivial intersection;

(ii) $\beta_{a_j}^x \subseteq \beta_{a_k}^x$ implies that there exists an $l \neq j, k$ such that $\beta_{a_j}^x \subseteq \beta_{a_l}^{\bar{x}} \subseteq \beta_{a_k}^x$, $\bar{x}$ being the complement of x.

LEMMA 6. *Let α cover every line in S_n. Then α contains a nested sequence $\{\beta_{a_i}^{x_i}\}$, $i = 1, \cdots, n/2$ of half blocks where, again, a_j is the biggest odd number less than or equal to $n/(2j)$.*

Proof. The proof is by induction on i. Let B_{a_i} be a set of disjoint a_i-blocks in α. By Lemma 4 and the fact that $ia_i \leqq n$, $|B_{a_i}| \geqq i$. When $i=1$, pick any half block of any β_{a_1}. Assume the lemma is true for $i-1$. Let $B = \{\beta_{a_j}^{x_j}\}_{j=1,\cdots,i-1}$ be the nested sequence constructed up to step $i-1$. We will show how to find a half block of B_{a_i} that preserves the *nestedness* of B.

For each $\beta_{a_i} \in B_{a_i}$ define

$$In(\beta_{a_i}) = \{j | \beta_{a_i}^{x_j} \cap \beta_{a_j}^{x_j} \neq \varnothing \wedge \beta_{a_i}^{\bar{x}_j} \not\subseteq B_{a_j}^{x_j}\}.$$

Intuitively, $In(\beta_{a_i})$ is the set of half blocks that prevent β_{a_i} from being properly nested.

CLAIM 1. $\{In(\beta_{a_i})\}$, $\beta_{a_i} \in B_{a_i}$, are pairwise disjoint. To see this, suppose that $j \in In(\beta_{a_i}) \cap In(\tilde{\beta}_{a_i})$ for some β_{a_i} and $\tilde{\beta}_{a_i}$ and without loss of generality assume that β_{a_i} appears in α to the left of $\tilde{\beta}_{a_i}$. Furthermore, assume that $\beta_{a_j}^{x_j}$ is of type 0, i.e., $x_j = 0$. Then, either $\beta_{a_j}^0 \cap \beta_{a_i} = \varnothing$ or if $\beta_{a_j}^0 \cap \beta_{a_i} \neq \varnothing$, then $\beta_{a_i}^1 \subseteq \beta_{a_j}^0$ (because of suffix and prefix properties of the blocks). In either case, we get $j \notin In(\beta_{a_i})$. If $x_j = 1$, a similar argument shows that $j \notin In(\tilde{\beta}_{a_i})$.

CLAIM 2. There exists a β_{a_i} such that $In(\beta_{a_i}) = \varnothing$. This is true by Claim 1 and the fact that there are i a_i-blocks in B_{a_i}.

Choose a β_{a_i} as in Claim 2 and consider a minimal (in the inclusion sense) $\beta_{a_j}^{x_j}$ such that $\beta_{a_j}^{x_j} \cap \beta_{a_i} \neq \varnothing$. If no such $\beta_{a_j}^{x_j}$ exists then β_{a_i} is disjoint to every $\beta_{a_j}^{x_j}$ so that we can pick any half block of it. Otherwise, $\beta_{a_i}^{\bar{x}_j} \subseteq \beta_{a_j}^{x_j}$ and letting $x_i = \bar{x}_j$ we have that $\{\beta_{a_j}^{x_j}\}_{j=1,\cdots,i-1} \cup \{\beta_{a_i}^{x_i}\}$ is properly nested. □

LEMMA 7. *Let $B = \{\beta_{a_j}^{x_j}\}_{j=1,\cdots,i}$ be a nested sequence. Then*

$$\left| \bigcup_{j=1}^{i} \beta_{a_j}^{x_j} \right| \geqq \sum_{j=1}^{i} a_j.$$

Proof. Without loss of generality, assume that the half blocks in B are ordered so that $\beta_{a_i}^{x_i}$ is not contained in $\beta_{a_j}^{x_j}$ for $j = 1, \cdots, i-1$. We proceed by induction on i. The case $i=1$ is obvious. Assume the lemma is true for $i-1$ so that $|\bigcup_{j=1}^{i-1} \beta_{a_j}^{x_j}| \geqq \sum_{j=1}^{i-1} a_j$.

By (ii) in the definition of nested sequences, we have that the maximal $\beta_{a_j}^{x_j} \subseteq \beta_{a_i}^{x_i}$ are of opposite type, i.e., $x_i = \bar{x}_j$. Let β be the union of the $\beta_{a_j}^{\bar{x}_i}$, which are maximal half blocks contained in $\beta_{a_i}^{x_i}$. For $\beta_{a_i}^{x_i}$ to have (C1) we need $|\beta_{a_i}^{x_i} - \bigcup_{j=1}^{i-1} \beta_{a_j}^{x_j}| \geqq a_i$ so that $|\bigcup_{j=1}^{i} \beta_{a_j}^{x_j}| \geqq \sum_{j=1}^{i} a_j$. □

THEOREM 2. *A universal sequence α for S_n satisfies $|\alpha| = \Omega(n \log n)$.*

Proof. The proof follows immediately from Lemmas 6 and 7. □

4. The complete graph. There is only one graph in $G(n-1, n)$, namely K_n. While connectivity is no longer an issue, the question of $U(n-1, n)$ is still surprisingly difficult and interesting.

The probabilistic construction of Aleliunas et al. [2] shows that $U(n-1, n) = O(n^5 \log n)$ as the upper bound for any d. In fact, a more specific probabilistic analysis of random walks in K_n shows that the expected length to visit all nodes is $O(n \log n)$. From this follows $U(n-1, n) = O(n^3 \log^2 n)$.

For the lower bound, we consider the following two-player game, played between D (the driver) and P (the passenger). There is a fixed integer n and the game is with a taxi moving on a graph in $G(d, n)$. The game starts at any node of the graph. At each step, P can either direct the taxi along a directed edge that has already been traversed before, or he/she may let D choose any untraversed edge. In particular, if the present node is being visited for the first time, then D moves. The game ends at the first time when all nodes of the graph have been visited. P pays D the number of steps the game took.

We denote by $DP(d, n)$ the value of this game. Our result is Theorem 3.

THEOREM 3.

$$n^2-3n+3 \geqq DP(n-1, n) \geqq \Omega\left(\frac{n \log^2 n}{\log \log n}\right).$$

COROLLARY 1.

$$U(n-1, n) \geqq \Omega\left(\frac{n \log^2 n}{\log \log n}\right). \qquad \square$$

Proof of Theorem 3. The upper bound is obvious. P plays a strategy according to which he lets D play all the time. This insures that no edge is traversed twice in the same direction, so when $(n-1)(n-2)+1$ steps elapse all nodes must be visited.

To prove the lower bound, we consider a strategy for D that is defined inductively. At any time T in the game there is a digraph B_T on $V(K_n)$ consisting of all directed edges traversed thus far. The induction hypothesis follows:

(*) D has a strategy that causes the game to last at least $T(n) = (n \log^2 n)/(30 \log \log n)$ steps in such a way that all indegrees in B_T do not exceed $\log^2 n$.

We proceed to show how the strategy is carried over from n to $2n$. In the first stage D applies (*) to the first n nodes, thus he stays there at least $T(n) = (n \log^2 n)/(30 \log \log n)$ steps, with no indegree exceeding $\log^2 n$. At the first time after $T(n)$ at which D gets the right to move, he moves to node $n+1$. Now for another $T(n)$ steps he stays at nodes $\{n+1, \cdots, 2n\}$ according to (*). After these two stages, which take more than $2T(n)$ steps, no indegree exceeds $\log^2 n$ and node $2n+1$ is still isolated. Now begins a merging stage. To carry out the induction we show that for $(n \log n)/(5 \log \log n)$ steps D can proceed in the game with no indegree exceeding $\log^2(2n)$. Since $T(2n)-2T(n) \leqq (n \log n)/(5 \log \log n)$ the induction hypothesis is maintained. We claim the following easy lemma.

LEMMA 8. *Let G be a digraph, $S \subseteq V(G)$, and let $d \geqq 3$ be the largest indegree in G. Then there are at least $|V|/2$ nodes u for which all paths from u to S have length at least* $(\log |V| - \log |S| - 1)/\log d$.

Now let us consider the set S of all nodes of largest outdegree. Whenever D is given the move, he chooses to go to a node whose indegree is strictly less than $\log^2(2n)$ and whose distance in B_T from S is as large as possible. Note that the average indegree is at most $T(2n)/(2n) = (\log^2(2n))/(30 \log \log (2n))$ so that all but $2n/(30 \log\log(2n)) = o(n)$ nodes have indegree strictly less than $\log^2(2n)$. Thus during the whole process there are always many nodes with small indegree. In our case, all indegrees are at most $\log^2(2n)$ and it follows by Lemma 8 that for every node u and for at least half of the nodes v, the distance from v to u is at least $\log n/(5 \log \log n)$. Therefore, if $|S|=1$ then with this strategy the game proceeds at least $\log n/(5 \log \log n)$ steps more before the maximum outdegree increases. Ignoring momentarily the possibility that $|S|$ is larger, no degree reaches $2n$ before $(n \log n)/(5 \log \log n)$ steps. As long as the maximum outdegree is less then $2n$ the missing node will not be reached and the induction hypothesis is established since $T(2n)-2T(n) \leqq (n \log n)/(5 \log \log n)$.

To complete the proof for arbitrary $|S| \geqq 1$, consider the number of steps needed to increase the outdegree by two. We investigate a segment of the game during which the largest outdegree in B_T increases from m to $m+2$. Let us concentrate on that step where for the first time some outdegree reaches $m+2$ and let us say that at this point the number of nodes of degree $m+1$ is k. If $k \geqq (2 \log n)/(5 \log \log n)$ then our claim about the number of steps remains valid since the increase of the outdegree of any node requires at least one move (of D). On the other hand, if $k<(2\log n)/(5\log\log n)$ then by Lemma 8 at the beginning of the stage there is a node whose distance from all k nodes of outdegree $m+1$ is at least $\log(n/k)/\log(\log^2 n)>(2\log n)/(5\log\log n)$ and, all of our previous arguments carry through. Thus $DP(n)\geqq T(n)=(n\log^2 n)/(30\log\log n)$ and the proof is complete. □

5. Bounds for arbitrary graphs. Given the $U(2,n)=\Omega(n\log n)$ result, it is a little tedious but not difficult to derive the same (or improved as a function of n) lower bounds for all degrees. We again note that $G(d,n)$ is nonempty for $2\leqq d\leqq n-1$ if and only if dn is even, so that we always assume that dn is even.

Aleliunas [1] has shown that $U(2,n)\leqq U(d,(d-1)n)$. We modify his construction to obtain Lemma 9.

LEMMA 9. $U(2,n)\leqq(2/d)U(d,(d-1)n)$.

Proof. We show how to derive a universal sequence s' for $G(2,n)$ from a universal sequence s for $G(d,(d-1)n)$. For any $a,b\in\{0,1,\cdots,d-1\}$ let $s(a,b)$ be the sequence obtained from s by replacing each a by 0, each b by 1 and deleting all other symbols. Now let a, b be the least frequently occurring symbols in s and define $s'=s(a,b)$ so that $|s'|\leqq(2/d)|s|$. We will show that s' is universal for $G(2,n)$.

Let C be any labeled n-cycle. We want to construct a labeled graph G_C in $G(d,(d-1)n)$ whose traversal by s will guarantee that s' covers C. Let $K^i_{d-1}=(V^i,E^i)$ for $0\leqq i\leqq n-1$ be n copies of the complete graph K_{d-1}. Say $V^i=\{v^i_1,\cdots,v^i_{d-1}\}$. Then $G_C=\langle\bigcup_i V^i,\bigcup_i E^i\cup D\rangle$, where

$$D=\{(v^i_j,v^{(i+1)(\mathrm{mod}\,n)}_j)|0\leqq i\leqq n-1,1\leqq j\leqq d-1\}.$$

Intuitively, the K^i_{d-1} correspond to nodes in C while the edges in D correspond to the edges in C. We label the edges in E^i by any labeling from $\{0,1,\cdots,d-1\}\backslash\{a,b\}$. We label the edges in D in a way which corresponds exactly to the labeling in C. That is, if $\langle i,i+1\rangle$ has label "0" (respectively, "1") in C then for all j, $\langle v^i_j,v^{i+1}_j\rangle$ has label a (respectively, b) and $\langle v^i_j,v^{i-1}_j\rangle$ has label b (respectively, a). Clearly G_C is in $G(d,(d-1)n)$. Now it should be clear that as s traverses the graph G_C, it is precisely the labels $\{a,b\}$ that cause a traversal between neighboring (in the cycle) copies of K_{d-1}. Thus s covers G_c implies s' covers C. □

LEMMA 10. $U(d_1,n)\leqq(d_1/d_2)U(d_2,(d_2-d_1+1)n)$ *for all* $d_1\leqq d_2$.

Proof. This is an immediate generalization of the construction in Lemma 9. □

LEMMA 11. $U(d,n)=\Omega(n\log n-n\log d)$.

Proof. If n was divisible by $d-1$, this lemma would follow immediately from Lemma 9 and Theorem 2. For arbitrary sufficiently large $n=q(d-1)+r$ we proceed as follows. If $r=0$ then we would follow the construction of Lemma 9 and form a "cycle" with $q=n/(d-1)$ copies of K_{d-1}. If $r>0$ we form $q-3$ copies of $K^i_{d-1}=(V^i,E^i)$ and a set of nodes W with $|W|=n-(q-3)(d-1)=3(d-1)+r$. As before, each copy K^i_{d-1} will play the role of a node in a cycle as will W. We only have to describe how to fit W into a cyclic structure.

Suppose we want W to have K^1_{d-1} and K^{q-3}_{d-1} as its cyclic neighbors. Since dn is even, it follows that $d|W|$ is even and $|W|\geqq d+1$. Thus we can form a d-regular graph (W,E) and remove $d-1$ node-disjoint edges, say (u_j,w_j). We connect W to K^1_{d-1} and K^{q-3}_{d-1} by edges $\{(u_j,v^1_j)\}$ and $\{(w_j,v^{q-3}_j)\}$ for $1\leqq j\leqq d-1$. We connect K^i_{d-1} to K^{i+1}_{d-1} $(1\leqq i\leqq q-3)$ as in Lemma 9. In this way we are able to construct a d-regular graph G_C on n nodes. And again, as in Lemma 9, for any universal sequence s for $G(d,n)$ we construct $s'=s(a,b)$, where a and b are the least frequently occurring symbols in s. By labeling the "cycle" in G_C to mimic the labeling in C we can argue that if s is started on some node in $K^{\lceil(q-3)/2\rceil}_{d-1}$, then s' would cover at least $\lceil(q-3)/2\rceil$ nodes in

C. Since C is an arbitrary member of $G(2, n)$, this insures that s' covers at least $\lceil (q-3)/2 \rceil$ nodes in any infinite labeled line. Therefore $U(2, n/(3d)) \leqq U(2, (q-3)/2) \leqq (2/d)\, U(d, n)$ for any sufficiently large n which together with Theorem 2 yields Lemma 11. □

Lemma 11 shows that for small d (say $d \leqq n^{\varepsilon}$, $\varepsilon < 1$), we have $U(d, n) = \Omega(n \log n)$. For large d, there is a simple way to achieve the same bound using the driver-passenger game introduced in § 4.

LEMMA 12. $U(d, n) \geqq DP(d, n) = \Omega(n \log d)$.

Proof. The driver's strategy is simply first to form a directed cycle on n nodes with the first n labels. Then whenever the driver has a free choice he chooses to direct the new edge to the nearest nonadjacent node on the cycle. This continues until some node has degree d. On the ith tour of the cycle, the driver takes at least $\lfloor n/i \rfloor$ steps and the game continues for at least d tours. Thus,

$$DP(d, n) \geqq \sum_{i=1}^{d} \left\lfloor \frac{n}{i} \right\rfloor = \Omega(n \log d). \qquad \square$$

In terms of n, the largest known lower bound is obtained for $d = n/2 - 1$ by another simple driver-passenger game.

LEMMA 13. *For* $d \leqq n/2 - 1$, $U(d, n) \geqq DP(d, n) \geqq d(n-d)$.

Proof. The game will construct $G = (V, E)$ in $G(d, n)$ with $V = V_1 \cup V_2$, $|V_1| = n - d - 1$, $|V_2| = d + 1$. G is constructed so that a given universal sequence will visit only nodes in V_1 within the first $d(n-d-1)$ steps. This is simply achieved by thinking of the driver generating a d-regular graph on V_1 minus an edge. We complete the construction of G by choosing any complete graph on V_2 minus an edge with an arbitrary labeling. Now if (u_1, w_1) (respectively, (u_2, w_2)) is missing from V_1 (respectively, V_2), then G is the union of these graphs on V_1 and V_2 joined by the two edges (u_1, u_2) and (v_1, v_2). The driver forces the sequence (passenger) to stay on V_1 until $d(n-d-1)$ steps have expired and then trivially forces another d steps to cover the nodes of V_2. □

THEOREM 4. *For all* $d \leqq (n/2 - 1)$,

$$U(d, n) = \Omega((n \log n) + d(n-d)).$$

Proof. The proof is immediate from Lemmas 11, 12, and 13. □

Our final result emphasizes the importance of the case $d = 3$. Theorem 5 below is based on Theorem 4.13 of Cook and Rackoff [4]. Let $G'(d, n)$ denote the class of all graphs with n nodes and all degrees less than or equal to d labeled by $\{0, 1, \cdots, d-1\}$. In this case a sequence s in $\{0, 1, \cdots, d-1\}^*$ and a node u_0 in a graph G in $G'(d, n)$ uniquely defines a sequence of nodes $u_0, u_1, \cdots, u_{|s|}$ in G with (u_{i-1}, u_i) labeled s_i if u_{i-1} has an out edge so labeled and $u_i = u_{i-1}$ if u_{i-1} does not have an out edge labeled by s_i. And now, as before, let $U'(d, n)$ denote the minimal length of a universal sequence for $G'(d, n)$.

The following is a direct consequence of Theorem 4.13 in Cook and Rackoff [4].

LEMMA 14. *There is a finite state transducer computing a function*

$$f: \{0, 1, 2\}^* \rightarrow \{0, 1, \cdots, d-1\}^*$$

with the property that if s is universal for $G'(3, (2d-1)n)$ then $f(s)$ is universal for $G'(d, n)$. Furthermore, $|f(s)| \leqq (1/\log d)\, |s|$ so that

$$U'(d, n) \leqq \frac{1}{\log d}\, U'(3, (2d-1)n).$$

In order to place this result within the context of regular graphs we need to justify our introductory comment that regularity is not a significant restriction.

LEMMA 15. *$U'(d, n) \leqq U(d, d'n)$, where $d' = d$ if d is even and $d' = d+1$ if d is odd.*

Proof. For any G' in $G'(d, n)$ we construct a G in $G(d, d'n)$ by taking d' copies of G'. If x has degree $\delta < d$ in G' we connect all d' copies of x by a graph in $G(d-\delta, d')$ (a perfect matching if $\delta = d-1$). The new edges are labeled by the labels missing at x in G'. It is easily verified that a sequence that covers G covers G' as well. If all the degrees in G' are d and $d-1$ a slight modification is needed, which we omit. □

THEOREM 5.

$$U(d, n) \leqq \frac{1}{\log d} U(3, 2(2d-1)n).$$

Furthermore, there is a finite state transducer computing a function

$$f: \{0, 1, 2\}^* \to \{0, 1, \cdots, d-1\}^*$$

such that if s is universal for $G(3, 2(2d-1)n)$ then $f(s)$ is universal for $G(d, n)$.

Proof. The Cook and Rackoff construction on which Lemma 14 is based produces graphs where every node has degree two or three. In this case, we can take $d' = d-1 = 2$ in the construction of Lemma 15. The theorem then follows immediately from Lemmas 14 and 15. □

6. Conclusion. Perhaps the main technical result of this paper is the proof of a nonlinear lower bound for $d = 2$, thus answering a specific challenge in Aleliunas et. al. [2]. However it is clear that even for this restricted case we are far from understanding the true nature of $U(2, n)$. We know that the crucial aspect of labeled lines is the parity of the blocks. (We claim that, within a factor of n, we can assume that every block has length 1 or 2.) It seems feasible to us that some of the ideas developed here will lead to an explicit polynomial length universal sequence for $G(2, n)$. We also expect to be able to narrow the gap between the lower and upper bounds for $U(2, n)$.

For the complete graph, many obvious questions remain. It seems reasonable to be able to explicitly construct a "good" universal sequence. At present, we only know the brute force approach that gives $n^{n^2} \approx |G(n-1, n)|$. It is not difficult to see that a sequence universal for $G(n-1, n)$ will traverse at least n nodes when applied to members of $G(n, n+1)$. But we cannot see how to use this fact to construct such sequences. It also seems reasonable that we can narrow the gap for $\mathrm{DP}(n-1, n)$.

Theorem 5 emphasizes the importance of $U(3, n)$. In particular, any explicit universal sequence beyond brute force for $G(3, n)$ would be of interest. It will also be of interest if we could find for $d \geqq 3$ a simple d-ary infinite graph that would play the role that the infinite line played for $d = 2$.

Finally, there are many alternative universal sequence formulations that could be used for determining graph connectivity. One formulation we find particularly interesting is to number the nodes $V = \{1, \cdots, n\}$ and consider sequences in $\{1, \cdots, n\}^*$. Now, a sequence command i causes a move to node i if there is an edge from the currently scanned node to node i. Otherwise, it remains in the current node. Random walk arguments again show the existence of polynomial length universal sequences.

Note added in proof. Bridgland [*J. Algorithms*, 8 (1987), pp. 395-404] has given another construction for a universal sequence of length $n^{O(\log n)}$. His construction differs from ours. We were informed also of a construction by Barrington [private communication] for the same problem. Since the completion of this research in June

1986 there has been much activity in this area, and some of our results have been improved. Istrail [*Proc.* 20*th Annual ACM Symposium on Theory of Computing*, 1988, pp. 491-503] presents an explicit construction of a sequence of polynomial length that is universal for $d=2$. The ideas required go beyond the ones presented here. Karloff, Paturi, and Simon [unpublished manuscript] have given an explicit construction of a sequence universal for complete graphs whose length is $n^{O(\log n)}$. The construction of a polynomial length sequence even for complete graphs remains open. Alon and Ravid [*Discrete Appl. Math.*, to appear] have improved our lower bound for $U(n-1, n)$ to $n^2/\log n$. Their bound does not apply to our driver-passenger game. Also in the closely related area of random walks on graphs there has been considerable progress. A special issue of the *Journal of Theoretical Probability*, D. Aldous, ed., will be dedicated to the subject. A paper by Kahn, Linial, Nisan, and Saks that will appear therein shows that the expected cover time for regular graphs is only $O(n^2)$. This yields an improvement on the upper bound for $U(d, n)$ over the one in [2].

REFERENCES

[1] R. ALELIUNAS, *A simple graph traversing problem*, M.Sc. thesis, University of Toronto, Toronto, Ontario, Canada, January 1978.

[2] R. ALELIUNAS, R. M. KARP, R. J. LIPTON, L. LOVÁSZ, AND C. RACKOFF, *Random walks, universal traversal sequences, and the complexity of maze problems*, in Proc. 20th Annual Symposium on Foundations of Computer Science, 1979, pp. 218-223.

[3] M. F. BRIDGLAND, *Universal traversal sequences for paths and cycles*, J. Algorithms, 8 (1987), pp. 395-404.

[4] S. COOK AND C. RACKOFF, *Space lower bounds by maze threadability on restricted machines*, SIAM J. Comput., 9 (1980), pp. 636-652.

[5] L. LOVÁSZ, *Combinatorial Problems and Exercises*, North-Holland, Amsterdam, 1979. p. 249 (solution to exercise 5.2).

[6] H. R. LEWIS AND C. H. PAPADIMITRIOU, *Symmetric space bounded computation*, Theoret. Comput. Sci., 9 (1982), pp. 161-187.

[7] W. J. SAVITCH, *Relationships between nondeterministic and deterministic tape complexities*, J. Comput. System Sci., 4 (1970), pp. 177-192.