Discrete Mathematics 16 (1976) 329–337.

LOWER BOUNDS ON THE LENGTHS OF NODE SEQUENCES IN DIRECTED GRAPHS*

George MARKOWSKY
IBM Thomas J. Watson Research Center, Yorktown Heights, NY 10598, U.S.A.

Robert Endre TARJAN
Computer Science Department, Stanford University, Stanford, CA 94305, U.S.A.

Received 19 June 1975
Revised 16 December 1975

A strong node sequence for a directed graph $G = (N, A)$ is a sequence of nodes containing every cycle-free path of G as a subsequence. A weak node sequence for G is a sequence of nodes containing every basic path in G as a subsequence, where a basic path $n_1, n_2, \ldots, n_k$ is a path from n_1 to n_k such that no proper subsequence is a path from n_1 to n_k. (Every strong node sequence for G is a weak node sequence for G.) Kennedy has developed a global program data flow analysis method using node sequences. Kwiatowski and Kleitman have shown that any strong node sequence for the complete graph on n nodes must have length at least $n^2 - O(n^{7/4+\epsilon})$, for arbitrary positive ϵ. Every graph on n nodes has a strong sequence of length $n^2 - 2n + 4$, so this bound is tight to within $O(n^{7/4+\epsilon})$. However, the complete graph on n nodes has a weak node sequence of length $2n - 1$. In this paper, we show that for infinitely many n, there is a reducible flow graph G with n nodes (all with in-degree and out-degree bounded by two) such that any weak node sequence for G has length at least $\frac{1}{2}\log_2 n - O(n \log \log n)$. Aho and Ullman have shown that every reducible flow graph has a strong node sequence of length $O(n \log_2 n)$; thus our bound is tight to within a constant factor for reducible graphs. We also show that for infinitely many n, there is a (non-reducible) flow graph H with n nodes (all with in-degree and out-degree bounded by two), such that any weak node sequence for H has length at least cn^2, where c is a positive constant. This bound, too, is tight to within a constant factor.

1. Reducible flow graphs with long node sequences

Let $G = (N, A)$ be a directed graph. If $s \in N$, G is a *flow graph with start node s* if there is a path from s to any node in G. A flow graph G with start node s is *reducible* [2] if it can be reduced, by a sequence of applications of the following two transformations, to the graph $(\{s\}, \emptyset)$.

* The Research of the second author was partially supported by National Science Foundation Grant GJ-35604X1, by a Miller Research Fellowship at University of California; and by National Science Foundation Grant DCR72-03752 A02 at Stanford University.

T_1: Delete a loop (edge of the form (v, v)) from G.

T_2: If (v, w) is the only edge of the form (x, w) and $w \neq s$, delete w and all incident edges from G. For each deleted edge of the form (w, y) such that (v, y) is not an edge of G, add (v, y) as an edge of G.

We construct a family of reducible flow graphs $G(i, k)$ for $i \geq 1$, $k \geq 1$, which have long weak node sequences. Each $G(i, k)$ will have a distinguished start node $s(i, k)$ and a distinguished finish node $f(i, k)$. Let $G(i, k) = (N(i, k), A(i, k))$ be defined recursively by the following rules:

$$N(1, k) = \{s(1, k)\}, \quad A(1, k) = \emptyset, \quad f(1, k) = s(1, k),$$

$$\begin{aligned} N(i+1, k) = {} & (N(i, k) \times \{1, 2\}) \\ & \cup \{s(i+1, k), t(i+1, k), u(i+1, k), v(i+1, k), w(i+1, k)\} \\ & \cup \{x(i+1, k, j)\colon i \leq j \leq k\}, \end{aligned}$$

$$\begin{aligned} A(i+1, k) = {} & \{((y, j), (z, j))\colon (y, z) \in A(i, k), j \in \{1, 2\}\} \\ & \cup \{(s(i+1, k), (s(i, k), 1)), ((s(i, k), 1), t(i+1, k)), \\ & (t(i+1, k), u(i+1, k)), ((f(i, k), 1), u(i+1, k)), \\ & ((f(i, k), 1), v(i+1, k)), (u(i+1, k), (s(i, j), 2)), \\ & ((f(i, k), 2), w(i+1, k)), ((f(i, k), 2), s(i+1, k)), \\ & (v(i+1, k), x(i+1, k, 1)), (w(i+1, k), x(i+1, k, 1))\} \\ & \cup \{(x(i+1, k, j), x(i+1, k, j+1))\colon 1 \leq j < k\}, \end{aligned}$$

$$f(i+1, k) = x(i+1, k, k).$$

Fig. 1 illustrates $G(i+1, k)$, which is formed by appropriately combining two copies of $G(i, k)$.

Lemma 1.1. *For each i and k, all vertices in $G(i, k)$ have in-degree and out-degree at most two, $s(i, k)$ has in-degree and out-degree at most one, and $f(i, k)$ has out-degree zero.*

Proof. Easy by induction on i.

Lemma 1.2. *For each i and k, $G(i, k)$ is a reducible flow graph.*

Proof. It is easy to prove by induction on i that every vertex in $G(i, k)$ is reachable from $s(i, k)$. We prove by induction on i that $G(i, k)$ is reducible. $G(1, k)$ is reducible by definition. Suppose $G(i, k)$ is reducible. Then $G(i+1, k)$ can be reduced in the following way: Reduce the first copy of $G(i, k)$ in $G(i+1, k)$ to the single node $(s(i, k), 1)$. Reduce the second copy of $G(i, k)$ in $G(i+1, k)$ to the

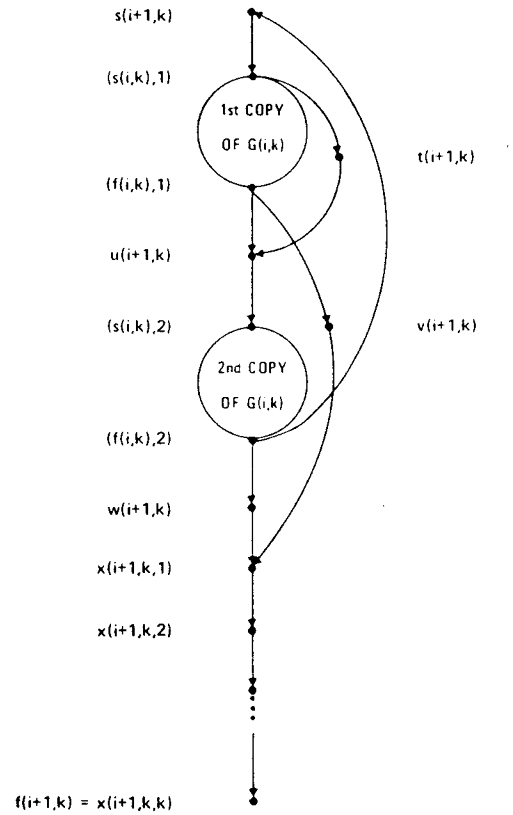

Fig. 1.

single node $(s(i, k), 2)$. Delete the following nodes in order using T_2, applying T_1 to remove loops as they are created:

$$\begin{gathered} x(i+1, k, k), x(i+1, k, k-1) \ldots x(i+1, k, 2), w(i+1, k), v(i+1, k), \\ t(i+1, k), (s(i, k), 1), u(i+1, k), (s(i, k), 2), x(i+1, k, 1). \end{gathered}$$

This reduces $G(i+1, k)$ to $(\{s(i+1, k)\}, \emptyset)$.

Let $n(i, k) = |N(i, k)|$ and $a(i, k) = |A(i, k)|$. The following equations follow from the definitions of $N(i, k)$ and $A(i, k)$:

$$n(1, k) = 1, \quad n(i+1, k) = 2n(i, k) + (5 + k),$$

$$a(1, k) = 0, \quad a(i+1, k) = 2a(i, k) + (9 + k).$$

Lemma 1.3. $n(i, k) = (6 + k)2^{i-1} - (5 + k)$ *and* $a(i, k) = (9 + k)2^{i-1} - (9 + k)$.

Proof. By induction:

$$n(1) = 1 = (6+k)\cdot 2^0 - (5+k),$$
$$n(i+1) = 2n(i,k) + (5+k)$$
$$= 2[(6+k)2^{i-1} - (5+k)] + (5+k)$$
$$= (6+k)2^i - (5+k);$$

and similarly for $a(i,k)$.

Lemma 1.4. *In $G(i,k)$ there is a basic path $p(i,k)$ from $s(i,k)$ to $f(i,k)$ containing $m(i,k) = (4+k)2^{i-1} - (3+k)$ nodes.*

Proof. Define $p(i,k)$ recursively as follows:

$$p(1,k) = s(1,k),$$
$$p(i+1,k) = s(i+1,k), p(i,k)\times\{1\}, u(i+1,k), p(i,k)\times\{2\}, w(i+1,k),$$
$$x(i+1,k,1), x(i+1,k,2), \ldots, x(i+1,k,k)$$

Note: $p(i,k)$ is a sequence of nodes in $G(i,k)$; if $p(i,k) = y_1, \ldots, y_m$, then $p(i,k)\times\{j\}$ denotes the sequence of nodes $(y_1,j), (y_2,j), \ldots, (y_m,j)$ in $G(i+1,k)$.

It is clear from Fig. 1 that if $p(i,k)$ is a basic path, so is $p(i+1,k)$. Furthermore, if $p(i,k)$ contains $m(i,k)$ vertices, then $m(1,k) = 1$ and $m(i+1,k) = 2m(i,k) + (3+k)$. We can prove by induction that $m(i,k) = (4+k)2^{i-1} - (3+k)$.

By a *restricted node sequence* for $G(i,k)$ we mean a sequence L of nodes such that every basic path in $G(i,k)$ *ending at* $f(i,k)$ is a restricted node sequence. Let $l(i,k)$ be the minimum number of nodes in a restricted node sequence for $G(i,k)$.

Lemma 1.5. $l(i,k) \geqslant (i-1)(4+k)\cdot 2^{i-2} + 1$.

Proof. First we show

(a) $l(1,k) \geqslant 1$,

(b) $l(i+1,k) \geqslant 2l(i,k) + m(i,k) + 2 + k$.

Clearly (a) holds. To prove (b), suppose L is any restricted node sequence for $G(i+1,k)$. Then L contains as disjoint subsequences restricted node sequences for the two copies of $G(i,k)$ contained in $G(i+1,k)$. Let L_1 be the restricted node sequence for

$$G(i,k)\times\{1\} = (N(i,k)\times\{1\}, \{((y,1),(z,1)) \colon (y,z)\in A(i,k)\})$$

which ends earliest in L. Similarly, let L_2 be the restricted node sequence for

$$G(i,k)\times\{2\} = (N_i\times\{2\}, \{((y,2),(z,2)) \colon (y,z)\in A(i,k)\})$$

which ends earliest in L.

For each basic path in $G(i,k)$ which ends at $f(i,k)$ there is a basic path in $G(i+1,k)$ consisting of a copy of this path in $G(i,k)\times\{1\}$ followed by $u(i+1,k)$ followed by a copy of $p(i,k)$ in $G(i,k)\times\{2\}$ followed by $w(i+1,k), x(i+1,k,1), \ldots, x(i+1,k,k)$. Thus the last node in L_1 must be followed in L by $u(i+1,k)$ copies in $G(i,k)\times\{2\}$ of the nodes in $p(i,k)$, $w(i+1,k)$, $x(i+1,k,1), \ldots,$ and $x(i+1,k,k)$. Similarly the last node in L_2 must be followed by $s(i+1,k)$, copies in $G(i,k)\times\{1\}$ of the nodes in $p(i,k)$, $v(i+1,k)$, $x(i+1,k,1), \ldots,$ and $x(i+1,k,k)$.

Thus L consists at least of L_1, L_2, and $m(i,k)+2+k$ additional nodes [$u(i+1,k)$, $m(i,k)$ nodes in $G(i,k)\times\{2\}$, $w(i+1,k)$, $x(i+1,k,1), \ldots, x(i+1,k,k)$ if L_1 ends after L_2; $s(i+1,k)$, $m(i,k)$ nodes in $G(i,k)\times\{1\}$, $v(i+1,k)$, $x(i+1,k,1), \ldots, x(i+1,k,k)$ if L_2 ends after L_1]. This gives (b).

Using (a) and (b) we can prove the lemma by induction:

$$l(1) \geqslant 1 = (0)(4+k)\cdot 2^{-2} + 1,$$
$$l(i+1,k) \geqslant 2l(i,k) + m(i,k) + 2 + k$$
$$\geqslant 2l(i,k) + (4+k)2^{i-1} - 1$$
$$\geqslant (i-1)(4+k)2^{i-1} + 2 + (4+k)2^{i-1} - 1$$
$$= i(4+k)2^{i-1} + 1.$$

Theorem 1.6. *For infinitely many n, there are reducible flow graphs with n nodes (all of in-degree and out-degree two or less) having no weak node sequences of length less than $\frac{1}{2}n\log_2 n - O(n\log\log n)$.*

Proof. For each n of the form $n = (6+i)2^{i-1} - (5+i)$, $G(i,i)$ is a reducible flow graph satisfying the conditions of the theorem. No weak node sequence for $G(i,i)$ has length less than

$$l(i,i) \geqslant (i-1)(4+i)2^{i-2} + 1.$$

Since $2^{i-1} = (n+5+i)/(6+i)$,

$$\ell(i,i) \geqslant \tfrac{1}{2}n\frac{(4+i)}{(6+i)}(i-1) \geqslant \tfrac{1}{2}n\frac{(i-2)}{i}(i-1)$$
$$\geqslant \tfrac{1}{2}ni - O(n).$$

Also,

$$i - 1 = \log_2[n + (5+i)] - \log(6+i)$$
$$\geqslant \log_2 n - O(\log\log n).$$

Hence

$$l(i,i) \geqslant \tfrac{1}{2}n\log_2 n - O(n\log\log n)$$

and the theorem holds.

Thus the Aho–Ullman bound is tight to within a constant factor.

2. Non-reducible flow graphs with long node sequences

We will construct a family of non-reducible sparse graphs which have long weak node sequences. First we show that we can use graphs with long strong node sequences and high in-degree and out-degree for our examples. Let G be any flow graph. Let G' be constructed from G using the following rule.

(a) Delete each arc (v, w) of G and replace it by a new node x and two new arcs (v, x) and (x, w). Repeat until all of G's original arcs are replaced.

Any elementary path $v_1, v_2, \ldots, v_k$ of G corresponds to (and is contained as a subsequence in) some basic path $v_1, x_1, v_2, x_2, \ldots, v_{k-1}, x_{k-1}, v_k$ of G'. This fact gives the following lemma.

Lemma 2.1. *Let G be a flow graph with n nodes and e arcs. There is a flow graph G' with $n + e$ nodes and $2e$ arcs such that any weak node sequence for G' is a strong node sequence for G.*

Let G'' be formed from G' using the following two rules:

(b) For each node v with three or more exiting arcs, say (v, a), (v, b), (v, c), delete two of these arcs, say (v, a) and (v, b), and replace them by a new node x and three new arcs (v, x), (x, a), (x, b). Repeat until all nodes have out-degree two or less.

(c) For each node v with three or more entering arcs, say (a, v), (b, v), (c, v), delete two of these arcs, say (a, v) and (b, v), and replace them by a new node x and three new arcs (v, x), (x, a), (x, b). Repeat until all nodes have in-degree two or less.

Any basic path of G' corresponds to (and is contained as a subsequence in) some basic path of G''. This gives the next lemma.

Lemma 2.2. *Let G be a flow graph with n nodes and e arcs. There is a flow graph G'' with at most $n + 3e$ nodes and $4e$ arcs, such that all nodes of G'' have in-degree and out-degree at most two, and any weak node sequence for G'' is a strong node sequence for G.*

The next result is crucial to the construction. Let $G = (N_1, A_1)$ be a graph with a distinguished start vertex s and a distinguished finish vertex f. By a *doubly restricted node sequence* for G we mean a sequence L of nodes such that every elementary path in G starting at s and ending at f is a subsequence of L. Every strong node sequence for G contains a doubly restricted node sequence. Let $H = (N_2, A_2)$ be any other directed graph. Let $G \otimes H = (N_3, A_3)$ be the directed graph given by

$$N_3 = N_1 \times N_2,$$

$$A_3 = \{((x, z), (y, z)): (x, y) \in A_1, z \in N_2\} \cup \{((f, y), (s, z)): (y, z) \in A_2\}.$$

Theorem 2.3. *Let l_1 be the minimum length of a doubly restricted node sequence for G. Let l_2 be the minimum length of a strong node sequence for H. Then every strong node sequence for $G \otimes H$ has length at least $l_1 \cdot l_2$.*

Proof. Let $L = w_1, w_2, \ldots, w_l$ be any strong node sequence for $G \otimes H$. From L we derive a strong node sequence L_H for H. Each node in L will correspond either to one node in L_H or to no nodes in L_H.

We define a sequence $z_1, \ldots, z_l$ from left-to-right. Suppose $z_1, \ldots, z_{j-1}$ have been defined. Consider $w_j = (x, z)$ and let $w_i = (y, z)$ be the node in L with $i < j$, i maximum, such that $z_i = z$ (let $i = 0$ if there is no such w_i). Let $z_j = z$ if the subsequence of L from w_{i+1} to w_j (inclusive) contains a doubly restricted node sequence of

$$G \times \{z\} = (N_1 \times \{z\}, \{((u, z), (v, z)): (u, v) \in A_1\}).$$

Otherwise let $z_j = \emptyset$.

The sequence $z_1, \ldots, z_l$ contains nodes from H and occurrences of $\emptyset$. Let L_H be formed from $z_1, \ldots, z_l$ by deleting all occurrences of $\emptyset$. Each node in L_H corresponds to a subsequence of L which is a doubly restricted node sequence for some copy of G. By the construction all these subsequences are disjoint. Thus, if l' is the length of L_H, $l \geq l_1 \cdot l'$.

Now all we must show is that L_H is a strong node sequence for H. Let $x_1, \ldots, x_k$ be any elementary path in H. Recursively define indices $a_1, b_1, a_2, b_2, \ldots, a_k, b_k$ as follows:

$$a_1 = 1,$$

$$a_{i+1} = b_i + 1,$$

b_i is the first position j such that the subsequence of L from w_{a_i} to w_j (inclusive) contains a doubly restricted node sequence of $G \times \{x_i\}$.

If it were not possible to define all the a_i, b_i, then we could construct from $x_1, \ldots, x_k$ by replacing each node x_i with an appropriate path from (s, x_i) to (f, x_i) in $G \times \{x_i\}$, a path in $G \otimes H$ which was not a subsequence of L. But L is a strong node sequence for L. Thus all the a_i, b_i can be defined. But by the construction of $z_1, \ldots, z_l$, some z_j with $a_i \leq j \leq b_i$ must have $z_j = x_i$. This is true for each i; thus $x_1, \ldots, x_k$ is a subsequence of L_H. Hence L_H is a strong node sequence for H, $l' \geq l_2$, and $l \geq l_1 \cdot l_2$.

Now we construct a family of non-reducible sparse graphs $H(i)$. Each $H(i)$ will have a distinguished start node s_i and a distinguished finish node f_i. Let $H(i) = (N(i), A(i))$ be defined recursively by the following rules:

$$N(1) = \{s_1\}, \quad A(1) = \emptyset, \quad f_1 = s_1,$$

$$N(i + 1) = (N(i) \times \{1, 2, \ldots, 2^i\}) \cup \{s_{i+1}, f_{i+1}\},$$

$$A(i + 1) = \{((y, j), (z, j)): (y, z) \in A(i) \text{ and } 1 \leq j \leq 2^i\}$$

$$\cup \{(s_{i-1}, (s_i, j)) : 1 \leq j \leq 2^i\}$$
$$\cup \{((f_i, j), f_{i+1}) : 1 \leq j \leq 2^i\}$$
$$\cup \{((f_i, j), (s_i, k)) : 1 \leq j \leq 2^i, 1 \leq k \leq 2^i\}.$$

Let $n(i) = |N_i|$ and $a(i) = |A_i|$. The following equalities are immediate from the definitions above:

$$n(1) = 1, \quad n(i+1) = 2^i n(i) + 2,$$
$$a(1) = 0, \quad a(i+1) = 2^i a(i) + 2^{2i} + 2^{i+1}.$$

Let $m(i) = 2^{\binom{i}{2}}$; $m(i)$ is defined recursively by $m(1) = 1$, $m(i+1) = 2^i m(i)$.

Lemma 2.4. *There are positive constants c_1, c_2 such that $n(i) \leq c_1 m(i)$, $a(i) \leq c_2 m(i)$ for all i.*

Proof.

$$\frac{a(i+1)}{m(i+1)} = \frac{2^i a(i) + 2^{2i} + 2^{i+1}}{2^i m(i)} = \frac{a(i)}{m(i)} + \frac{2^i + 2}{m(i)}$$
$$= \frac{a(i)}{m(i)} + (2^i + 2)/2^{\binom{i}{2}}.$$

Thus

$$\frac{a(i)}{m(i)} \leq \frac{a(1)}{m(1)} + \sum_{i=1}^{\infty} (2^i + 2)/2^{\binom{i}{2}} \leq c_2$$

for a suitable constant c_2. A similar argument works for $n(i)$.

Lemma 2.5. *Let $l(i)$ be the minimum number of nodes in a doubly restricted node sequence for $H(i)$. Then $l(i) \geq c_3 n(i)^2$ for some suitable positive constant c_3.*

Proof. The graph $H(i+1)$ contains as a subgraph the graph $H(i) \otimes C(2^i)$, where $C(2^i)$ is the complete directed graph on 2^i vertices. Furthermore, every elementary path in $H(i) \otimes C(2^i)$ is a subsequence of an elementary path in $H(i+1)$ starting at s_{i-1} and ending at f_{i+1}. By Theorem 2.3 and the Kwiatowski–Kleitman result [5], picking $\varepsilon = \frac{1}{8}$, there is a positive constant c_4 such that $l(i+1) \geq (2^{2i} - c_4 2^{(15/8)i}) l(i)$. Then

$$\frac{l(i+1)}{[m(i+1)]^2} \geq \frac{(2^{2i} - c_4 2^{(15/8)i}) l(i)}{2^{2i} m(i)^2}.$$

Thus for any $i_0 < i$,

$$\frac{l(i)}{[m(i)]^2} \geq \left(\prod_{j=i_0}^{i} \left[1 - \frac{c_4}{2^{j/8}}\right] \right) \frac{l(i_0)}{[m(i_0)]^2} \geq \left(\prod_{j=i_0}^{\infty} \left[1 - \frac{c_4}{2^{j/8}}\right] \right) \frac{l(i_0)}{[m(i_0)]^2}$$
$$\geq \left(\prod_{j=i_0/8}^{\infty} \left[1 - \frac{c_4}{2^j}\right]^8 \right) \frac{l(i_0)}{[m(i_0)]^2}.$$

Choose i_0 such that $c_4/2^{i_0/8} \leq \frac{1}{2}$. Then

$$\ln \prod_{j=i_0/8}^{\infty} \left[1 - \frac{c_4}{2^j}\right]^8 = 8 \sum_{j=i_0/8}^{\infty} \ln \left[1 - \frac{c_4}{2^j}\right] = \sum_{j=i_0/8}^{\infty} \sum_{k=1}^{\infty} -\frac{1}{k} \left(\frac{c_4}{2^j}\right)^k$$
$$\geq 8 \sum_{j=i_0/8}^{\infty} -\frac{c_4}{2^{j-1}}$$
$$\geq -16.$$

Thus

$$\frac{l(i)}{[m(i)]^2} \geq e^{-16} \frac{l(i_0)}{[m(i_0)]^2}, \qquad l(i) \geq c_5 [m(i)]^2$$

for all $i > i_0$ and a suitable positive constant c_5. Using the fact that $l(i)$ is always positive and applying Lemma 2.4 gives the desired result.

Theorem 2.6. *For infinitely many values of a, there exists a graph H with a arcs such that H has no strong node sequences of length less than $c_6 a^2$, for a suitable positive constant c_6.*

Proof. Immediate from Lemma 2.5.

Lemmas 2.1 and 2.2 give the following corollary.

Corollary 2.7. *For infinitely many values of n, there is a graph with n nodes (all with in-degree and out-degree bounded by two) such that no weak node sequences of the graph have length less than $c_7 n^2$, for a suitable positive constant c_7.*

References

[1] A.V. Aho, and J.D. Ullman, Node listings for reducible flow graphs, Proc. Seventh ACM Symposium on Theory of Computing (1975) 177–185.

[2] M.S. Hecht and J.D. Ullman, Characterizations of reducible flow graphs, J. ACM 21 (3) (1974) 367–375.

[3] K.W. Kennedy, Node listings applied to data flow analysis, Proc. Second ACM Conference on Principles of Programming Languages (1975) 10–21.

[4] P.J. Koutas and T.C. Hu, Shortest string containing all permutations, Discrete Math. 11 (2) (1975) 125–132.

[5] D.J. Kwiatowski and D.J. Kleitman, A lower bound on the length of a sequence containing all permutations as subsequences, J. Combin. Theory, Ser. A, to appear.

[6] G. Markowsky, Lower bounds on lengths of basic and elementary node sequences for reducible flow graphs, Technical Report RC-5330, IBM Thomas J. Watson Research Center, Yorktown Heights, NY (1975).

[7] M. Newey, Note on a problem involving permutations as subsequences, STAN-CS-73-340, Department of Computer Science, Stanford University, Stanford, CA (1973).