JOURNAL OF COMPUTER AND SYSTEM SCIENCES 12, 134–146 (1976)

Complexity Measures for Regular Expressions

ANDRZEJ EHRENFEUCHT AND PAUL ZEIGER

Department of Computer Science, University of Colorado, Boulder, Colorado 80302

Received July 15, 1974

Several measures of the complexity of a regular expression are defined. (Star height and number of alphabetical symbols are two of them.) Upper and lower estimates for the complexities of expressions for certain sets of paths on graphs are derived.

1. INTRODUCTION

Some of our colleagues have considered using a regular expression that represents sequences of instructions executed by a program as a tool in static analysis of the program. The question of how large or complex one might expect such an expression to be naturally arose. This paper is intended to shed some light on the issue.

The four measures of expression complexity considered are:

N = size, the number of alphabetical symbols;

H = star height, the depth of nesting of stars;

L = length, the length of the longest nonrepeating path through the expression;

W = width, the maximum number of symbols unioned (dual to L).

TABLE I

Measure	Alphabetical symbol	$E \cup F$	$E \cdot F$	E^*
N	1	$N(E) + N(F)$	$N(E) + N(F)$	$N(E)$
H	0	$\max(H(E), H(F))$	$\max(H(E), H(F))$	$H(E) + 1$
L	1	$\max(L(E), L(F))$	$L(E) + L(F)$	$L(E)$
W	1	$W(E) + W(L)$	$\max(W(E), W(F))$	$W(E)$

See Table I for inductive definitions of these measures. The complexity of a regular *set* with respect to any of these measures is taken to be the minimal measure over all expressions for that set. The sets examined are:

1. the set of all paths between two specified nodes on the complete graph on n nodes, where "complete" means there is a directed arc between every node pair (self-loops included), and each arc bears a distinct label;

2. the set of all paths of length $\leqslant k$ arcs between two specified nodes on the above graph; and

3. the set of all paths from node 1 to node n on the graph that has a distinctly labeled, directed arc running from every node to every higher-numbered node, hereafter called the *half-complete graph.*

All logarithms are to the base 2. We are disappointed that so many of the lower bounds are worse than polynomial. For the first set we find:

(1) $2^{(n-1)} \leqslant N$, and (2) $N \leqslant 6 \cdot 4^{(n-2)}$;

(3) $H = n$;

(4) $W \leqslant n(2n+1)^{\lceil \log(n/2) \rceil}$;

(5) $L \leqslant [(2n)^{\log(n+2)}]^{1/2}$.

For the second set:

(6) $n^{\log k} \cdot 4^{(1-k)} \leqslant N$, and (7) $N \leqslant (2n+1)^{\lceil \log k \rceil}$.

For the third set:

(8) $(n-2)^{(2/3)(\log((1/3)\cdot\log(n-2))-1)} \leqslant N$, and (9) $N \leqslant (2n+1)^{\lceil \log n \rceil}$;

(10) $H = 0$;

(11) $(n-2)^{(2/3)(\log((1/3)\cdot\log(n-2))-(5/2))} \leqslant W$, and (12) $W \leqslant n(2n+1)^{\lceil \log(n/2) \rceil}$;

(13) $L = n - 1$.

TABLE II

Result	Section
(1)	4
(2)	10
(3)	4, 10
(4)	12
(5)	11
(6)	7
(7)	8
(8)	6
(9)	13
(10)	— (trivial)
(11)	6
(12)	12, 13
(13)	5

For convenience in finding individual items, the distribution of results over sections of the paper is given in Table II. This distribution is the result of our grouping the results according to the methods used to derive them.

2. Normality

Henceforth, if E is a regular expression, $|E|$ will be the set of strings that E represents.

We call an expression E *normal* with respect to an arc-labeled graph G if there is a pair of functions, init and fin, from subexpressions of E to nodes of G, for which:

1. If $F \cup G$ is a subexpression of E, then $\text{init}(F) = \text{init}(G) = \text{init}(F \cup G)$ and $\text{fin}(F) = \text{fin}(G) = \text{fin}(F \cup G)$.

2. If $F \cdot G$ is a subexpression of E, then $\text{init}(F \cdot G) = \text{init}(F)$, $\text{fin}(F \cdot G) = \text{fin}(G)$, and $\text{fin}(F) = \text{init}(G)$.

3. If F^* is a subexpression of E, then $\text{init}(F) = \text{init}(F^*) = \text{fin}(F) = \text{fin}(F^*)$. In this case we call this common value the *base point* of the star.

4. If F is any subexpression of E, $|F| \subset$ the set of all label sequences of paths from $\text{init}(F)$ to $\text{fin}(F)$.

Note that for each of the sets defined in the preceding section, *all* expressions for the set are normal with respect to its graph, by virtue of the distinctness of the arc labels.

3. Index

An expression E *covers* a path p on a graph G if E is normal with respect to G, and there is a string q in $|E|$ of which the label sequence of p is a contiguous substring. If there is a greatest integer n for which E covers p^n, we call it $I_p(E)$, the *index* of p in E, and say that E is p-finite; otherwise we define $I_p(E) = \infty$ and say that E is p-infinite. One checks easily that index satisfies the difference equations:

$$\begin{aligned}
&I_p\,(\text{alphabetical symbol}) = 0 \text{ or } 1,\\
&I_p\,(F \cup G) = \max(I_p(F), I_p(G)),\\
&I_p\,(F \cdot G) \leqslant I_p(F) + I_p(G) + 1,\\
&I_p\,(F^*) = \sup[I_p(F^k)\colon k \geqslant 0].
\end{aligned} \tag{3.1}$$

From these, we find that the difference equations for $2 \cdot L$ dominate those for $1 + I_p$, if I_p is finite, so:

$$I_p(E) < 2L(E) \text{ if } E \text{ is } p\text{-finite,}$$

and as a corollary,

$$I_p(E) < 2N(E) \text{ if } E \text{ is } p\text{-finite.} \tag{3.2}$$

4. Lower Bounds on N and H for the Complete Graph

Consider the complete graph on n nodes with arcs bearing distinct labels, and nodes labeled 1 through n. The desired bound is an immediate corollary of the following theorem.

Theorem 4.1. *There is a loop p passing through node 1 for which, given any expression E covering p, $N(E) \geqslant 2^{(n-1)}$.*

Proof. We proceed by induction on n; the assertion is trivial for $n = 1$, taking p to be the self-loop. Suppose we have a loop p passing through node 1 on the complete graph on $n - 1$ nodes satisfying:

$$\text{For each } E \text{ covering } p,\ N(E) \geqslant 2^{n-2}.$$

In the complete graph on n nodes, make p_k from p by cyclically permuting the nodes, replacing node 1 with node k. Thus, for each k, p_k is a copy of p passing through node k and missing node $k - 1$ (the minus being taken modulo n to keep the indices in range). Now consider the loop

$$g = p_1{}^m a_{12} p_2{}^m a_{23} \cdots p_n{}^m a_{n1} ,$$

where $m = 2^n$ and a_{ij} is the arc from node i to node j. Take any E covering g, and note that for each k, $I_{p_k}(E) \geqslant 2^n$, by definition of g. By Eq. (3.2), either $N(E) \geqslant 2^{n-1}$ (and we are done), or E is p_k-infinite for all k.

We now deal with the latter case; in what follows, "minimal," when applied to subexpressions, means "minimal with respect to the relation 'subexpression of.'" The set of all p_k-infinite subexpressions of E is a subset of the finite set of all subexpressions of E and hence has minimal elements; each such minimal element is a star. For each k, choose a minimal p_k-infinite subexpression $F_k{}^*$ and among all these, choose a minimal one and call it F^*.

F^* has a base point (by normality of E), say node j. Choose G^* from among the $F_k{}^*$ covering P_{j+1} (so that the loop that G^* covers misses node j, and the loop that F^* covers passes through node j). Let ϵ be the expression representing the set whose sole element is the null string. Now if, in E, we replace F^* with ϵ, G^* still covers P_{j+1} after the replacement, since all that has been lost is loops on a point that P_{j+1} misses anyway. Thus,

$N(G^*) \geqslant 2^{n-2}$ after the substitution (by inductive hypothesis), and

$N(F^*) \geqslant 2^{n-2}$ before the substitution (again by inductive hypothesis), from which

$N(E) \geqslant 2^{n-1}$ before the substitution. Q.E.D. (of result (1)).

That $H = n$ for the complete graph was first proved by Cohen [1]. One-half ($H \leqslant n$) of this theorem is easy (and will follow from results in Section 10), while the other half can, alternatively, be proved by the methods used above. The analog of Theorem 4.1 is:

For each integer m, there is a loop p passing through node 1 on the complete graph on n nodes for which, given any expression E covering p, either

$$N(E) \geqslant m$$

or

$$H(E) \geqslant n \qquad \text{and} \qquad I_p(E) = \infty.$$

The pertinent loop is built from the loop a_{11}^m in the case $n = 1$ by taking

$$g_m = (p_1 a_{12} p_2 a_{23} \cdots p_n a_{n1})^{2m}$$

instead of $p_1{}^m a_{12} \cdots p_n{}^m a_{n1}$.

With this analog of Theorem 4.1 in hand, we consider any expression E_{1n} for the set of all paths from node 1 to node n. (This expression is normal, by the comment at the end of Section 2.) For all integers m, E_{1n} must cover g_m; since E_{1n} is of finite length, $H(E_{1n}) \geqslant n$.

So far, we have proved inequality (1) and sketched a proof of equality (3).

5. Contractions

For expressions containing stars, normality and index have sufficed as tools for isolating the desired features of subexpressions. For the star-free expressions arising in the remaining cases to be treated, a different tool is needed, and is introduced here.

A *contraction* of a regular expression E is an expression obtained from E by applying the following transformations to E until the resulting expression is free of unions and stars.

$T1$: Replace $F \cup G$ with F or with G,

$T2$: Replace an F^* with an n-fold concatenation of F with itself, $n \geqslant 0$.

One observes immediately:

Lemma 5.1. *A string w is in $|E| \Leftrightarrow$ there is a contraction C of E for which $|C| = \{w\}$.*

The virtue of this seemingly unnatural redefinition of elementhood for regular sets is that the resulting C's can be written as trees and used in counting arguments, as will be done in the next section. We make a simple preliminary application here.

Theorem 5.1. *If E is an expression for the set of all paths from node 1 to node n in the half-complete graph on n nodes, then $L(E) \geqslant n - 1$.*

Proof. First, observe that any starred subexpression of E can be replaced with ϵ, since $| E |$ is finite; thus only $T1$ is involved in contracting E. Since $T1$ does not increase L, $L(E) \geqslant L(C)$ for any contraction C of E, and if $| C |$ = the longest path from 1 to n, then $L(C) = n - 1$. Q.E.D.

Since writing out E as a union of concatenations gives $L(E) \leqslant n - 1$, we have proved result (13).

6. Lower Bounds on N and W for the Half-Complete Graph

We need the following preliminary facts.

Lemma 6.1. *The number M of (binary) concatenations in an expression E is* $\leqslant N(E) - 1$.

Proof.

$$M \text{ (alphabetical symbol)} = 0,$$
$$M(E \cup F) = M(E) + M(F),$$
$$M(E \cdot F) = M(E) + M(F) + 1,$$
$$M(E^*) = M(E).$$

Comparing these equations with those for N, we find that $1 + M$ is dominated by N. Q.E.D.

Lemma 6.2. *If E is normal with respect to a graph G, so is any contraction of E.*

Proof. There is little to prove. We need only extend the functions init and fin to the new subexpressions created when replacing a star by an n-fold concatenation; their values are taken to be the base point of the star, and all the conditions for normality survive. Q.E.D.

Let p be any path from node 1 to node n on the half-complete graph for which the number of nodes that p passes through (excluding node 1 and node n) is k. If E is any expression for the set of all paths through the half-complete graph, then by Lemma 5.1 there is a contraction of E that describes the path p, and this contraction can be written as a binary tree whose nodes are concatenations, and whose tips are the branch labels of p. Since, by the remark at the end of Section 2, E is normal, we can associate with each of these concatenations a node of the half-complete graph, namely fin (F) where F is the left factor in the concatenation (see Section 2).

The situation now is that we have the path p of length $k + 1$ arcs represented by a binary tree having k nontip (tree) nodes and $k + 1$ tips, each of whose nontip (tree) nodes is labeled with one of the $n - 2$ (graph) node numbers. We next derive an upper estimate for the number of these trees by first counting unlabeled trees, and then multiplying by the number of labelings. The number of unlabeled trees is at most 4^k, since each child of each nontip (tree) node is either a tip or not, and this set of k quaternary choices fixes the unlabeled tree. The number of labelings of nontip (tree) nodes is, of course, $\leqslant (n - 2)^k$, but we need an estimate that involves the number N of alphabetical symbols in the expression E. To get such an estimate, consider the longest path through the tree; at its tip lies one of the N symbols of the expression E, and once this symbol is chosen, the labeling of that path through the tree is fixed (see Section 2 and Lemma 6.2). This tree path passes through at least $\log(k + 1)$ nontip (tree) nodes so there are at most $(n - 2)^{(k-\log(k+1))}$ ways of labeling the remaining nodes of the tree. Putting this all together, we find that the number of possible trees is at most

$$4^k \cdot N \cdot (n - 2)^{(k-\log(k+1))}.$$

On the other hand, the number of paths p of length $k + 1$ through the half-complete graph is the binomial coefficient $\binom{n-2}{k}$, since choosing the path's interior nodes fixes the path. We will underestimate this number as $(n - 2)^k/k^k$ to get the inequality

$$(n - 2)^k/k^k \leqslant 4^k \cdot N \cdot (n - 2)^{(k-\log(k+1))}.$$

By canceling the common factor of $(n - 2)^k$ and rearranging, we find:

$$N \geqslant (n - 2)^{\log(k+1)}/(4k)^k \geqslant (n - 2)^{\log k}/(4k)^k$$

or

$$N \geqslant \exp_2[(\log(n - 2) - k) \log k - 2k].$$

Choosing $k = \frac{1}{3} \log n$, we find, after a little algebra:

$$N \geqslant (n - 2)^{(2/3)[\log((1/3)\log(n-2))-1]} \qquad \text{(result (8))}.$$

A lower bound on W is found by observing that $L \times W \leqslant N$, and since every expression for the half-complete graph has $L = n - 1$, then

$$W \geqslant (n - 2)^{(2/3)[\log((1/3)\log(n-2))-(5/2)]} \qquad \text{(result (11))}.$$

7. A Lower Bound on N for Set of All Paths of Length $\leqslant l$ between Given Nodes in the Complete Graph on n Nodes

The method of the preceding section works in this case as well, with minor modifications. Since in this case, a path of length $l = k + 1$ may touch up to k nodes, excluding end points, and may touch a node more than once, $\binom{n-2}{k}$ is replaced with n^k on the left side of the main inequality. Furthermore, we do not get to choose k, so we are left with:

$$n^k \leqslant 4^k \cdot N \cdot n^{(k-\log(k+1))},$$

so

$$N \geqslant 4^{(1-l)} n^{\log l}$$

This gives result (6).

8. An Upper Bound on N for the Set of All Paths of Length $\leqslant k$

Consider a finite, directed graph on n nodes with at most one (labeled) branch between any pair of nodes. Let T be the matrix of one-step transitions, i.e., the i, j-element of T = the label on the branch that goes from node i to node j, if there is such a branch, and, equals 0 otherwise.

Let I be the identity matrix, and "lift" the algebra of regular expressions to matrices of regular expressions in the obvious way (union acts like addition, concatenation like multiplication, and ϵ like unity).

Let $F_k(T) = I \cup T \cup T^2 \cup \cdots \cup T^k$. As is well known, $F_k(T)$ is a matrix whose i, j-element is a regular expression for the set of all paths of length $\leqslant k$ from node i to node j.

Let $g(k)$ be the number of symbols in the largest entry of matrix T^k; using $T^k = [T^{(k+1)/2}] \cdot [T^{(k-2)/2}]$ for k odd and $T^k = T^{k/2} \cdot T^{k/2}$ for k even, we find $g(k) \leqslant n\{[g(k+1)/2] + [g(k-1)/2]\}$ for k odd and $g(k) \leqslant 2ng(k/2)$ for k even, and $g(1) = 1$.

Claim. $g(k) \leqslant (2n)^{\lceil \log k \rceil}$.

Proof goes by induction and rests on the fact that if k is odd,

$$\lceil \log(k+1) \rceil = \lceil \log k \rceil.$$

Let $f(k)$ be the number of symbols in the largest entry of matrix

$$F_k(T) = I + T + T^2 + \cdots + T^k;$$

since

$$F_k(T) = T^{(k+1)/2} F_{(k-1)/2}(T) + F_{(k-1)/2}(T) \qquad \text{for} \quad k \text{ odd},$$

and

$$F_k(T) = T^{k/2}(F_{k/2}(T) - I) + F_{k/2}(T) \qquad \text{for} \quad k \text{ even},$$

we have

$$f(k) \leqslant n\left[g\left(\frac{k-1}{2}\right) + f\left(\frac{k-1}{2}\right)\right] + f\left(\frac{k-1}{2}\right) \qquad \text{for} \quad k \text{ odd}$$

and

$$f(k) \leqslant n[g(k/2) + f(k/2)] + f(k/2) \qquad \text{for} \quad k \text{ even.}$$

CLAIM. $f(k) \leqslant (2n + 1)^{\lceil \log k \rceil}$ *(result* (7)).

Again, the proof goes by induction, using the fact that if k is odd,

$$\lceil \log(k + 1) \rceil = \lceil \log k \rceil.$$

We note in passing that the resulting expressions preserve ambiguity.

9. PATH DECOMPOSITION

At several points in what follows, we shall need the fact that each path on a graph of n nodes can be written as:

(path on $\leqslant n/2$ nodes) (loop) (path on $\leqslant n/2$ nodes),

where integer division by 2 is intended, e.g., $\frac{3}{2} = 1$. For, by choosing the shortest prefix of the path that hits $n/2 + 1$ distinct nodes, and the shortest suffix that hits $n/2 + 1$ distinct nodes, we force some node to be hit by both.

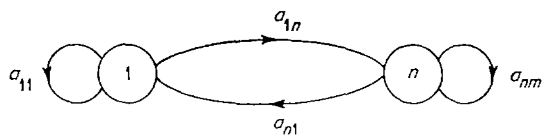
10. UPPER BOUNDS ON N AND H FOR THE COMPLETE GRAPH ON n NODES

A standard [2] method for deriving a regular expression for the set of all paths between two nodes of a graph proceeds as follows.

1. Number the nodes so that the node pair for which the paths are being represented is $(1, n)$. Let the expression labeling the arc from node i to node j be a_{ij}.

2. Choose k so that $1 < k < n$, and for each $i \neq k$, $j \neq k$, replace a_{ij} with $a_{ij} \cup a_{ik}a_{kk}^*a_{kj}$. (This explicity represents all paths passing through node k elsewhere in the graph, thus eliminating the need for node k.)

3. Continue eliminating nodes until only 1 and n are left. Then use the fact that all paths from 1 to n on



are represented by

$$(a_{11} + a_{1n}a_{nn}^*a_{n1})^* \, a_{1n}a_{nn}^* \, .$$

When this method is used on the complete graph on n nodes, symmetry guarantees that at each step of the way, all arc expressions will have the same length. Since each node elimination quadruples the size of each remaining arc expression, and since the last step sextuples the size, we have

$$N \leqslant 6 \cdot 4^{(n-2)} \qquad \text{(result (2)).}$$

That $H \leqslant n$ follows immediately from the same construction (half of result (3)).

11. An Upper Bound on L for the Complete Graph on n Nodes

Let $P_{ij}(n)$ be the set of all paths from node i to node j on the complete graph on n nodes, and let $P_{ij}^{(k)}(n)$ be the set of all paths in $P_{ij}(n)$ that do not pass through node k. We shall abbreviate $L(P_{ij}(n))$ as $L_{\mathrm{d}}(n)$ when $i \neq j$ (d stands for "different"), and abbreviate $L(P_{ij}(n))$ as $L_{\mathrm{s}}(n)$ when $i = j$ (s stands for "same"). The same abbreviations will be used for W as well as L.

Since

$$P_{ii}(n) = \left(\bigcup_{jk} a_{ij}P_{jk}^{(i)}(n)\, a_{ki}\right)^* \cup \{a_{ii}\},$$

we have

$$L_{\mathrm{s}}(n) \leqslant 2 + L_{\mathrm{d}}(n-1).$$

By representing each path as

$$\text{(path on} \leqslant n/2 \text{ nodes) (loop) (path on} \leqslant n/2 \text{ nodes),}$$

we get

$$P_{ij}(n) = \left[\bigcup_{\substack{\text{all pairs } R,\, S \text{ of} \\ \text{subsets containing} \\ \text{at most } n/2 \text{ nodes}}} T(R, S)\right],$$

where each term $T(R, S)$ is of the form

$$\bigcup_k P_{ik}(n/2)\, P_{kk}(n)\, P_{kj}(n/2),$$

from which

$$L_{\mathrm{d}}(n) \leqslant 2L_{\mathrm{d}}(n/2) + L_{\mathrm{s}}(n).$$

Substituting the former equation in the latter, we get

$$L_d(n) \leqslant 2L_d(n/2) + L_d(n-1) + 2.$$

We substitute this equation $n/2$ times into itself in the $L_d(n-1)$ position to get

$$L_d(n) \leqslant n(L_d(n/2) + 1) \leqslant 2n(L_d(n/2)),$$

so

$$L_d(n) \leqslant [(2n)^{\log n + 2}]^{1/2} \qquad \text{(result (5))}.$$

12. An Upper Bound on Width for the Complete Graph on n Nodes

In the preceding section we made L small by using a very "wide" (large union of terms) formula. Here we shall go the other way and make W small by using a very "long" (large L) formula. We first consider paths from a node to itself:

$$P_{ii}(n) = \left[\bigcup_{\substack{\text{all pairs } R,\, S \text{ of} \\ \text{subsets containing} \\ \text{at most } n/2 \text{ nodes}}} T(R, S) \right]^*,$$

where each term $T(R, S)$ is of the form

$$\bigcup_k P_{ik}(n/2)\, P_{kk}^{(i)}(n)\, P_{ki}(n/2).$$

Next, convert this large (starred) union to a large product by the identity

$$\left(\bigcup_i A_i\right)^* = (\Pi_i A_i{}^*)^*,$$

so that the width of the entire formula is at most the width of the widest term of the form $P_{ik}(n/2)\, P_{kk}^{(i)}(n)\, P_{ki}(n/2)$. Since $P_{ik}(n/2)$ and $P_{ki}(n/2)$ are of width $W_d(n/2)$, while $P_{kk}^{(i)}(n)$ is of width $W_s(n-1)$ we have

$$W_s(n) \leqslant \max(W_s(n-1),\, W_d(n/2)).$$

To get a second inequality involving W_s and W_d, use

$$P_{ij}(n) = \bigcup_k Q_{ik}(n)\, P_{kk}(n)\, Q_{kj}(n),$$

where the Q's, although possibly involving paths through any node, contain only paths of length at most $n/2$, and hence have expressions of size $\leqslant (2n+1)^{\lceil \log(n/2) \rceil}$ by result (7), so

$$W_d \leqslant n \max[(2n+1)^{\lceil \log(n/2) \rceil}, W_s(n)].$$

By repeatedly replacing $W_s(n)$ with $\max[W_s(n-1), W_d(n/2)]$ we get

$$W_d \leqslant n \max[(2n+1)^{\lceil \log(n/2) \rceil}, W_d(n/2)].$$

Now if the first argument of the max ever dominates the second, it will continue to do so for all larger n, since

$$(n/2)(2(n/2)+1)^{\lceil \log(n/4) \rceil} \leqslant (2n+1)^{\lceil \log(n/2) \rceil}.$$

Since the first argument already dominates for $n = 2$, we have

$$W_d \leqslant n(2n+1)^{\lceil \log(n/2) \rceil} \qquad \text{(result (4))}.$$

13. Upper Bounds on N and W for the Half-Complete Graph on n Nodes

Each of these bounds follows easily from facts derived earlier. By substituting $k = n$ in the formula for all paths of length $\leqslant k$ we find $N \leqslant (2n+1)^{\lceil \log n \rceil}$. We have not improved on the bound $W \leqslant n(2n+1)^{\lceil \log(n/2) \rceil}$ derived for the complete graph.

(Since the half-complete graph can be made from the complete graph by annulling edges, the same bound applies in this case. Alternatively, the same bound can be derived directly, via a matrix calculation like the one used in Section 8.)

14. Comments and Conclusions

It is unfortunate that most of these measures are worse than polynomial in n. Incidentally, if difference of sets is allowed as an operation, everything becomes polynomial; for example, $\bigcup_{ij} P_{ij}(n)$, the set of all paths on the n-node complete graph, can be written in terms of the set A of all arcs, as $A^* - \bigcup A^*abA^*$, where the union is taken over all pairs (a, b) of edges for which b does not follow a. A similar construction works for intersection of sets.

We would be pleased if some other researchers were inspired to narrow some of the gaps between upper and lower bounds.

Acknowledgments

The authors thank their more practically oriented colleagues, Professors Lee Osterweil and Lloyd Fosdick, for asking the questions partially answered here. Significant assistance in improving the exposition was provided by Professor Harold Gabow and by the referee, Professor Nancy Lynch.

References

1. R. S. Cohen, Star height of certain families of regular events, *J. Comput. System Sci.* **4** (1970), 281–297.
2. J. E. Hopcroft and J. D. Ullman, "Formal Languages and Their Relation to Automata," p. 39, Addison-Wesley, Reading, Mass., 1969.