Algorithm 457

Finding All Cliques of an Undirected Graph [H]

Coen Bron* and Joep Kerbosch† [Recd. 27 April 1971 and 23 August 1971]
* Department of Mathematics † Department of Industrial Engineering, Technological University Eindhoven, P.O. Box 513, Eindhoven, The Netherlands

Present address of C. Bron: Department of Electrical Engineering, Twente University of Technology, P.O. Box 217, Enschade, The Netherlands.

Key Words and Phrases: cliques, maximal complete subgraphs, clusters, backtracking algorithm, branch and bound technique, recursion
CR Categories: 3.71, 5.32
Language: Algol

Description

Introduction. A maximal complete subgraph (clique) is a complete subgraph that is not contained in any other complete subgraph.

A recent paper [1] describes a number of techniques to find maximal complete subgraphs of a given undirected graph. In this paper, we present two backtracking algorithms, using a branch-and-bound technique [4] to cut off branches that cannot lead to a clique.

The first version is a straightforward implementation of the basic algorithm. It is mainly presented to illustrate the method used. This version generates cliques in alphabetic (lexicographic) order.

The second version is derived from the first and generates cliques in a rather unpredictable order in an attempt to minimize the number of branches to be traversed. This version tends to produce the larger cliques first and to generate sequentially cliques having a large common intersection. The detailed algorithm for version 2 is presented here.

Description of the algorithm—Version 1. Three sets play an important role in the algorithm. (1) The set *compsub* is the set to be extended by a new point or shrunk by one point on traveling along a branch of the backtracking tree. The points that are eligible to extend *compsub*, i.e. that are connected to all points in *compsub*, are collected recursively in the remaining two sets. (2) The set *candidates* is the set of all points that will in due time serve as an extension to the present configuration of *compsub*. (3) The set *not* is the set of all points that have at an earlier stage already served as an extension of the present configuration of *compsub* and are now explicitly excluded. The reason for maintaining this set *not* will soon be made clear.

The core of the algorithm consists of a recursively defined extension operator that will be applied to the three sets just described. It has the duty to generate all extensions of the given configuration of *compsub* that it can make with the given set of candidates and that do not contain any of the points in *not*. To put it differently: all extensions of *compsub* containing any point in *not* have already been generated. The basic mechanism now consists of the following five steps:

Step 1. Selection of a candidate.
Step 2. Adding the selected candidate to *compsub*.
Step 3. Creating new sets *candidates* and *not* from the old sets by removing all points not connected to the selected candidate (to remain consistent with the definition), keeping the old sets in tact.
Step 4. Calling the extension operator to operate on the sets just formed.
Step 5. Upon return, removal of the selected candidate from *compsub* and its addition to the old set *not*.

We will now motivate the extra labor involved in maintaining the sets *not*. A necessary condition for having created a clique is that the set *candidates* be empty; otherwise *compsub* could still be extended. This condition, however, is not sufficient, because if now *not* is nonempty, we know from the definition of *not* that the present configuration of *compsub* has already been contained in another configuration and is therefore not maximal. We may now state that *compsub* is a clique as soon as both *not* and *candidates* are empty.

If at some stage *not* contains a point connected to all points in *candidates*, we can predict that further extensions (further selection of candidates) will never lead to the removal (in Step 3) of that particular point from subsequent configurations of *not* and, therefore, not to a clique. This is the branch and bound method which enables us to detect in an early stage branches of the backtracking tree that do not lead to successful endpoints.

A few more remarks about the implementation of the algorithm seem in place. The set *compsub* behaves like a stack and can be maintained and updated in the form of a global array. The sets *candidates* and *not* are handed to the extensions operator as a parameter. The operator then declares a local array, in which the new sets are built up, that will be handed to the inner call. Both sets are stored in a single one-dimensional array with the following layout:

| *not* | *candidates*

index values: 1.....*ne*...............*ce*....

The following properties obviously hold:

1. $ne \leq ce$
2. $ne = ce$:empty (*candidates*)
3. $ne = 0$:empty (*not*)
4. $ce = 0$:empty (*not*) and empty (*candidates*) = clique found

If the selected candidate is in array position $ne + 1$, then the second part of Step 5 is implemented as $ne := ne + 1$.

In version 1 we use element $ne + 1$ as selected candidate. This strategy never gives rise to internal shuffling, and thus all cliques are generated in a lexicographic ordering according to the initial ordering of the candidates (all points) in the outer call.

For an implementation of version 1 we refer to [3].

Description of the algorithm—Version 2. This version does not select the candidate in position $ne + 1$, but a well-chosen candidate from position, say s. In order to be able to complete Step 5 as simply as described above, elements s and $ne + 1$ will be interchanged as soon as selection has taken place. This interchange does not affect the set *candidates* since there is not implicit ordering. The selection does affect, however, the order in which the cliques are eventually generated.

Now what do we mean by "well chosen"? The object we have in mind is to minimize the number of repetitions of Steps 1–5 inside the extension operator. The repetitions terminate as soon as the bound condition is reached. We recall that this condition is formulated as: there exists a point in *not* connected to all points in *candidates*. We would like the existence of such a point to come about at the earliest possible stage.

Let us assume that with every point in *not* is associated a counter, counting the number of candidates that this point is not connected to (*n*umber *o*f *d*isconnections). Moving a selected candidate into *not* (this occurs after extension) decreases by one all counters of the points in *not* to which it is disconnected and introduces a new counter of its own. Note that no counter is ever

Fig. 1. Random graphs show the computing time per clique (in ms) versus dimension of the graph (in brackets: total number of cliques in the test sample).

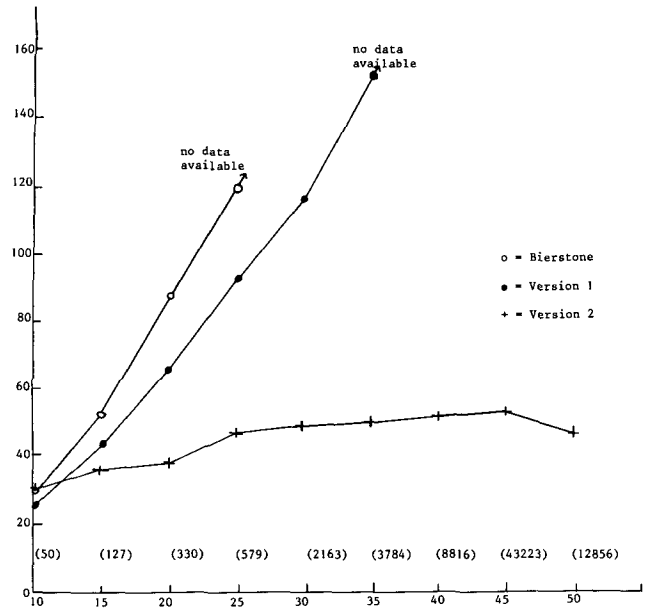
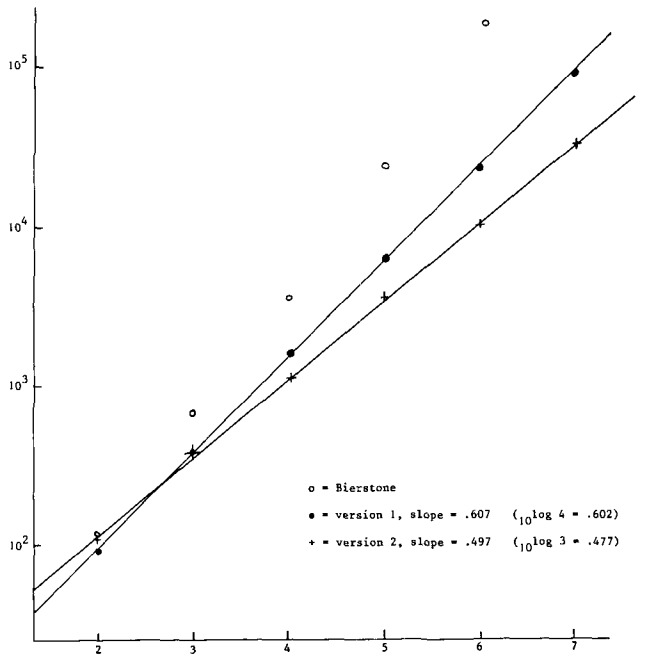


Fig. 2. Moon-Moser graphs show the computing time (in ms) versus k. Dimension of the graph = $3k$. Plotted on logarithmic scale.



decreased by more than one at any one instant. Whenever a counter goes to zero the bound condition has been reached.

Now let us fix one particular point in *not*. If we keep selecting candidates disconnected to this fixed point, the counter of the fixed point will be decreased by one at every repetition. No other counter can go down more rapidly. If, to begin with, the fixed point has the lowest counter, no other counter can reach zero sooner, as long as the counters for points newly added to *not* cannot be smaller. We see to this requirement upon entry into the extension operator, where the fixed point is taken either from *not* or from the original *candidates*, whichever point yields the lowest counter value after the first addition to *not*. From that moment on we only keep track of this one counter, decreasing it for every next selection, since we will only select disconnected points.

The Algol 60 implementation of this version is given below.

Discussion of comparative tests. Augustson and Minker [1] have evaluated a number of clique finding techniques and report an algorithm by Bierstone [2] as being the most efficient one.

In order to evaluate the performance of the new algorithms, we implemented the Bierstone algorithm[1] and ran the three algorithms on two rather different testcases under the Algol system for the EL-X8.

For our first testcase we considered random graphs ranging in dimension from 10 to 50 nodes. For each dimension we generated a collection of graphs where the percentage of edges took on the following values: 10, 30, 50, 70, 90, 95. The cpu time per clique for each dimension was averaged over such a collection. The results are graphically represented in Figure 1.

The detailed figures [3] showed the Bierstone algorithm to be of slight advantage in the case of small graphs containing a small number of relatively large cliques. The most striking feature, however, appears to be that the time/clique for version 2 is hardly dependent on the size of the graph.

[1] Bierstone's algorithm as reported in [1] contained an error. In our implementation the error was corrected. The error was independently found by Mulligan and Corneil at the University of Toronto, and reported in [6].

The difference between version 1 and "Bierstone" is not so striking and may be due to the particular Algol implementation. It should be borne in mind that the sets of nodes as they appear in the Bierstone algorithm were coded as one-word binary vectors, and that a sudden increase in processing time will take place when the input graph is too large for "one-word representation" of its subgraphs.

The second testcase was suggested by the referee and consisted of regular graphs of dimensions $3 \times k$. These graphs are constructed as the complement of k disjoint 3-cliques. Such graphs contain 3^k cliques and are proved by Moon and Moser [5] to contain the largest number of cliques per node.

In Figure 2 a logarithmic plot of computing time versus k is presented. We see that both version 1 and version 2 perform significantly better than Bierstone's algorithm. The processing time for version 1 is proportional to 4^k, and for version 2 it is proportional to $(3.14)^k$ where 3^k is the theoretical limit.

Another aspect to be taken into account when comparing algorithms is their storage requirements. The new algorithms presented in this paper will need at most $\frac{1}{2}M(M+3)$ storage locations to contain arrays of (small) integers where M is the size of largest connected component in the input graph. In practice this limit will only be approached if the input graph is an almost complete graph. The Bierstone algorithm requires a rather unpredictable amount of store, dependent on the number of cliques that will be generated. This number may be quite large, even for moderate dimensions, as the Moon-Moser graphs show.

Finally it should be pointed out that Bierstone's algorithm does not report isolated points as cliques, whereas the new algorithm does. Either algorithm can, however, be modified to produce results equivalent to the other. Suppression of 1-cliques in the new algorithm is the simplest adaption.

Acknowledgments. The authors are indebted to H.J. Schell for preparation of the test programs and collection of performance statistics. Acknowledgments are also due to the referees for their valuable suggestions.

References

1. Augustson, J.G., and Minker, J. An analysis of some graph theoretical cluster techniques, *J. ACM 17* (1970), 571–588.
2. Bierstone, E. Unpublished report. U of Toronto.
3. Bron, C., Kerbosch, J.A.G.M., and Schell, H.J. Finding cliques in an undirected graph. Tech. Rep. Technological U. of Eindhoven, The Netherlands.
4. Little, John D.C., et al. An algorithm for the traveling salesman problem. *Oper. Res. 11* (1963), 972–989.
5. Moon, J.W., and Moser, L. On cliques in graphs. *Israel J. Math.* 3 (1965), 23–28.
6. Mulligan, G.D., and Corneil, D.G. Corrections to Bierstone's algorithm for generating cliques. *J. ACM 19* (Apr. 1972), 244–247.

Algorithm

```
procedure output maximal complete subgraphs 2(connected, N);
  value N; integer N;
  Boolean array connected;
comment The input graph is expected in the form of a symmetrical
  Boolean matrix connected. N is the number of nodes in the
  graph. The values of the diagonal elements should be true;
begin
  integer array ALL, compsub[1 : N];
  integer c;
  procedure extend version 2(old, ne, ce);
    value ne, ce; integer ne, ce;
    integer array old;
  begin
    integer array new[1 : ce];
    integer nod, fixp;
    integer newne, newce, i, j, count, pos, p, s, sel, minnod;
    comment The latter set of integers is local in scope but need
      not be declared recursively;
    minnod := ce; i := nod := 0;
DETERMINE EACH COUNTER VALUE AND LOOK FOR MINIMUM:
    for i := i + 1 while i ≤ ce ∧ minnod ≠ 0 do
    begin
      p := old[i]; count := 0; j := ne;
COUNT DISCONNECTIONS:
      for j := j + 1 while j ≤ ce ∧ count < minnod do
        if ¬ connected[p, old[j]] then
        begin
          count := count + 1;
SAVE POSITION OF POTENTIAL CANDIDATE:
          pos := j
        end;
TEST NEW MINIMUM:
      if count < minnod then
      begin
        fixp := p; minnod := count;
        if i ≤ ne then s := pos
        else
        begin s := i; PREINCR: nod := 1 end
      end NEW MINIMUM;
    end i;
    comment If fixed point initially chosen from candidates then
      number of disconnections will be preincreased by one;
BACKTRACKCYCLE:
    for nod := minnod + nod step −1 until 1 do
    begin
INTERCHANGE:
      p := old[s]; old[s] := old[ne + 1];
      sel := old[ne + 1] := p;
FILL NEW SET not:
      newne := i := 0;
      for i := i + 1 while i ≤ ne do
        if connected[sel, old[i]] then
          begin newne := newne + 1; new[newne] := old[i] end;
FILL NEW SET cand:
      newce := newne; i := ne + 1;
      for i := i + 1 while i ≤ ce do
        if connected[sel, old[i]] then
        begin newce := newce + 1; new[newce] := old[i] end;
ADD TO compsub:
      c := c + 1; compsub[c] := sel;
      if newce = 0 then
      begin
        integer loc;
        outstring(1, 'clique = ');
        for loc := 1 step 1 until c do
          outinteger(1, compsub[loc])
      end output of clique
      else
      if newne < newce then extend version 2(new, newne, newce);
REMOVE FROM compsub:
      c := c − 1;
ADD TO not:
      ne := ne + 1;
      if nod > 1 then
      begin
SELECT A CANDIDATE DISCONNECTED TO THE FIXED POINT:
        s := ne;
LOOK: FOR CANDIDATE:
        s := s + 1;
        if connected[fixp, old[s]] then go to LOOK
      end selection
    end BACKTRACKCYCLE
  end extend version 2;
  for c := 1 step 1 until N do ALL[c] := c;
  c := 0; extend version 2(ALL, 0, N)
end output maximal complete subgraphs 2;
```

Remark on Algorithm 323 [G6]

Generation of Permutations in Lexicographic Order [R.J. Ord-Smith, *Comm. ACM 11* (Feb. 1968), 117]

Mohit Kumar Roy [Recd. 15 May 1972]
Computer Centre, Jadavpur University, Calcutta 32, India

In presenting Algorithm 323, *BESTLEX*, for generating permutations in lexicographic order, the author has mentioned the number of transpositions. It may be remarked here that equal numbers of transpositions are required by both *BESTLEX* and the previously fastest algorithm, Algorithm 202 [1]. The exact number of transpositions (T_n) necessary to generate the complete set of $n!$ permutations is given by

$T_n = n!\,(\psi_{n-1}) - (n+1)/2$, if n is odd, and
$T_n = n!\,(\psi_{n-2}) - n/2$, if n is even,

where $\psi_{2n} = 1 + \frac{1}{2!} + \frac{1}{4!} + \cdots + \frac{1}{(2n)!} \doteq 1.543$ for $n \geq 3$.

The above expressions do not include the few extra transpositions (equal to the integral part of $n/2$) required by *BESTLEX* to generate the initial arrangement from the final one, as this portion has

not been included in Algorithm 202. Therefore, the number of transpositions has no importance in the context of the claim that *BESTLEX* is more than twice as fast as Algorithm 202.

The main factor contributing to the speed of *BESTLEX* is the substantial reduction in the number of comparisons required, by the introduction of the **own integer array** q. Taking into account only those comparisons which involve array elements, the number of comparisons (C_n) required to generate all the $n!$ permutations can be shown to be equal to

$$C_n \text{ (Algorithm 202)} = \frac{n!}{2}[1 + 3\varphi_{n-2}] + n,$$

$$C_n \text{ (BESTLEX)} = n!\,[\tfrac{1}{2} + \varphi_{n-1}],$$

where $\varphi_n = 1 + \frac{1}{2!} + \frac{1}{3!} + \cdots + \frac{1}{n!} \doteq 1.718$ for $n \geq 6$.

This shows that the number of comparisons required by *BESTLEX* is lower by $.859(n!)$ (approximately) in the case of the generation of all the $n!$ arrangements.

Finally, a modification of the *BESTLEX* algorithm is suggested which will reduce the number of comparisons again by $(n!)/2$. The modification involves replacement of lines 2–14 of Algorithm 323 by the following.

```
begin own integer array q[3:n]; integer k, m;
  real t; own Boolean flag;
comment Own dynamic arrays are not often implemented. The
  upper bound will have to be given explicitly;
if first then
begin first := false; flag := true
for m := 3 step 1 until n do q[m] := 1
end of initialization process;
if flag then
begin flag := false;
  t := x[1]; x[1] := x[2]; x[2] := t;
  go to finish
end;
flag := true;
for k := 3 step 1 until n do
```

References

1. Shen, Mok-Kong. Algorithm 202, generation of permutations in lexicographical order. *Comm. ACM* 6 (Sept. 1963), 517.

Added in proof: An improved version of *BESTLEX*, viz. Algorithm 323A, Generation of Permutation Sequences: Part 2, by R.J. Ord-Smith [*Comp. J. 14*, 2 (May 1971), 136–139], which also incorporates the modification suggested here, has come to the author's attention.

Remark on Algorithm 408 [F4]

A Sparse Matrix Package (Part I)
[John Michael McNamee, *Comm. ACM 14* (Apr. 1971), 265–273]

E.E. Lawrence [Recd. 1 February 1972, 12 March 1973]
Central Application Laboratory, Mullard Limited, New Road, Mitcham, Surrey CR4 4XY, England

The subroutines constituting Algorithm 408 were, with the exception of *MVSPMX* and *WRSPMX*, tested on an IBM 360/65 using CALL/360-0S. The author's alteration (iii) was introduced, i.e. declaration of the M-array to be half length. Other changes were introduced in order: (a) to make the algorithm more conversational in a time shared environment; and (b) to improve the speed of the sorting procedure in *PERCOL*.

The following deficiencies in the algorithm were noted

1. The dimensional parameters of *ACSPMX*, *ADSPMX*, and *MUSPMX* are incomplete. As an illustration of this consider the two matrices

$$A = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 4 \end{bmatrix}$$

$$B = \begin{bmatrix} 0 & 0 & 0 & 4 \\ 0 & 0 & 3 & 0 \\ 0 & 2 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}$$

each of which has four nonzero elements.

Then the sum matrix has eight such elements, and in general, for two matrices with n_1 and n_2 nonzero elements, the number of nonzero elements, n_3, in the sum matrix is in the range $0 \leq n_3 \leq n_1 + n_2$.

However in *ADSPMX* the condition used is $n_1 = n_2 = n_3$.

Similar arguments apply to *ACSPMX* and *MUSPMX*.

To correct this requires extensions to the parameter lists and dimension statements, and also it changes the conditional statements within the subroutines concerned.

This shows up with the CALL/360-0S system since the compiler performs subscript checking. It would not be evident on most compilers including the IBM Fortran IV G compiler. It is, however, bad practice to rely on default effects of compilers.

2. There are three, probably copying, errors in *MUSPMX* (page 270).

(i) Line 33 should be:
```
IF(NCA.EQ.NCB) GO TO 3
```

(ii) Line 55 should be:
```
DO 14 J = 1, NRB
```

(iii) Line 102 should be:
```
CALL IPK(NRB,MC,2,NM)
```

Remark on Algorithm 420 [J6]

Hidden-Line Plotting Program
[Hugh Williamson, *Comm. ACM 15* (Feb. 1972), 100–103]

Hugh Williamson [Recd. 9 Oct. 1972]
National Con-Serv, Incorporated, Austin, Texas

The input quantities to subroutine *HIDE* referred to in the following paragraphs (e.g. *N1*, *NFNS*, "input curve to be plotted") are described in the block of comment statements at the beginning of *HIDE* as originally published.

If $N1 < 0$, *DO* loop 71 is not executed properly, since the upper limit, *N1*, is less than the lower limit, 2. This affects only checking for monotonicity in the input abscissa array; otherwise, if the inputs are correct, the performance of the program is not affected.

The error is corrected if the first 11 executable statements are replaced by the following (the first executable statement of the original program, which is not changed, is listed for clarity):

```
IF(MAXDIM.LE.0) RETURN
IFPLOT = 1
IF(N1.GT.0) GO TO 76
N1 = -N1
IFPLOT = 0
```

Fig. 1. Without verticals.

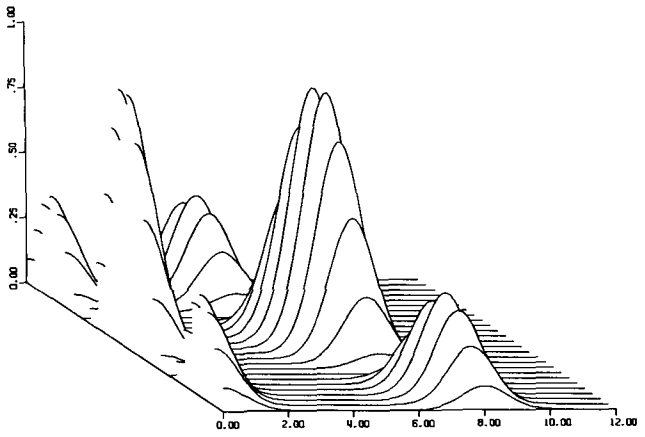
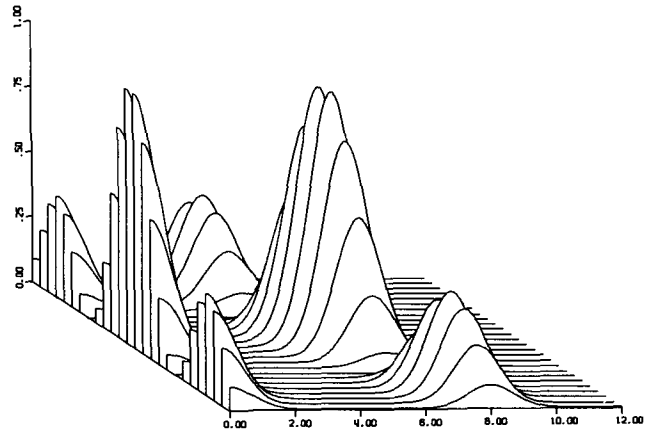


Fig. 2. With verticals to aid visualization.




```
76  DO 71 I = 2,N1
      IF(X(I-1).LT.X(I)) GO TO 71
      MAXDIM = 0
      GO TO 75
71  CONTINUE
    IF(NG.GT.0) GO TO 5000
```

On computers in which all variables are not automatically set to zero before execution, *FNSM1* is not properly initialized if *NFNS* ≤ 0. To correct this, simply insert the statement

```
FNSM1 = 0.
```

before the statement

```
IF(NFNS.LE.0) GO TO 46
```

The latter is the sixth statement after Fortran statement number 74.

FNSM1 will still be improperly defined if *NFNS* = 1. If only one curve is to be plotted, however, translating to simulate stepping in the depth dimension will not be done, so set *NFNS* = −1 for only one curve to be plotted.

In some cases, the three-dimensional surface is easier to visualize if (nearly) vertical lines are drawn at the left edge of each curve; this effect is illustrated by Figures 1 and 2. The verticals are added by inserting (*XMIN*-ϵ, *YMIN*) as the first point in each input curve to be plotted, where ϵ is a small positive number ($10^{-4} \times DELTAX$ would be appropriate).

The author appreciates very much the comments received form readers of Communications regarding implementation of *HIDE* on different computers.

Remark on Algorithm 429 [C2]

Localization of the Roots of a Polynomial [C2]

[W. Squire, *Comm. ACM 15* (Aug. 1972), 776–777]

H.B. Driessen and E.W. LeM. Hunt [Recd. 13 Oct. 1972, 29 Jan. 1973]

Supreme Headquarters Allied Powers of Europe, Technical Center, P.O. Box 174, The Hague, The Netherlands

There seems to be an error in this algorithm. If we take the polynomial:

$$z^4 + a_2z^2 + a_3z^3 + a_4z + a_5 = 0,$$

then after the second pass through the K-loop of the logical function *HRWTZR*(C, N), the term $(a_2a_3 - a_4)a_4 - a_5a_2$ is tested for a minus sign. However, the term which should be tested according to the Routh-Hurwitz criterion is $(a_2a_3 - a_4)a_4 - a_5a_2^2$. If this term is negative then there are no roots with positive real parts.

As an example, if the polynomial

$$z^4 + 5.6562\,z^3 + 5.8854\,z^2 + 7.3646\,z + 6.1354 = 0$$

is studied with the help of Algorithm 429 one will find as output:

Roots are in an annulus of inner radius .454 $E + 00$ and outer radius .836 $E + 01$;

There are no real positive roots;

The negative roots (if any) are between $-.454\ E + 00$ and $-.836\ E + 01$;

There are no roots with positive real parts.

However, if one calculates the roots of this equation, one will find approximately:

$$z_1 = -1.0001$$
$$z_2 = -4.7741$$
$$z_{3,4} = +0.0089 \pm 1.1457\,i$$

Statement 20 + 1 in the logical function *HRWTZR*(C,N), which was originally "$C1 = C(1)$", should be amended to read "$C1 = C(1)/C1$".

As a by-product of our investigation, it turns out that the structure of the logical function *HRWTZR* can be simplified by abandoning the logically redundant steps $C(K) = C(K+1)$.

The following listing incorporates both the correction and the simplifications. The function has been parameter tested on a CDC-6400.

```
      LOGICAL FUNCTION HRWTZR (C,N)
      DIMENSION C(N)
      HRWTZR = .FALSE.
      IF (C(1) .LE.0..OR.C(N).LE.0.) RETURN
      C1 = C(1)
      M = N - 1
      DO 30 I = 2,M
      DO 20 K = I,M,2
   20 C(K) = C(K) - C(K+1)/C1
      C1 = C(I)/C1
      IF (C1.LE.0.) RETURN
   30 CONTINUE
      HRWTZR = .TRUE.
      RETURN
      END
```