Discrete Applied Mathematics 44 (1993) 205–220
North-Holland

Subtree and substar intersection numbers

Yi-Wu Chang*

University of Illinois, Urbana, IL 61801, USA

Michael S. Jacobson**

University of Louisville, Louisville, KY 40292, USA

Clyde L. Monma

Bell Communications Research, Morristown, NJ 17960-1910, USA

Douglas B. West***

University of Illinois, Urbana, IL 61801, USA

Received 15 August 1990
Revised 22 July 1991

Abstract

Chang, Y.-W., M.S. Jacobson, C.L. Monma and D.B. West, Subtree and substar intersection numbers, Discrete Applied Mathematics 44 (1993) 205–220.

We introduce the *star number* [*tree number*] of a graph, being the minimum t such that G is the intersection graph of unions of t substars [subtrees] of a host tree. We study bounds on these parameters, compare them with interval number, and characterize the graphs with star number 1.

Keywords. Intersection graphs, interval number.

1. Introduction

In an *intersection representation* of a graph, each vertex is assigned a set such that vertices are adjacent if and only if their sets intersect. The sets may, for example, be intervals on the real line, in which case the resulting graph is an *interval graph*.

Correspondence to: Professor D.B. West, University of Illinois, Urbana, IL 61801, USA.
* Research supported by ONR Grant N00014-85K0570 and by University of Illinois Research Board.
** Research supported by ONR Grant N00014-85-0694.
*** Research supported by ONR Grant N00014-85K0570 and by NSA/MSP Grant MDA904-90-H-4011.

More generally, the sets may be unions of t intervals (called t-intervals) or may be boxes in d dimensions (called d-boxes). The minimum t such that a graph G has a t-interval intersection representation is called its *interval number* $i(G)$, and the minimum d such that it has a d-box intersection representation is called its *boxicity*. Early results on these parameters are summarized in [14].

We can also view discrete intervals as subpaths of a host path. We can generalize this by letting the sets used to represent the vertices be subtrees of a host tree. It is well known that the graphs obtainable as intersection graphs of subtrees of a tree are precisely the *chordal graphs*, defined to be those having no chordless cycle. Intermediate between these are the *path graphs*, studied in [3, 5, 6], which are the intersection graphs of paths in a host tree.

In the same way that complexity parameters based on intervals were introduced above, we can also introduce such parameters based on trees. The *chordality* of a graph G, mentioned originally in [1], is the minimum number of chordal graphs whose intersection is G, just as boxicity can be interpreted as the minimum number of interval graphs whose intersection is G. By analogy with interval number, here we introduce the *tree number* of G, denoted $t(G)$, which is the minimum t such that G has an intersection representation in which each vertex is assigned a set consisting of the union of (at most) t subtrees of a tree. Note that the chordal graphs are the graphs with tree number 1.

Since interval representations are subtree representations, we always have $t(G) \le i(G)$. The interval number of a chordal graph can be arbitrarily high, but this requires large cliques. In general, we prove that $i(G) \le (\omega(G)-1)t(G)+1$, where $\omega(G)$ is the size of the largest complete subgraph of G. The bound is tight for $\omega(G)=2$, and for larger $\omega(G)$ it can be achieved within a factor of $\log(\omega(G))$.

We also introduce the *star number* $s(G)$, defined to be the minimum t such that G is the intersection graph of unions of at most t stars in a host tree. The graphs with star number 1 are called the *substar graphs*. We show that $s(G)$ and $i(G)$ are independent parameters; there are substar graphs with arbitrarily high interval number and interval graphs with arbitrarily high star number. We also observe that $s(G) \le n/3$, although the best bound on the star number of an n-vertex graph remains an open question. Several other elementary bounds on star number are also presented, such as the fact that the star number of a planar graph is at most 3.[1]

The most difficult result of the paper is a forbidden subgraph characterization of substar graphs. Consider first the special class obtained by requiring the host tree itself to be a star. Since every substar contains the center or consists of a single leaf, the intersection graph G of distinct substars of a star is a *split graph*, meaning that it has a clique and an independent set that together cover the vertices. The clique

[1] Note added in proof: Y.-W. Chang has now proved that $s(G) \le \lceil (n+1)/4 \rceil$, which is best possible and that 3 is the best possible upper bound on the star number of a planar graph.

arises from the substars containing the center in the representation. Conversely, any split graph has such a representation, in which the vertices of the independent set are assigned leaves of the host star and the vertices of the clique are assigned the center and the leaves corresponding to their neighbors. If we enlarge the class of split graphs by allowing *vertex duplication* (adding a vertex adjacent to x with the same closed neighborhood as x), then we have the class of intersection graphs of substars of a star (because we are allowed to assign the same leaf more than once). When we allow a general host tree, we obtain a more interesting class of graphs. The substar graphs are chordal graphs avoiding a specified finite set of induced subgraphs, all with at most 13 vertices. The proof of the characterization can be converted into a recognition algorithm for substar graphs. Using the fact that chordal graphs can be recognized and represented in time $O(|V|+|E|)$, the running time of the recognition algorithm is quadratic in the number of vertices.

In discussing intersection representations more formally, we think of a representation as a function f that assigns each vertex of G a set. We use $n(G)$, $e(G)$, $\omega(G)$, $\Delta(G)$, $\delta(G)$ for the number of vertices, number of edges, clique number, and maximum and minimum vertex degrees of G. We will say "interval representation" in place of "multiple-interval intersection representation", and similarly for "substar representation" and "subtree representation". An *optimal* representation of a graph is one achieving the value of the parameter $i(G)$, $s(G)$, or $t(G)$ under discussion.

2. Graphs with large tree number

Given a new graph parameter, the first tasks are to show that it is well defined and that it can be arbitrarily large. Since $t(G) \le i(G)$, $t(G)$ is well defined. To show that it can grow, we strengthen the standard lower bound argument for $i(G)$ to apply also to $t(G)$.

Theorem 2.1. *If G is a triangle-free graph, then $t(G) \ge (e(G)+1)/n(G)$. More generally,*

$$t(G) \ge \frac{e(G)+\binom{\omega(G)}{2}}{n(G)(\omega(G)-1)}.$$

Proof. Let f be a subtree representation. Choose an arbitrary root u_0 of the host tree. For any subtree T used in f, let $u(T)$ denote the root, i.e., the vertex of T closest to u_0. Let $T_1, \dots, T_m$ be the subtrees used in f, indexed according to nondecreasing distance between u_0 and $u_i = u(T_i)$. Note that $u_j \in T_i$ if $i<j$ and $T_i \cap T_j \ne \emptyset$. Let $D_k = \{j<k\colon T_j \cap T_k \ne \emptyset\}$. For each $j \in D_k$, T_j contains u_k; hence $|D_k| \le \omega(G)-1$. This implies that the introduction of each successive subtree T_k in $\{T_1, \dots, T_m\}$ creates at most $\min\{k-1, \omega(G)-1\}$ new edges (intersections) in the representation. Altogether we have $t(G)n(G)$ subtrees in an optimal representation, where $n(G) >$

$\omega(G)$. If we add $\binom{\omega(G)}{2}=\sum_{k=1}^{\omega(G)}(\omega(G)-k)$ for $T_1,\ldots,T_{\omega(G)}$ to bring the contribution of each tree up to at most $\omega(G)-1$, then we obtain $t(G)n(G)(\omega(G)-1)\geq e(G)+\binom{\omega(G)}{2}$. □

Corollary 2.2. $t(K_{m,n})=\lceil(mn+1)/(m+n)\rceil$.

Proof. Trotter and Harary [13] provided an interval representation with this many intervals per vertex. □

Given the inequality $t(G)\leq i(G)$, which as noted above is tight for some graphs, it is natural to consider how bad the inequality can be. Although the interval number can be much larger than the tree number, the ratio is bounded by the clique number. We construct an interval representation from a subtree representation by using the subtree indexing argument described in the proof of Theorem 2.1.

Theorem 2.3. $i(G)\leq(\omega(G)-1)t(G)+1$, *and this bound is best possible when* $\omega=2$.

Proof. Let f be an optimal subtree representation of G. Given the indexing scheme $T_1,\ldots,T_m$ and sets $D_k=\{j<k\colon T_j\cap T_k\neq\emptyset\}$ as described in the proof of Theorem 2.1, recall that $|D_k|\leq\omega(G)-1$. To construct an interval representation from this, begin with one interval $I(u)$ for each vertex u, all disjoint. Then, for each k and each $j\in D_k$, with $T_k\subset f(x)$ and $T_j\subset f(y)$, create a tiny interval for x in $I(y)$.

Now consider triangle-free graphs. Observe that adding a pendant edge to a graph (new vertex x adjacent to old vertex y) cannot increase its tree number, because we can add a new vertex to the host tree, let it be $f(x)$ and be in $f(y)$, and extend the host tree from a vertex of $f(y)$ to include the new vertex. On the other hand, adding pendant vertices to a graph can increase the interval number. In particular, the complete bipartite graph $G=K_{t,2t^2+1}$ has interval number t (in general, $i(K_{m,n})=\lceil(mn+1)/(m+n)\rceil$ [13]), but adding a pendant edge to each vertex of the large part increases the interval number to $t+1$ (see [11], for example). Since the lower bound on $t(G)$ equals the upper bound on $i(G)$ when G is a complete bipartite graph, the parameters differ by one on this augmentation. □

For larger clique number, the bound of Theorem 2.3 is not tight. Scheinerman [10] proved that the maximum interval number of a chordal graph with clique size ω is asymptotic to $c\omega/\log\omega$. Hence the bound is not tight for large ω, even if $t(G)=1$. Perhaps the argument for Theorem 2.3 can be strengthened to save a factor of $\log\omega$.

3. Substar representations

Like interval representations, substar representations have less power than subtree representations, and we have $s(G) \geq t(G)$. The star number obeys some of the same bounds as the interval number, but it can be larger or smaller than the interval number. The simplest example of a graph with $s(G) < i(G)$ is the tree obtained by subdividing each edge of $K_{1,3}$ once; call this graph Y. A tree has interval number 2 if and only if it contains Y [8, 13]. On the other hand, every tree G has star number 1; subdivide every edge of G to obtain a host tree T in which the star assigned to v consists of all edges in T incident to v. We postpone the discussion of interval graphs that are not substar graphs.

Most of the results of this section are upper bounds on $s(G)$. Several of these use the following simple remark.

Lemma 3.1. *If θ is a multiple intersection parameter using sets that are subtrees of a host tree and $G = \bigcup G_i$, then $\theta(G) \leq \sum \theta(G_i)$.*

Proof. Disjoint host trees can be combined into a single host tree by the addition of edges. □

Corollary 3.2. *$s(G) \leq \max_{H \subseteq G} e(H)/(n(H) - 1)$; in particular, $s(G) \leq 3$ if G is planar.*

Proof. The quantity $\max_{H \subseteq G} e(H)/(n(H) - 1)$ is equal to the *arboricity* $\Upsilon(G)$, meaning the minimum number of forests whose union is G, as proved by Nash-Williams [9] and later by Edmonds [2]. The edge bound for planar graphs gives the arboricity at most 3. Any bound on arboricity is a bound on star number because every forest has star number 1. □

The main result of [11] is that $i(G) \leq 3$ for planar graphs; the same bound holds trivially for $s(G)$. We will later exhibit planar graphs with star number 2.

In discussing intersection representations, it often helps to consider the *depth* of a representation f, which is the maximum over $x \in \bigcup_{v \in V(G)} f(v)$ of the number of vertices v such that $x \in f(v)$. The *depth-r* star number [interval number] is the minimum t such that G has a depth-r substar representation [interval representation]. We denote this by $s_r(G)$ [$i_r(G)$]. The star number obeys several of the bounds on interval number because they are bounds on $i_2(G)$ and $s(G) \leq s_2(G) \leq i_2(G)$, as we prove next. The construction given for trees showed that $s_2(G) = 1$ for any forest G, which implies that the inequality can be strict and that the corollary about planar graphs in fact holds for $s_2(G)$, not just $s(G)$.

Theorem 3.3. *For any graph G, $s_2(G) \leq i_2(G)$.*

Proof. Given an interval representation f of depth 2, we construct a substar representation g of depth 2 for the same graph G by converting each interval in f

into a substar in g in a host tree that is a long path together with many pendant edges attached at every vertex with even index.

Let f' be obtained from f by discarding the intervals that are entirely contained in another interval. For the intervals $I_1, \ldots, I_m$ of f', the order of left endpoints is the same as the order of right endpoints. Let $x_1, \ldots, x_{2m}$ be the long path in the host tree for g. Corresponding to I_j, we establish a star S in g with center at x_{2j}, assigned to the vertex $v \in V(G)$ such that $I_j \subset f(v)$. The star S also contains all the pendant edges at x_{2j}. Furthermore, S contains x_{2j-1} if I_j intersects I_{j-1}, and it contains x_{2j+1} if I_j intersects I_{j+1}. Each remaining interval in f is contained in some interval in f'. If some such interval in $f(v)$ is contained in I_j, then in g we assign v a singleton star at a pendant neighbor of x_{2j} not assigned to any other such vertex. The result is a depth-2 substar representation establishing the same edges as f. □

This inequality does not extend to higher depth. In particular, when we present interval graphs with large star number, we will find a graph with $i(G) = i_3(G) = 1 < s(G) = 2$. Meanwhile, Theorem 3.3 implies

Corollary 3.4. *The star number of a graph obeys the following bounds:*
(1) $s(G) \leq s_2(G) \leq \lceil (1 + \Delta(G))/2 \rceil$, *with equality for triangle-free regular graphs.*
(2) $s(G) \leq s_2(G) \leq 1 + \max_{H \subseteq G} \delta(H)$.

Proof. The bounds follow from Theorem 3.3 and bounds on $i_2(G)$. The maximum degree bound appears in [8], with a short proof in [15]. The bound $i_2(G) \leq 1 + \max_{H \subseteq G} \delta(H)$ appears in [12]. □

The next upper bound on $s(G)$ does not depend on limited depth.

Theorem 3.5. *For any graph G, $s(G)$ is bounded by the minimum number of cliques whose (vertex) deletion leaves an independent set.*

Proof. Let $Q_1, \ldots, Q_m$ be cliques such that $G - Q$ has no edges, where $Q = \bigcup Q_i$. Suppose $G - Q_i$ has l_i vertices. Establish a host tree containing m disjoint stars S_i, where S_i has l_i leaves. Assign the leaves of S_i to the distinct vertices of $G - Q_i$. For each $v \in V(G - Q_i)$, assign its leaf of S_i also to each neighbor of v in Q_i. The vertices of Q_i are also assigned the center of S_i. The result is a substar representation in which each vertex is assigned one substar of each S_i, so $s(G) \leq m$. □

For a complete r-partite graph with parts of size m, Theorem 3.5 establishes m as an upper bound, and Theorem 2.1 yields $\lceil (m+1)/2 \rceil$ as a lower bound. The lower bound is achieved if $m = 2$ or $r = 2$, but for larger m and r we do not know the value of $s(K_{m,\ldots,m})$.

We next consider bounds in terms of the number of vertices alone. The construction in the proof of Theorem 3.5 implies that $s(G)$ is bounded by the minimum size of a maximal matching, which is at most $n/2$. As a lower bound, Theorems 2.1 and 3.3 imply that $s(G)=i(G)=\lceil (n+1)/4 \rceil$ when $G=K_{\lfloor n/2 \rfloor, \lceil n/2 \rceil}$. Griggs [7] proved that this is the largest interval number for n-vertex graphs. The following weaker bound is easy to prove, given that we can invoke Griggs' result.

Theorem 3.6. *For an n-vertex graph,* $s(G) \le \lceil n/3 \rceil$.

Proof. If G is triangle-free, then Griggs' result and Theorem 3.3 imply $s(G) \le s_2(G) \le i_2(G) = i(G) \le \lceil (n+1)/4 \rceil \le \lceil n/3 \rceil$. For graphs with triangles, we use induction on n; $s(K_3)=1$. Suppose $n(G)>3$, and let $T=uvw$ be a triangle in G. We express G as the union of two graphs and invoke Lemma 3.1. Let H_1 be the subgraph consisting of all edges incident to T, and let $H_2=G-T$. By Theorem 3.5, $s(H_1)=1$, and by induction, $s(H_2) \le \lceil n/3 \rceil - 1$, so Lemma 3.1 implies $s(G) \le \lceil n/3 \rceil$. □

Despite these similarities, there is no direct relationship between interval number and star number; either can be larger. As remarked in the introduction, split graphs are intersection graphs of substars of a star, so split graphs have star number 1. However, there are split graphs with arbitrarily high interval number. This was first observed by Trotter (unpublished), as cited in [10]. Trotter's construction can be simplified as follows: define a split graph G_n in which the independent set corresponds to the elements of $[n]=\{1,\dots,n\}$, the clique vertices correspond to the subsets of $[n]$, and the edges between them correspond to the membership relation. In a t-interval representation f of G_n, the nt intervals for singletons have some specified order. There are at most $\binom{nt}{2t}$ subsets of these intervals that can be the subset intersected by a t-interval. Hence $2^n > \binom{nt}{2t}$ implies $i(G_n)>t$, and $i(G_n)$ grows at least as fast as $n/(2 \lg n)$. (This counting argument is essentially the same as that given by Edgar Ramos (private communication) for a related problem.) It is also easy to observe that the number of labeled n-vertex split graphs with clique number $(n/2)(\binom{n}{n/2}2^{(n/2)^2})$ grows faster than the number of labeled t-interval representations $(\binom{2nt}{2t,\dots,2t})$ for any fixed t. Our next task is to construct interval graphs with arbitrarily high star number.

Theorem 3.7. *There exist interval graphs with arbitrarily large star number.*

Proof. Let P_k be a poset whose cover diagram is a tree with $k+1$ levels and a single maximal element. The tree is full, in the sense that every maximal chain in P_k has length k. Every element of P_k covers k elements, except that the elements of the penultimate level cover $k(k-1)$ elements, and of course the minimal elements cover none. Let G_k be the comparability graph of P_k. Using tiny intervals for the minimal elements and an interval for each nonminimal element that contains the intervals for all its descendants, we obtain an interval representation for P_k.

We claim that $s(G_k)=k$. The upper bound is easy. Corollary 3.2 yields $s(G_k) \le$

$1+k$. We can save one star explicitly by using a host tree with a large star for each nonminimal element of the tree, all disjoint. For each vertex of G_k, place a singleton star at a leaf of the large star assigned to each of its ancestors in P_k, obtaining a substar representation with depth 2. The minimal elements of P_k and their parents are assigned k stars.

For the lower bound, suppose that f is a substar representation of G_k with at most $k-1$ stars per vertex. Let u_1 be the unique maximal element of P_k. Since the k subtrees obtained by deleting u_1 have no relations between them, they induce subgraphs of G_k with no edges between them. Since $f(u_1)$ consists of only $k-1$ stars, we can select a subtree in which no element is assigned a star containing a center of any star in $f(u_1)$; let u_2 be the root (maximal element) of this subtree. By the same reasoning, we define $u_3, \ldots, u_k$ to be vertices such that u_i is a child of u_{i-1} in the tree and no element of the subtree rooted at u_i is assigned the center of any star in $f(u_{i-1}), \ldots, f(u_1)$. Let U be the collection of $k(k-1)$ stars assigned to $u_1, \ldots, u_k$.

Let $S=\{v_1, \ldots, v_{k(k-1)}\}$ be the collection of minimal elements of P_k under u_k. By the choice of $u_1, \ldots, u_k$, each of $f(v_1), \ldots, f(v_{k(k-1)})$ contains a leaf from each of $f(u_1), \ldots, f(u_k)$. Since each v_i has k such requirements but only $k-1$ stars assigned to it, some star assigned to v_i must intersect two stars in U, via a path P_i with at most three vertices. For $v_i \neq v_j$, the paths P_i and P_j are disjoint, since S is an independent set in G_k. Hence these paths can be viewed as corresponding to $k(k-1)$ disjoint edges of a multigraph on $k(k-1)$ vertices corresponding to the stars listed in U. With this many edges, there must be a cycle, which translates back into a cycle in the host tree. □

In fact, the graph G_k is critical for star number k, in that any proper induced subgraph of it has star number at most $k-1$.

Theorem 3.8. *The graph G_k constructed for Theorem 3.7 is a minimal forbidden induced subgraph for graphs with star number less than k.*

Proof. We construct a substar representation for $G_k - v$ with at most $k-1$ stars per vertex. Let $u_1, \ldots, u_r, v$ be the maximal chain in P_k from the root u_1 down to v. Let $W=\{u_1, \ldots, u_r\}$. Let V be the $k(k-1)-1$ remaining leaves under u_r if $r=k$; otherwise $V=\emptyset$. Let $Z=W \cup V \cup \{v\}$. For each nonminimal element of $P-Z$, establish a single star with many leaves. For $i \leq \min\{r, k-1\}$, assign u_i the $k-1$ stars that have been assigned to its immediate children, together with many additional pendant leaves. If $i=r=k$, nothing has been assigned to the children of u_i; in this case, simply establish $k-1$ stars for u_r with two vertices each. For each $x \in P-Z$, let $w(x)$ be the nearest ancestor of x in W, and assign x a leaf in a nontrivial star for each ancestor of x other than $w(x)$. If x is not a child of $w(x)$, then this assigns x a leaf in a star contained in $f(w(x))$; otherwise, $f(w(x))$ already contains the nontrivial star for x. We have assigned at most $k-1$ stars to each vertex of $G_k - Z$ and accounted for all the adjacencies involving these vertices.

If $r<k$, simply take one of the extra leaves from each $f(u_i)$ and identify all of these into a single vertex; this takes care of the adjacencies within W and completes the representation of $G-v$. If $r=k$, we currently have $k-1$ components of the representation for each $u_i \in W$. Each such component contains a substar assigned to a child of u_i, some leaves assigned to descendants of u_i, and some leaves assigned only to u_i. Let C be an Eulerian circuit for the doubly-directed complete directed graph on k vertices $a_1, \dots, a_k$. Begin a traversal of C at a_1, considering one of the stars assigned to u_1 as the current star. Traverse C, except for the last edge. For each edge $a_i a_j$ encountered, identify an unused leaf of the current star for u_i with an unused leaf in a star for u_j that has not yet been visited, and consider this new star the current star. Since C visits each vertex exactly $k-1$ times before returning to a_1 the last time, each new star is available as requested. The result is a long caterpillar in which $k(k-1)-1$ vertices are assigned to two vertices of W. Assign one of these "double-duty" vertices to each $y \in V$. Complete $f(y)$ by adding a leaf from a star for its $k-2$ other neighbors in W. This assigns $k-1$ stars to vertices of V and completes the representation. □

4. Characterization of substar graphs

Theorems 3.7 and 3.8 show that G_2 is a minimal forbidden induced subgraph for substar graphs; G_2 can also be written as $K_1 \vee 2P_3$, where P_n henceforth denotes the n-vertex path. In this section, we provide a forbidden induced subgraph characterization of the substar graphs. First we describe the induced subgraphs other than the chordless cycles that are forbidden. Label G_2 by letting z be the vertex of degree 6 and a,x,u and v,y,b be the two paths. We use $\leftrightarrow$ and $\not\leftrightarrow$ for adjacency and nonadjacency. Note that $K_1 \vee P_5$ is the graph obtained by identifying u,v into a single vertex w and deleting the extra copy of the edge zw. We will show that the graphs in Fig. 1 are not substar graphs. We have drawn and labeled them in a way suggested by the proof that any chordal graph that is not a substar graph must contain one of these. Although there seem to be many forbidden induced subgraphs, we will see that they arise from a relatively small number of configurations. Indeed, it would be more compact to represent them by their subtree intersection representations, which are suggested in Fig. 2.

(1) $H_1 = K_1 \vee P_5$.

(2) $H_2 = G_2 = K_1 \vee 2P_3$. Furthermore, we also include the graphs obtained from G_2 by adding edges from u to a possibly empty initial segment of v,y,b and from v to a possibly empty terminal segment of a,x,u. Although our argument will show uniformly that all these graphs are not substar graphs, several of them contain $K_1 \vee P_5$ as induced subgraphs and therefore must be deleted to obtain the minimal set of forbidden induced chordal subgraphs. In particular, if we add only the edge uv to obtain G, then $G-b=G-a=K_1 \vee P_5$. If we add uv and uy but not ub, then $G-y=K_1 \vee P_5$. Hence at least one of u, v must have all three possible neighbors to

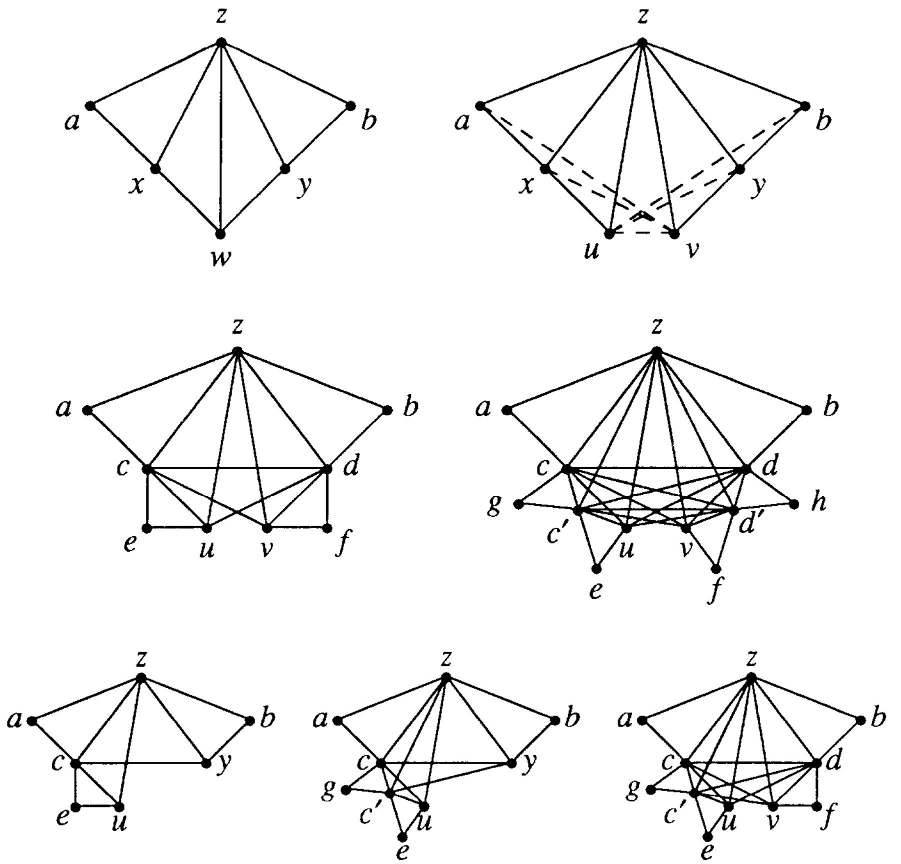

Fig. 1. Forbidden induced subgraphs for substar graphs.

avoid $K_1 \vee P_5$, which means there are three *minimal* forbidden induced subgraphs of Type 2.

(3) H_3. The 9-vertex graph obtained from G_2 by changing the name of x to c and y to d, adding two new vertices e, f, and adding the trail d, u, e, c, d, f, v, c.

(4) H_4. The 13-vertex graph obtained from H_3 by deleting the edges ce and df, adding the vertices g, c', d', h with neighborhoods $\{c', c\}$, $\{g, e, z, d', d, v, u, c\}$, $\{h, f, z, c', c'u, v, d\}$, and $\{d', d\}$, respectively.

(5) H_5. The 7-vertex graph obtained from the subgraph of H_3 induced by $\{a, c, u, e, z\}$ and the subgraph of H_2 induced by $\{z, b, y\}$ by identifying the two vertices labeled z and adding the edge yc.

(6) H_6. The 9-vertex graph obtained from the subgraph of H_4 induced by $\{a, c, c', u, e, z\}$ and the subgraph of H_2 induced by $\{z, b, y\}$ by identifying the two vertices labeled z and adding the edges yc and yc'.

(7) H_7. The 11-vertex graph obtained from the subgraph of H_4 induced by $\{a, c, c', u, e, g, z\}$ and the subgraph of H_3 induced by $\{z, b, y, d, v, f\}$ by identifying

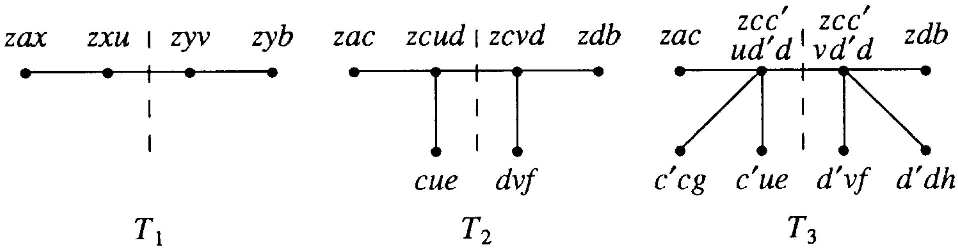

Fig. 2. Forced subtree representations for minimal nonsubstar graphs.

the two vertices labeled z and adding all edges between $\{c, c', u\}$ and $\{d, v\}$ except uv.

We separate one direction of the characterization as a lemma. Let $\boldsymbol{S}$ be the set of graphs listed above. First we make the well-known observation, useful for both directions of the characterization, that any subtree representation of a chordal graph G can be shrunk to a "minimal" representation in which there is a bijection between the maximal cliques of G and the vertices of the host tree T. For completeness, we include the argument. The vertices of each maximal clique are assigned pairwise intersecting subtrees, which by the Helly property have a common vertex. The vertices for distinct maximal cliques must of course be distinct. If any $q \in V(T)$ corresponds to a nonmaximal clique Q, then the vertices of Q are also assigned the neighbor q'' of q on the q, q'-path in T, where q' is a vertex corresponding to a maximal clique Q' containing Q. We can then shrink the edge qq'' of T, placing the combined vertex in all subtrees that were assigned q''. The same argument applies to maximal cliques assigned more than one vertex. Note that this shrinkage does not change the intersection graph or increase the diameter of any assigned subtree. (Since the existence of a perfect elimination order guarantees that an n-vertex chordal graph has at most n maximal cliques, this observation also guarantees that the host tree does not need more than n vertices.)

Lemma 4.1. *The graphs in* $\boldsymbol{S}$ *are not substar graphs.*

Proof. For each $G \in \boldsymbol{S}$, we assume a minimal substar representation f and obtain a contradiction. By the remark, we have a bijection between maximal cliques of G and vertices of the host tree T.

Consider the graphs of Types 1 or 2 in $\boldsymbol{S}$. We can argue uniformly by referring to the middle vertex of the path $H_1 - z$ as u or v. For each such graph, we have $(a \nleftrightarrow u, y)$, $(x \nleftrightarrow y)$, and $(x, v \nleftrightarrow b)$. As a result, zax, zxu, zvy, and zyb are contained in distinct maximal cliques. If α, β, γ, δ, respectively, are the vertices of the host tree representing these cliques, then we have $\alpha, \beta \in f(x)$ and $\gamma, \delta \in f(y)$. Since $x \nleftrightarrow y$ in each graph, this puts two independent edges into $f(z)$, which is impossible when $f(z)$ is a star.

For the other five graphs, we begin by listing the maximal cliques.

	Q_1	Q_2	Q_3	Q_4	Q_5	Q_6	Q_7	Q_8
H_3:	zac	$zcud$	$zcvd$	zdb	cue	dvf		
H_4:	zac	$zc'cudd'$	$zc'cvdd'$	zdb	$c'cg$	$c'ue$	$d'vf$	$d'dh$
H_5:	zac	zcu	zcy	zyb	cue			
H_6:	zac	$zc'cu$	$zc'cy$	zyb	$c'cg$	$c'ue$		
H_7:	zac	$zc'cud$	$zc'cvd$	zdb	$c'cg$	$c'ue$	dvf	

Let q_i be the vertex of T corresponding to Q_i. In each case, $f(z)=\{q_1,q_2,q_3,q_4\}$. The center of $f(z)$ cannot be q_1 or q_4, because a and b are nonadjacent to some other vertex appearing in two of the other cliques involving z. We show first that we may reduce the possible substar representations for each graph to the assumption that q_3 is the center of $f(z)$.

For H_5 and H_6, the center of $f(z)$ cannot be q_2, because $f(y)\cap f(z)$ is connected. For H_7, if the center of $f(z)$ is q_2, then q_4 and q_3 are both adjacent to q_2, but then $f(dvf)$ extends q_4,q_2,q_3 to a longer path in $f(d)$. For H_3 and H_4, we may use symmetry to complete the reduction.

For H_3 and H_5, q_5 now extends q_1,q_3,q_2 to a longer path in $f(c)$. For H_4,H_6,H_7, the assumption that $f(c)$ and $f(c')$ are stars now implies that $\{f(zac),q_3,q_2,f(c'ue)\}$ induce P_4 in that order, with q_3 as the center of $f(c)$ and q_2 the center of $f(c')$. Now we cannot place $f(c'cg)$ at a vertex adjacent to the center of both $f(c)$ and $f(c')$. □

Theorem 4.2. *A graph is a substar graph if and only if it has no chordless cycle and no induced subgraph in* $\boldsymbol{S}$.

Proof. We have proved necessity. For the converse, let G be a chordal graph such that G is not a substar graph, but all proper induced subgraphs of G are substar graphs; we prove $G\in\boldsymbol{S}$. Since G is chordal, G can be represented as an intersection graph of subtrees of a host tree T; we choose an appropriate representation f.

First, we may assume that every $f(v)$ is a subtree of T having diameter 2 or 3 (i.e., not containing P_5). To see this, let u be a vertex of G belonging to exactly one maximal clique, Q; u can be chosen as the first vertex in a perfect elimination order for G. Let f' be a substar representation for $G'=G-u$. In the host tree T' for f' there is a vertex q' corresponding to a clique Q' of G' that contains $Q-u$. Add a leaf to T' and include it in the subtrees assigned to vertices of Q. Since no subtree in f' contains P_4, the result is a subtree representation of G with no subtree containing P_5.

Among the subtree representations of G in which no subtree contains P_5, choose one assigning the minimum number of nonstars. Take this representation and reduce it to a minimal representation as discussed before the lemma. In the resulting representation f there is a bijection between maximal cliques of G and vertices of T. This does not disturb our previous assumptions on the representation, because shrinking an edge does not increase the diameter of any subtree. This minimality implies that for any edge qq' of T with corresponding cliques Q, Q' in G, both $Q-Q'$ and $Q'-Q$ are nonempty. The minimality of G ensures that also $Q\cap Q'$ is nonempty.

Since G is not a substar graph, we may choose $z\in V(G)$ such that $f(z)$ contains P_4. Let Q_1, Q_2, Q_3, Q_4 be the distinct maximal cliques of G corresponding to these four vertices q_1, q_2, q_3, q_4 in order. We may choose $a\in Q_1-Q_2$ and $b\in Q_4-Q_3$, and we may choose $u\in Q_2-Q_1$ and $v\in Q_3-Q_4$. Since each vertex is assigned a tree, we have $u\nleftrightarrow a\nleftrightarrow b\nleftrightarrow v$. Suppose there exists $x\in(Q_1\cap Q_2)-Q_3$ and $y\in(Q_4\cap Q_3)-Q_2$. If $u=v$, then $\{z,a,x,u,y,b\}$ induce $H_1=K_1\vee P_5$. Otherwise, $\{z,a,x,u,v,y,b\}$ induce a forbidden subgraph of Type 2 ($f(u)$ may extend to contain q_3 and/or q_4, and $f(v)$ may extend to contain q_2 and/or q_1). This case is illustrated as T_1 in Fig. 2. In Fig. 2 vertices of the host tree are labeled by vertices of G to which they are assigned.

Note that every P_4 in $f(z)$ contains q_2 and q_3, though there may be alternate choices for q_1 and q_4. Suppose that $(Q_1\cap Q_2)\subset Q_3$ for every choice of q_1 among the neighbors of q_2 in $f(z)$, so that x cannot be chosen as described above for any P_4 in $f(z)$. Fixing q_1 and a as above, call a neighbor of q_2 in T *switchable* if it is assigned to at least one vertex of $Q_1\cap Q_2$ and to no vertex of Q_2-Q_3. Consider the host tree T' obtained by replacing qq_2 by qq_3 for every switchable vertex $q\in V(T)$. In terms of vertices, let $f'(w)=f(w)$ for all $w\in V(G)$; this guarantees that f' is an intersection representation of G. Furthermore, since switchable vertices are assigned to nothing in Q_2-Q_3, every $f'(w)$ is a subtree of T'.

Suppose q is a neighbor of q_2 in $f(z)$ other than q_1 or q_3; q cannot be assigned to a vertex of Q_2-Q_3, because then we could use q as q_1 and obtain the left half of T_1. Furthermore, $z\in Q_1\cap Q_2$, so every neighbor of q_2 in $f(z)$ is switchable. Hence $f'(z)$ is a star centered at q_3. We claim that either the switch to T' does not increase the diameter of any subtree, contradicting the choice of f as having the minimum number of nonstars, or we obtain the left half of T_2 or T_3 in Fig. 2.

The diameter of a subtree $f(w)$ can increase under the switch only if $f(w)$ contains a switchable vertex q and a nonswitchable vertex q' as neighbors of q_2 (with associated cliques Q, Q' in G). The switch to f' increases the distance between q and q' in $f(w)$. Choose $e\in Q'-Q_2$. The nonswitchability of q' requires a vertex $u\in(Q'\cap Q_2)-(Q_1\cup Q_3)$; $u\notin Q_1$ follows from the emptiness of $(Q_1\cap Q_2)-Q_3$. (Note that u has the properties specified for the earlier "u".) If $q_1\in f(w)$, then set $c=w$; we now have the configuration on the left side of T_2 in Fig. 2.

If $q_1\notin f(w)$, then the definition of switchability guarantees a vertex $w'\in Q_1\cap Q$. If $w'=z$, forget the choices of q_1 and a, set $c=w$ (recall $w\in Q\cap Q'$), and choose $a\in Q-Q_2$, letting q play the role of q_1. We now again have the configuration on the left side of T_2 in Fig. 2. Finally, if $w'\neq z$, we set $c=w'$, set $c'=w$, and choose $g\in Q-Q_2$; we now have the configuration on the left side of T_3 in Fig. 2.

The symmetric argument implies that either we can select $y\in(Q_4\cap Q_3)-Q_2$, as on the right side of T_1 in Fig. 2, or we can select d, v, f, b on the right side of T_2 or d, d', v, f, h, b on the right side of T_3. As we have remarked, we obtain a graph in S if both x and y can be chosen. For each of the five remaining ways to pick a left side and a right side from T_1, T_2, T_3, with the left side (by symmetry) being chosen with at least as many vertices as the right side, we obtain one of the five forbidden subgraphs $H_3, \ldots, H_7$ of S. □

We close this section with informal comments on turning the characterization proof into a recognition algorithm; being very formal would essentially require repeating the proof of the theorem. Several algorithms are known for recognizing triangulated graphs in time $O(|V|+|E|)$ and producing a perfect elimination order (see [4, Chapter 4]). We begin with a perfect elimination order $v_1, \ldots, v_n$, and we try to build a successive substar representation for the subgraphs G_i induced by $\{v_n, \ldots, v_i\}$. When we are ready to add v_i, we begin with a substar representation of G_{i+1} and obtain one for G_i or find one of the forbidden configurations described in Fig. 2.

Assume we have a substar representation of the G_i; this includes a listing of the vertices in the clique associated with each vertex of the host tree in the representation. Since the neighbors of v_i induce a clique, there is one vertex of the host tree at which they appear. We extend those subtrees to a new vertex of the host tree assigned also to v_i. We now have a substar representation, unless one or more of the extended subtrees now contains P_4. Let z be a vertex such that $f(z)$ now contains P_4, corresponding to the four cliques Q_1, Q_2, Q_3, Q_4 of G. We can choose $v_i = a \in Q_1$ and u, v, b from Q_2, Q_3, Q_4 as described. If vertices described as x, y exist (check the vertex sets of the specified cliques), then G is not a substar graph. If x does not exist, we determine the set of switchable neighbors of q_2, again by examining the vertices in Q_1, Q_2, Q_3. If the switch does not decrease the diameter of any subtree, then we have reduced the number of P_4's in the representation, and we can repeat the analysis with any that remain.

Otherwise, we obtain the configuration on the left half of T_2 or T_3 in Fig. 2, as discussed in the proof. If y exists, we now have a forbidden subgraph. If y does not exist, we consider the switchable neighbors of q_3. This either brings us closer to a substar representation or produces the other half of a forbidden configuration.

The full examination of vertices for the addition of v_i (including the possibility of repeated reductions in the number of P_4's until all are eliminated) runs in linear time. Hence the algorithm runs in time $O(n^2)$. With care in implementation, this can probably be reduced to $O(|V|+|E|)$.

5. Powers of caterpillars, etc.

One graph in S that contains three of the dashed lines in Fig. 1 is in fact P_7^3, where the kth *power* of a graph G is the graph G' on the same vertices such that $(u,v) \in E(G')$ if and only if $d_G(u,v) \leq k$, where d_G is the distance function. It is reasonable to think that the star number can be made arbitrarily large by taking large powers of long paths. Surprisingly, this is not true. In fact, we prove the following stronger result. A *caterpillar* is a tree containing a path that intersects every edge.

Theorem 5.1. *If G is a caterpillar, then $s(G^k)\leq 2$ for all k.*

Proof. Let $P=x_1,\dots,x_n$ denote the path intersecting every edge. A caterpillar is a tree, so $s(G)=1$. For $k\geq 2$, the vertices not in P that are adjacent to the same vertex in P become duplicates, so we may assume for each i that there is at most one vertex y_i not in P that is adjacent to x_i. It suffices to prove the result when there is exactly one y_i for each i.

Let $S_j=\bigcup_{i=jk+1}^{(j+1)k}\{x_i,y_i\}$. The vertices $S_j-\{y_{jk+1}\}$ induce a clique in G^k. For each $j\geq 0$, we create a star in which every vertex of $S_j-\{y_{jk+1}\}$ is assigned the center. Let the leaves of the star be $q_0,\dots,q_k$. Assign q_0 to all of $S_j-\{y_{(j+1)k}\}$. For $1\leq i\leq k$, assign q_i to $\{x_{jk+i},\dots,x_{(j+1)k}\}\cup\{y_{jk+i+1},\dots,y_{(j+1)k}\}\cup\{x_{(j+1)k+i},y_{(y+1)k+i-1}\}$. For $i=k$, the sequence of y's described is empty. Note that the vertices of $f^{-1}(q_i)$ induce a clique, and that each vertex of S_j is assigned a nontrivial substar of the star associated with S_j, a leaf in the star associated with S_{j-1}, and no other vertex. The adjacencies of a vertex of S_j within S_j or in S_{j+1} are established in the star associated with S_j, its adjacencies in S_{j-1} are established in the star associated with S_{j-1}, and it has no other neighbors. □

The powers of trees are all chordal graphs. Let us remark once again the problem of the maximum value of the star number for a chordal graph on n vertices. More generally, let $\sigma(t,n)$ be the maximum value of $s(G)$ for an n-vertex graph with tree number t. The graphs G_k show that $\sigma(1,n)\in\Omega(\log n/\log\log n)$. As an upper bound, we can show that the growth is sublinear. This is true for arbitrary t, but we present the argument only for $t=1$. Note that for arbitrary n-vertex graphs, the growth is linear, since $s(K_{\lfloor n/2\rfloor,\lceil n/2\rceil})=\lceil (n+1)/4\rceil$.

Theorem 5.2. *If $t(G)=1$ and ε is any positive constant, then $s(G)\leq\varepsilon n+1/\varepsilon$.*

Proof. Let $r=\lceil 1/\varepsilon\rceil$. If G contains no $(r+1)$-clique, then $s(G)\leq r-1$, by induction on n. If $n\leq r-1$, then we can establish one nontrivial star for each vertex and also assign each vertex a leaf in the star for each of its neighbors. For $n\geq r$, any chordal graph G contains a "simplicial" vertex x, meaning its neighbors induce a clique. The graph $G-x$ also has no $(r+1)$-clique, so we obtain a representation f with at most r stars per vertex. Now introduce a new leaf to a star for each neighbor of x in G and let these leaves be $f(x)$. Since G has no $(r+1)$-clique, x has at most $r-1$ neighbors. Note that $r-1\leq 1/\varepsilon\leq n+1/\varepsilon$.

If G does contain an $(r+1)$-clique Q, then the subgraph of edges incident to Q can be represented with one star per vertex. By Lemma 3.1, induction, and the fact that $r>\varepsilon^{-1}$, we have $s(G)\leq s(G-Q)+1\leq\varepsilon(n-r-1)+\varepsilon^{-1}+1\leq\varepsilon n+\varepsilon^{-1}$. □

Acknowledgement

We thank one referee for correcting a slight error in the proof of Theorem 3.5, and the other for pointing out the relationship between the number of n-vertex split graphs and n-vertex graphs with bounded interval number.

References

[1] M.B. Cozzens and F.S. Roberts, On dimensional properties of graphs, Graphs Combin. 5 (1989) 29–46.
[2] J. Edmonds, Minimum partition of a matroid into independent sets, J. Res. Nat. Bur. Standards (B) 69 (1965) 67–72.
[3] F. Gavril, A recognition algorithm for the intersection graphs of paths in trees, Discrete Math. 23 (1978) 211–227.
[4] M.C. Golumbic, Algorithmic Graph Theory and Perfect Graphs (Academic Press, New York, 1980).
[5] M.C. Golumbic and R.E. Jamison, Algorithmic aspects of paths in a tree, Discrete Math. 55 (1985) 151–159.
[6] M.C. Golumbic and R.E. Jamison, Edge and vertex intersection of paths in a tree, J. Combin. Theory Ser. B 38 (1985) 8–22.
[7] J.R. Griggs, Extremal values of the interval number of a graph II, Discrete Math. 28 (1979) 37–47.
[8] J.R. Griggs and D.B. West, Extremal values of the interval number of a graph I, SIAM J. Algebraic Discrete Methods 1 (1980) 1–7.
[9] C.St.J. Nash-Williams, Decomposition of finite graphs into forests, J. London Math. Soc. 39 (1964) 12.
[10] E.R. Scheinerman, On the interval number of a chordal graph, J. Graph Theory 12 (1988) 311–316.
[11] E.R. Scheinerman and D.B. West, The interval number of a planar graph: Three intervals suffice, J. Combin. Theory Ser. B 35 (1983) 224–239.
[12] J. Spinrad, G. Vijayan and D.B. West, An improved edge bound on the interval number of a graph, J. Graph Theory 11 (1987) 447–449.
[13] W.T. Trotter and F. Harary, On double and multiple interval graphs, J. Graph Theory 2 (1978) 137–142.
[14] D.B. West, Parameters in partial orders and graphs: Packing, covering, and representation, in: I. Rival, ed., Graphs and Order, Proceedings NATO ASI, Banff, 1984 (Reidel, Dordrecht, 1985) 267–350.
[15] D.B. West, A short proof of the degree bound for interval number, Discrete Math. 73 (1989) 309–310.