JOURNAL OF COMPUTER AND SYSTEM SCIENCES 20, 219–230 (1980)

The Node-Deletion Problem for Hereditary Properties Is NP-Complete

JOHN M. LEWIS*

Yale University, New Haven, Connecticut 06520

AND

MIHALIS YANNAKAKIS†

Princeton University, Princeton, New Jersey 08540

We consider the family of graph problems called node-deletion problems, defined as follows: For a fixed graph property Π, what is the minimum number of nodes which must be deleted from a given graph so that the resulting subgraph satisfies Π? We show that if Π is nontrivial and hereditary on induced subgraphs, then the node-deletion problem for Π is NP-complete for both undirected and directed graphs.

1. INTRODUCTION

This paper deals with the class of graph problems defined as follows: For a fixed graph property Π, find the minimum number of nodes (or vertices) which must be deleted from a given graph so that the result satisfies Π. We call this the *node-deletion problem* for Π. Our results show that if Π is a "nontrivial" property which is "hereditary" on induced subgraph, then the node-deletion problem for Π is NP-hard. Furthermore, if we add the condition that testing for Π can be performed in polynomial time, then our results imply that the node-deletion problem for Π is NP-complete. (See [4] for an exposition of NP-completeness.)

A graph property is *nontrivial* if it is true for infinitely many graphs and false for infinitely many graphs. It is easy to see that if this condition does not hold, then the node-deletion problem is always polynomial. For example, if Π is true for only finitely many graphs $G_1, \ldots, G_n$, then we can find in polynomial time the largest G_i which is contained in a given graph. Hence the condition that the property be nontrivial is essential to our results.

A property is *hereditary* on induced subgraphs if, for any graph satisfying the property,

* Portions of the research of this author have been supported by the National Science Foundation under Grant MCS76-11460.

† Portions of the research of this author have been supported by the National Science Foundation under Grant MCS76-15255.

0022-0000/80/020219-12$02.00/0

all vertex-induced subgraphs also satisfy the property. Examples of well-known hereditary properties are planar, outerplanar, bipartite, acyclic, degree-constrained, interval graph, comparability graph, chordal, complete, independent set, and line invertible. (See [2] or [6] for definitions of these terms.) Our results thus imply the NP-completeness of a large number of specific node-deletion problems. Some of these had previously been shown to be NP-complete [3, 5, 8, 9]. The significance of the present paper lies in the fact that we exhibit techniques which can be used to prove the NP-completeness of an infinite family of problems. Note that the hereditary condition is also necessary for our results to hold. For example, the property biconnected is not hereditary, and the node-deletion problem for it is linear [14].

If a graph G does not satisfy some hereditary property, then any supergraph of G also fails to satisfy the property. We call G a *forbidden subgraph* for the property. Any hereditary property is definable by its set of forbidden subgraphs, and, conversely, any set of graphs is the set of forbidden subgraphs of some hereditary property. If the graphs all have edges, the property will be nontrivial. Our proofs rely heavily on this correspondence between hereditary properties and sets of graphs.

Since the number of sets of graphs is uncountable, so is the number of hereditary properties. But the set of algorithms can be enumerated [7] and is therefore countable. It follows that not all node-deletion problem are in NP (or even recursive). For this reason we add the stipulation that Π be a property which can be tested for in polynomial time. Then the node-deletion problem for Π is in NP, and our results show that it is NP-complete.

The paper is organized as follows. In Section 2 we show NP-hardness for node-deletion problems for properties whose forbidden subgraphs are all biconnected. In Section 3 we extend this to include node-deletion problems for all nontrivial hereditary properties on undirected graphs. Section 4 proves the same result for digraphs.

We introduce the following notation. For any graph G and graph property Π, we let $\gamma_\pi(G)$ be the minimum number of nodes which must be deleted from G in order that the resulting subgraph satisfy Π. When there is no possibility of any ambiguity, we will drop the subscript Π.

2. Properties Whose Forbidden Subgraphs Are Biconnected

A number of common graph properties can be defined by biconnected forbidden subgraphs. Examples are planar, bipartite, and acyclic. This section treats node-deletion problems for such properties.

For any $N \geqslant 3$, let SATN denote the satisfiability problem for sets of clauses in conjunctive normal form in which each clause has exactly N literals. It is known that SAT3 is NP-complete [5], and it is easy to show that SATN is NP-complete for $N \geqslant 4$ [10]. For any property Π definable by biconnected forbidden subgraphs, we will show that the node-deletion problem for Π is NP-complete by showing that SATN can be reduced to it for some N.

Let G be a biconnected graph with $N \geqslant 3$ vertices and e any edge of G. Given clauses

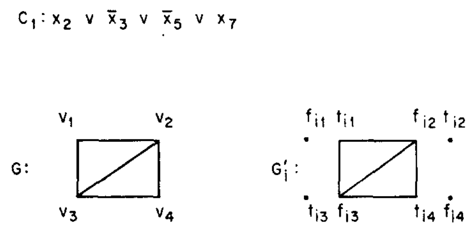


FIG. 1. The labeling of the vertices G'_i. t_{i1}, f_{i1} are associated with x_2, $\bar{x}_2$; t_{i2}, f_{i2} are associated with x_3, $\bar{x}_3$, etc.

$C_1, \ldots, C_n$, each with exactly N literals from among the set $x_1, \ldots, x_l, \bar{x}_1, \ldots, \bar{x}_l$, as input for SAT$N$, construct the graph Γ, called the *Γ-graph for* G, e, $C_1, \ldots, C_n$, as follows.

Begin with G, and for each vertex v_j, add a new, isolated vertex v'_j. This graph G' has $2N$ vertices. Let $G'_1, \ldots, G'_n$ be n disjoint copies of G', and form $G'_1 + \cdots + G'_n$, the graph consisting of the disjoint union of the n copies. For each G'_i, $i = 1, \ldots, n$, label the $2N$ vertices as follows: If the jth literal of clause C_i is x_k, let v_j be t_{ij} and v'_j be f_{ij}; if the jth literal is $\bar{x}_k$, let v_j be f_{ij} and v'_j be t_{ij}. Say that t_{ij} is *associated with* x_k, f_{ij} is *associated with* $\bar{x}_k$. A vertex labeled t will be called a t-vertex (t for true); one labeled f will be called an f-vertex (f for false). An example of this labeling for a particular clause and graph is shown in Fig. 1.

Note that if v is any of $t_{i1}, f_{i2}, f_{i3}, t_{i4}$, in Fig. 1, then $G'_i - \{v\}$ does not contain G as a subgraph, whereas if v is any of the isolated vertices of G'_i, then $G'_i - \{v\}$ still contains G. A truth assignment for $C_1, \ldots, C_n$ will correspond in the Γ-graph to removing from $G'_1 + \cdots + G'_n$ for each pair t_{ij}, f_{ij}, the vertex which is associated with a literal assigned the value 1. So if x_3 is assigned 1 in the above example, t_{i2} would be removed from G'_i, and if $\bar{x}_5$ is assigned 1, then f_{i3} would be removed. The idea is that if the truth assignment satisfies $\{C_1, \ldots, C_n\}$, these deletions will result in a graph that contains no subgraph isomorphic to G. That is, if $\{C_1, \ldots, C_n\}$ is satisfiable, there will be a set of $2nN - nN = nN$ vertices in $G'_1 + \cdots + G'_n$ whose induced subgraph does not contain G as a subgraph. Unfortunately, the converse of this statement is not true, for $G'_1 + \cdots + G'_n$ will always have an nN-vertex subgraph not containing G; for example, the subgraph induced by the nN isolated vertices. The remaining part of the construction of Γ is designed to ensure that if Γ has a subgraph of a certain size not containing G, then this subgraph will imply a valid truth assignment satisfying $\{C_1, \ldots, C_n\}$.

The $2nN$ vertices of $G'_1 + \cdots + G'_n$ can be partitioned into l sets $V_1, \ldots, V_l$, where V_k consists of vertices associated with x_k or $\bar{x}_k$. Let the number of vertices in V_k be $2p_k$, so that either x_k or $\bar{x}_k$ occurs p_k times in $C_1, \ldots, C_n$. Then V_k contains p_k^2 distinct pairs of the form t_{ij}, f_{rs}. To each such pair, attach a copy of G by identifying the two vertices with the endpoints of e. This adds to the graph $M = (N-2)\sum_{1 \leqslant k \leqslant l} p_k^2$ vertices which will be denoted $u_1, \ldots, u_M$ and referred to as u-vertices. The resulting graph Γ is the Γ-graph for G, e, $C_1, \ldots, C_n$. A partial construction of a Γ-graph is depicted in Fig. 2.

It is clear that Γ can be constructed in time polynomial in n and N. The following lemmas give the properties of Γ that allow SATN to be reduced to a node-deletion

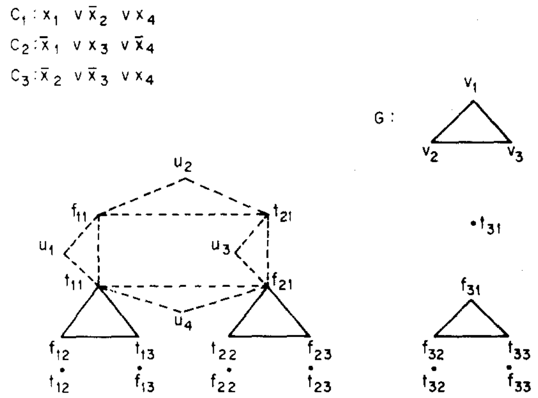


FIG. 2. Partial construction of the Γ-graph for G, e, C_1, C_2, C_3. The graph $G_1' + G_2' + G_3'$ has been formed, and the dotted edges are the attached copies of G for the set V_1 of vertices associated with x_1, $\bar{x}_1$.

problem on Γ. For a set S of vertices in a graph G, let $G \mid S$ be the subgraph induced by S.

LEMMA 1. *If $\{C_1, \ldots, C_n\}$ is satisfiable, then there is a subset $V' \subseteq V(\Gamma)$ of size $|V'| = nN + M$ such that $\Gamma \mid V'$ has no biconnected component with more than $N - 1$ vertices.*

Proof. Let $\phi: \{x_1, \ldots, x\} \to \{0, 1\}$ be a truth assignment which satisfies $C_1, \ldots, C_n$. For each set of vertices V_k, let

$$\begin{aligned} W_k &= \{t_{ij} \in V_k\}, \quad &\text{if } \phi(x_k) = 1 \\ &= \{f_{ij} \in V_k\}, \quad &\text{if } \phi(x_k) = 0. \end{aligned}$$

That is, W_k is the subset of V_k consisting of vertices associated with literals assigned the value 1. Let $V' = V(\Gamma) - \bigcup W_k$. Let $\Gamma' = \Gamma \mid V'$. In other words, delete all vertices associated with literals assigned the value 1.

For a pair t_{ij}, f_{rs} belonging to the same V_k, only one is in V'. Call this vertex v. If u is any u-vertex in the copy of G attached to t_{ij}, f_{rs}, then any path from u to a vertex not in this attached copy must contain v. Therefore v is a cutpoint of Γ'; its removal disconnects the $N - 2$ vertices in the attached copy containing u from the rest of Γ. It follows that neither u, nor any of $u_1, \ldots, u_M$, is in a biconnected component of Γ' with more than $N - 1$ vertices. Among the remaining vertices r_{ij}, f_{ij}, all edges which resulted from the attached copies of G are not present in Γ', since one of the end points of each such edge has been deleted. The only remaining edges to consider are those from the original graph $G_1' + \cdots + G_n'$. Since ϕ satisfies $C_1, \ldots, C_n$, at least one vertex from each of the n copies of G in this graph has been deleted, as argued previously, so none of the t- or f-vertices of Γ' is in a biconnected component with more than $N - 1$ vertices. Thus *no* vertex of Γ' is in a biconnected component with more than $N - 1$ vertices, so the lemma is proved. ∎

LEMMA 2. *Suppose Γ has a set $V' \subseteq V(\Gamma)$ of size $|V'| = nN + M$ such that G is not a subgraph of $\Gamma \mid V'$. Then $\{C_1, \ldots, C_n\}$ is satisfiable.*

Proof. Let $\Gamma' = \Gamma \mid V'$. Since G is not a subgraph of Γ', at least one vertex from each of the original n copies of G in $G_1 + \cdots + G_n$ must have been deleted; that is, a vertex corresponding to at least one literal in each clause has been deleted. The claim is that for each $k = 1, \ldots, l$, either V' contains only t-vertices of V_k, or it contains only f-vertices. If this claim is true, then the deleted vertices will determine a truth assignment ϕ defined by

$$\begin{aligned} \phi(x_k) &= 1, \quad \text{if } V' \cap V_k \text{ contains only } f\text{-vertices} \\ &= 0, \quad \text{if } V' \cap V_k \text{ contains only } t\text{-vertices.} \end{aligned}$$

and this truth assignment will satisfy $C_1, \ldots, C_n$.

To see that the above claim holds, note first that it can be assumed without loss of generality that all of $u_1, \ldots, u_M$ are in V'. For suppose u_k is not in V'. u_k is in a copy of G attached to some pair t_{ij}, f_{rs}. If one of these is in V', remove it from V' and add u_k. If neither of these is in V', add u_k and remove any t- or f-vertex from V'. Call the new set V'', and consider $\Gamma'' = \Gamma \mid V''$. If Γ'' contains G as a subgraph, then u_k would have to be in this subgraph. But by construction at least one of the vertices t_{ij}, f_{rs} connecting u_k to the rest of the graph has been deleted, so as argued in the proof of Lemma 1, u_k is in a biconnected component of at most $N - 1$ vertices. Thus G is not a subgraph of Γ''. Continuing this process verifies the assumption.

Now suppose for some k, $V' \cap V_k$ contains both a t-vertex t_{ij} and an f-vertex f_{rs}. Then t_{ij} and f_{rs}, plus the u-vertices in the copy of G attached to them could constitute a subgraph isomorphic to G, a contradiction. Therefore, for each k, $V' \cap V_k$ contains only t-vertices or only f-vertices, proving the claim and the lemma. ∎

These facts about the relationship between SATN and the Γ-graph lead directly to the following result.

THEOREM 3. *Let $G_1, G_2, \ldots$ be a collection of biconnected graphs such that $|V(G_i)| \geqslant |V(G_1)| = N \geqslant 3$ for all i. Let Π be the property having $G_1, \ldots$ as forbidden subgraphs. Then the node-deletion problem for Π is* NP-*complete.*

Proof. Let $e \in E(G_1)$ be arbitrary. Given N-clauses, $C_1, \ldots, C_n$ over the literals $x_1, \ldots, x_l, \bar{x}_1, \ldots, \bar{x}_l$, construct the Γ-graph for $G_1, e, C_1, \ldots, C_n$. As before, let p_k be the number of occurrences of either x_k or $\bar{x}_k$ in $C_1, \ldots, C_n$, let $M = (N - 2) \sum_{1 \leqslant k \leqslant l} p_k^2$. We show that $\{C_1, \ldots, C_n\}$ is satisfiable if and only if $\gamma_\pi(\Gamma) \leqslant nN$.

Suppose $\{C_1, \ldots, C_n\}$ is satisfiable. By Lemma 1, there is a subset $V' \subseteq V(\Gamma)$ of size $|V'| = nM + M$, obtainable by deleting from Γ those t- and f-vertices corresponding to literals assigned the value 1, such that $\Gamma \mid V'$ has no biconnected component with more than $N - 1$ vertices. This implies that none of $G_1, G_2, \ldots$ can be a subgraph of $\Gamma \mid V'$. Therefore $\gamma_\pi(\Gamma) \leqslant nN$.

Conversely, suppose $\gamma_\pi(\Gamma) \leqslant nN$. Then there is a set V' of size $|V'| = nM + M$ such that none of $G_1, G_2, \ldots$ is a subgraph of $\Gamma \mid V'$. In particular, G_1 is not a subgraph of V'. Then by Lemma 2, $\{C_1, \ldots, C_n\}$ is satisfiable. ∎

3. The General Node-Deletion Problem for Undirected Graphs

By $\alpha_0(G)$ we denote the node-cover number of G, i.e., $\alpha_0(G) = \gamma_\pi(G)$, with $\pi =$ "independent set of nodes."

Theorem 4. *The node-deletion problem for nontrivial graph-properties that are hereditary on induced subgraphs is* NP-*complete.*

Proof. For all m, n there is a number $r(m, n)$ (the so-called Ramsey number), such that every graph with no fewer than $r(m, n)$ nodes contains either a clique of m nodes (denoted K_m) or an independent set of n nodes (denoted $\bar{K}_n$) [13]. Let Π be any graph-property satisfying the conditions of the theorem. We claim that either all cliques or all independent sets of nodes (or both) satisfy Π. Suppose, to the contrary, that there are m, n such that K_m and $\bar{K}_n$ do not satisfy Π. Since Π is a nontrivial property, there is a graph satisfying Π, with more than $r(m, n)$ nodes, and since Π is hereditary on induced subgraphs either K_m or $\bar{K}_n$ has no satisfy Π. Define a complementary property $\bar{\Pi}$ as follows: A graph G satisfies $\bar{\Pi}$ iff its complement $\bar{G}$ satisfies Π. Clearly $\bar{\Pi}$ satisfies also the assumptions of the theorem (since the complement of a subgraph is a subgraph of the complement), and the two node-deletion problems are equivalent (if the input domain of graphs is unrestricted, or at least closed under complementation).

Suppose from now on without loss of generality that all independent sets of nodes satisfy Π; otherwise consider the equivalent problem for $\bar{\Pi}$.

Let G be a graph with connected components $G_1, G_2, ..., G_t$. For each G_i take a curpoint c_i and sort the components of G_i relative to c_i according to their size. This gives a sequence $\alpha_i \triangleq \langle n_{i1}, n_{i2}, ..., n_{ij_i} \rangle$, with $n_{i1} \geqslant \cdots \geqslant n_{ij_i}$, and assume that c_i is the cutpoint of G_i that gives the lexicographically smallest such α_i. (If G_i is biconnected, then c_i is any node of it, and $\alpha_i = \langle n_i \rangle$, where $n_i = | G_i |$.) Sort the sequences of α_i's according to the lexicographic ordering and let $\beta_G = \langle \alpha_1, \alpha_2, ..., \alpha_t \rangle$, where $\alpha_1 \geqslant_L \alpha_2 \geqslant_L a_3 \cdots \geqslant_L \alpha_t$.

The sequences of β_G induce a total ordering R among the graphs. (We may, however, have $\beta_G = \beta_H$ for two nonisomorphic graphs G and H.) Take J to be a least graph in this ordering that cannot be repeated arbitrarily many times without violating Π; i.e., there exists a number $k \geqslant 1$, such that k independent copies of J (without any interconnecting edges) violate Π, k-1 independent copies of J satisfy Π, and any number of independent copies of every H with $\beta_H <_R \beta_J$ satisfies Π. (For example, if $\Pi =$ complete p-partite graph, then J consists of a single edge and $k = 2$.) By nontriviality of Π, there exists such a graph J, and furthermore, since all independent sets of nodes satisfy Π, J has at least one component of size no less than 2.

Let $J_1, ..., J_t$ be the components of J sorted according to their α_i's, c_1 the cutpoint of J_1 that gave α_1, J_0 the largest component of J_1 relative to c_1, J_1' and J' the graphs obtained from J_1 and J, respectively, by deleting all nodes of J_0 except c_1, and d any node of J_0 other than c_1. (There exists such a node d since J_1, and consequently J_0, too, has at least two nodes.)

Now, given a graph G, input to the node cover problem, let G^* consist of nk inde-

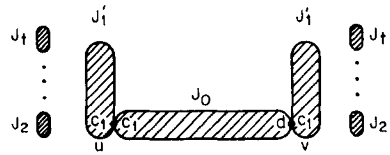


FIGURE 3

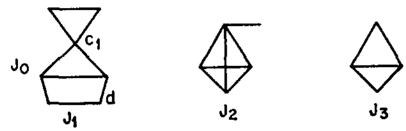


FIG. 4a. A graph J.

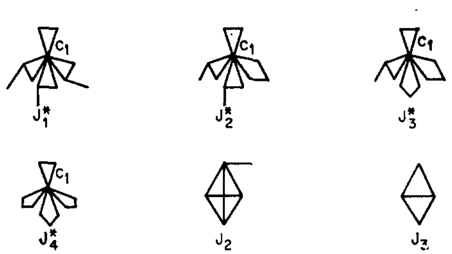


FIG. 4b. The corresponding graph J^*.

pendent copies of G, where n is the size of G. For each node u of G^* create a copy of J' and attach it to u by identifying c_1 with u. Replace every edge (u, v) of G^* by a copy of J_0, attached to u and v by its nodes c_1 and d. (See Fig. 3.) (It does not matter how we identify the nodes c_1, d with the nodes u and v.) Let G' be the resulting graph. We will show that $\alpha_0(G) \leqslant l \Leftrightarrow \gamma(G') \leqslant nkl$.

(1) Let V be a node cover for G, $|V| \leqslant l$. Delete V from each copy of G. Every connected component of the resulting subgraph of G' is either (a) a component J_i of J other than J_1, or (b) a graph formed by taking one copy of J_1' and several copies of J_0, deleting either c_1 or d from each copy of J_0 and attaching it by the other node (d or c_1) to node c_1 of the copy of J_1' (see Fig. 4 for an example), or (c) $J_1 - J_0$ (with c_1 deleted), or (d) $J_0 - c_1$, d (or the connected components of them, in case that the corresponding deletions have disconnected them. However, this does not affect our arguments, since as it will become obvious in a minute, the worst-case is when they are all connected.). Thus the remaining graph can be regarded as a subgraph of repetitions of the following graph J^*: J^* has $t + s - 1$ components, if s is the number of graphs of the form (b) for the possible choices of the node (c_1 or d) deleted from each copy of J_0: these are J_2, $J_3, \ldots, J_t$ and the s graphs J_i^*, $i = 1, \ldots, s$ of the form (b). For example, if J is the graph of Fig. 4a, and the maximum degree of G is 3, then J^* is as shown in Fig. 4b.

For all i, the components of J_i^* relative to c_1 are (a) those of J_1 except J_0 and (b) J_0 with one of the nodes c_1, d deleted. Since each component of the second kind has size less than $|J_0|$, the cutpoint c_1 gives an α-sequence for J_i^* which is lexicographically less than that of J_1, and consequently $\alpha_{J_i^*} <_L \alpha_{J_1}$, for all i.

Therefore $\beta_{J^*} <_R \beta_J$. (In our example $\beta_J = \langle\langle 5, 3\rangle, \langle 4, 2\rangle, \langle 4\rangle\rangle$ and $\beta_{J^*} = \langle\langle 4, 4, 4, 3\rangle, \langle 4, 4, 4, 3\rangle, \langle 4, 4, 4, 3\rangle, \langle 4, 4, 4, 3\rangle, \langle 4, 2\rangle, \langle 4\rangle\rangle$.)

By our choice of J, any number of independent copies of J^* satisfies Π, and by hereditariness the remaining graph does so too. Therefore, $\gamma(G') \leqslant nkl$.

(2) Suppose that $\alpha_0(G) \geqslant l + 1$, and let V be a solution to the node-deletion problem. Let m be the number of copies of G, from which $G' - V$ contains J as an induced subgraph. Since k independent copies of J violate Π and since Π is hereditary on induced subgraphs, $m < k$. That is, from at least $(n - 1)k + 1$ copies of G, $G' - V$ does not contain J as an induced subgraph. Let G_i be such a copy of G and define $V_i' = \{v \in N_i \mid V$ contains a node from the copy of J' attached to v (possibly v itself) or a node from the copy of J_0 that replaced an edge (v, u) with $v < u$ (the ordering of nodes is arbitrary)$\}$. Clearly $|V_i'| \leqslant |V \cap N_i|$. Suppose that there is an edge (v, u) of G_i such that $v, u \notin V_i'$. Then V does not contain any node from the copies of J' attached to v and u, or from the copy of J_0 that replaced (v, u) (since otherwise the smaller of v, u would belong to V_i'). Consequently (see Fig. 3) $G_i' - [V \cap N_i]$ contains J as an induced subgraph (regardless of how the nodes c_1 and d were identified to v and u). Therefore V_i' is a node cover for G. Thus V must contain at least $l + 1$ nodes from each of $(n - 1)k + 1$ copies of G, i.e., $|V| \geqslant [(n - 1)k + 1](l + 1) = nkl + l + 2 + k(n - 1 - l) \Rightarrow \gamma(G') > nkl$, since $n > l + 1$. ∎

COROLLARY 5. *The node-deletion problem restricted to planar graphs for graph-properties that are hereditary on induced subgraphs and nontrivial on planar graphs is* NP-*complete.*

Proof. For every n, there is an $r(n)$ (may take, for example, $r(n) = 4n$), such that all planar graphs with $r(n)$ or more nodes contain an independent set of n nodes. Since Π is nontrivial on planar graphs, all independent sets of nodes satisfy Π. The node-cover problem restricted to planar graphs is NP-complete [5]. Now note that if the original graph G and the graph J defined in the proof of Theorem 4 are planar, and in addition the two attachment points c_1, d of J_0 lie on a common face in an embedding of it on the plane, then the resulting graph G' is also planar. Since Π is nontrivial on planar graphs, we can carry through the proof of Theorem 4 and find such a planar graph J. Moreover we can choose node d to lie on a common face with c_1. ∎

COROLLARY 6. *The node-deletion Π problem for the following properties Π is* NP-*complete*: $\Pi =$ (1) *planar*, (2) *outerplanar*, (3) *line-graph*, (4) *chordal*, (5) *interval*, (6) *without cycles of specified length* l, (7) *without cycles of length* $\leqslant l$, (8) *degree-constrained with maximum degree* $t \geqslant 1$, (9) *acyclic* (*forest*), (10) *bipartite*, (11) *comparability graph*, (12) *complete bipartite.*

Furthermore, the restriction to planar graphs for properties (2)–(12) is also NP-complete.

4. The Node-Deletion Problem for Digraphs

Regarding now digraph properties, note that the first argument used in the proof of Theorem 4 does not hold in the case of digraphs, i.e., it may be the case that neither Π nor $\overline{\Pi}$ is satisfied by an independent set of nodes.

THEOREM 7. *The node-deletion problem for nontrivial digraph-properties that are hereditary on induced subgraphs is* NP-*complete.*

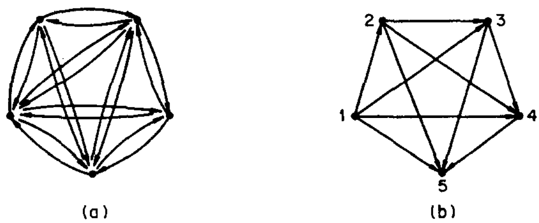
Proof. Recall Ramsey's (generalized) theorem [13]: Let $p_1, \ldots, p_m, t$ be any integers with $t \geqslant 1$; $p_1, \ldots, p_m \geqslant t$, and suppose that for every set S with n elements, the family of all subsets of S containing exactly t elements is partitioned into m disjoint families $f_1, \ldots, f_m$. Then there is a number $r(p_1, p_2, \ldots, p_m, t)$, such that for every set S with $n \geqslant r(p_1, \ldots, p_m, t)$ elements, there exist an i, $1 \leqslant i \leqslant m$, and a subset A_i of S with p_i elements all of whose t-subsets are in the family f_i.

If $D = (N, E)$ is a digraph with $N = \{u_1, u_2, \ldots, u_n\}$ partition the subsets of N with two elements in 4 classes, as follows:

$$\begin{aligned}
f_1 &= \{\{u_i, u_j\} \mid (u_i, u_j), (u_j, u_i) \notin E\},\\
f_2 &= \{\{u_i, u_j\} \mid (u_i, u_j), (u_j, u_i) \in E\},\\
f_3 &= \{\{u_i, u_j\} \mid (u_i, u_j) \in E, (u_j, u_i) \notin E, i < j\},\\
f_4 &= \{\{u_i, u_j\} \mid (u_i, u_j) \notin E, (u_j, u_i) \in E, i < j\}.
\end{aligned}$$

From Ramsey's theorem we conclude, that for every p_1, p_2, p_3, p_4, there is a number $r(p_1, p_2, p_3, p_4)$, such that all digraphs D with no fewer than $r(p_1, p_2, p_3, p_4)$ nodes contain either an independent set of p_1 nodes, or a complete symmetric (henceforth abbreviated as cs.) digraph on p_2 nodes (see Fig. 5a), or a complete antisymmetric transitive (abbreviated as c.a.t.) digraph on $\min(p_3, p_4)$ nodes (i.e., a digraph $D' = (N', E')$ whose nodes can be ordered in such a way that $E' = \{(u, v) \mid u < v;\ u, v \in N'\}$—see Fig. 5b; note also that both families f_3 and f_4 give such a digraph).

Since Π is nontrivial and hereditary on induced subgraphs, it is satisfied either (i) by all independent sets of nodes, or (ii) by all c.s. digraphs, or (iii) by all c.a.t. digraphs. The proof of Theorem 4 works for cases (i) and (ii) (in case (ii) the construction is carried out for $\overline{\Pi}$). It remains, therefore, to show the result for case (iii).



(a) (b)

FIG. 5. Complete symmetric (antisymmetric transitive) digraph of size 5.

Let s be the largest number such that s independent c.a.t. digraphs of any size satisfy Π, i.e., there exist numbers $k_1, k_2, \ldots, k_{s+1}$ such that $s+1$ independent c.a.t. digraphs of size $k_1, \ldots, k_{s+1}$ violate Π. (There exists such a number s if Π is not satisfied by all independent sets of nodes, and $s \geqslant 1$.) Since Π is hereditary on induced subgraphs there exists a number k such that $s+1$ independent c.a.t. digraphs of size k violate Π. (We can take, for example, $k = \max[k_1, k_2, \ldots, k_{s+1}]$.)

Given a graph $G = (N, E)$, input to the node-cover problem, with $N = \{u_1, \ldots, u_n\}$, $E = \{e_1, \ldots, e_m\}$, let $r = m(k-1)n$. Form a digraph $D' = (N', E')$ as follows: $N' = \{u_{ij} \mid 1 \leqslant i \leqslant n, 1 \leqslant j \leqslant r\} \cup \{e_{ijh} \mid 1 \leqslant i \leqslant m, 1 \leqslant j \leqslant r, 1 \leqslant h \leqslant s-1\}$ (if $s = 1$ there are no e_{ijk} nodes) and $E' = \{(u_{ij}, u_{gh}) \mid j < h$, or ($j = h$ and $i < g$); $(u_i, u_g) \notin E\} \cup \{(e_{ijh}, e_{fgh}) \mid j < g$ or ($j = g$ and $i < f$)$\}$.

Note that D' is formed by r copies of G, with every edge $e_f = (u_i, u_j)$ replaced by $s+1$ independent nodes: $\{u_{ig}, u_{jg}\} \cup \{e_{fgh} \mid 1 \leqslant h \leqslant s-1\}$ and the addition of some interconnecting edges.

We claim that $\gamma(D') \leqslant rl \Leftrightarrow \alpha_0(G) \leqslant l$.

(1) Let V be a node cover for G and V' the set of the r copies of V. $D' - V$ consists of s independent complete antisymmetric transitive digraphs. The $s-1$ of them have node set $N_h = \{e_{ijh} \mid 1 \leqslant i \leqslant m, 1 \leqslant j \leqslant r\}$ and the sth has node set $N_s = \{u_{ij} \mid 1 \leqslant i \leqslant n, 1 \leqslant j \leqslant r, u_i \notin V\}$. Since V is a node cover, no two nodes of N_s are copies of adjacent nodes of G and therefore $\langle N_s \rangle$ has the above-stated form. Thus, $\alpha_0(G) \leqslant l \Rightarrow \gamma(D') \leqslant rl$.

(2) Suppose that $\alpha_0(G) \geqslant l+1$, and let V be a solution to the node-deletion problem. For an edge $e_f = (u_i, u_j)$ of G, let $K_f = \{g \mid u_{ig}, u_{jg}, e_{fg1}, \ldots, e_{fg,s-1} \notin V\}$. The nodes that replaced edge e_f in the copies of G with index in K_f, form $s+1$ independent complete antisymmetric transitive digraphs of size $|K_f|$. By our choice of s and k and since Π is herditary on induced subgraphs, we must have $|K_f| \geqslant k-1$, if $D' - V$ is to satisfy Π. Therefore from at least $r - m \cdot (k-1) = (n-1)\,m(k-1)$ copies of G, there is at least one of the $s+1$ nodes that replaced each edge of G deleted. Arguing as in the proof of Theorem 4, V must contain at least $l+1$ nodes from each of these copies, i.e., $|V| \geqslant (n-1)\,m(k-1) \cdot (l+1) = m(k-1)[nl + n - (l+1)] \Rightarrow \gamma(G) > rl$, since $n > l+1$. ∎

The proof of Corollary 5 is valid for digraph properties, too.

COROLLARY 8. *The node-deletion problem restricted to acyclic digraphs for digraph-properties that are hereditary on induced subgraphs and nontrivial on acyclic digraphs is* NP-*complete.*

Proof. Since Π is nontrivial on acyclic digraphs we have to consider only cases (i) and (iii). For case (iii) the digraph D' constructed in the proof of Theorem 7 is clearly acyclic. For case (i), if in the construction of the proof of Theorem 4, J is acyclic and in the substitution of every edge (u, v), with $u < v$, by J_0, c_1 is identified with u and d with v, the digraph G' constructed there is also acyclic. (Recall that as we mentioned there, the way that the nodes c_1 and d of J_0 are identified with the endpoints of the edge is irrevelent.)

Since Π is nontrivial on acyclic digraphs J can be chosen as the R-least acyclic digraph satisfying the conditions stated in the proof of Theorem 4. ∎

COROLLARY 9. *The node-deletion problem for the following digraph-properties Π is* NP-*complete*: Π = (1) *acyclic (feedback-node set)*, (2) *transitive*, (3) *symmetric*, (4) *anti-symmetric*, (5) *line-digraph*, (6) *with maximum outdegree r*, (7) *with maximum indegree r*, (8) *without cycles of length l*, (9) *without cycles of length $\leqslant l$.*

Furthermore, the restriction of all problems to planar digraphs, and the restriction of problems 2, 3, 5, 6, 7 to acyclic digraphs is also NP-complete.

5. CONCLUSIONS

In this paper we saw how a set of similar problems—the node-deletion problems—can be attacked in a systematic way to prove the NP-completeness of all the members of the set. It would be interesting to find other classes of problems for which a similar result holds. In particular it would be nice if the same kind of techniques could be applied to the edge-deletion problems (of course for an appropriately restricted class of properties). Unfortunately we suspect that this is not the case—known reductions for certain properties do not seem to fall into a pattern [15]. A class of problems which seems more likely to be amenable to such a treatment is the class of polynomial and integer divisibility problems [11, 12], where most of the NP-completeness proofs employ similar reductions.

Regarding the class of node-deletion problems, two questions suggest themselves: (1) How much can we expand the class of properties for which the problem remains NP-complete, (2) the reduction schemes we described in Sections 3 and 4 show that the node-deletion problem (without the connectivity requirement) has at least as rich a structure (in the combinatorial sense—see also [1]) as the node cover problem. It is an immediate corollary of the proofs that any ϵ-approximate algorithm for any of the node-deletion problems could be used to derive an ϵ-approximate algorithm for the node-cover problem. What can we say in the other direction, and what are the interrelationships among the various problems in the class with respect to their combinatorial structure? This is very interesting, in view of the fact that there is, for example, no known approximation algorithm with bounded worst-case ratio for the feedback-node set (or any other problem of the class), whereas the node cover problem can be easily approximated within ratio 2, but also because it would shed more light into the nature of NP-complete problems from the combinatorial point of view and into their behavior with respect to approximation algorithms.

Note. The paper reports on work done independently by the two authors. The proof given in Section 2, and various extensions to other hereditary properties, was discovered by the first author [10]. The proof of NP-completeness of the general node-deletion problem for both undirected and directed graphs was found by the second author [15].

ACKNOWLEDGMENTS

We wish to thank J. D. Ullman and David Dobkin for generous encouragement and support.

REFERENCES

1. G. AUSIELLO, A. D'ATRI, AND M. PROTASI, On the structure of combinatorial problems and structure preserving reductions, *in* "Proceedings, Fourth Colloquium on Automata, Languages, and Programming," pp. 45–57, Lecture Notes in Computer Science No. 52, Springer–Verlag, Berlin/New York, 1977.
2. C. BERGE, "Graphs and Hypergraphs," North–Holland, Amsterdam, 1973.
3. S. A. COOK, The complexity of theorem-proving procedures, *in* "Proceedings, Third Annual ACM Symposium on Theory of Computing, 1971," pp. 151–158.
4. M. R. GAREY AND D. S. JOHNSON, "Computers and Intractability: A Guide to the Theory of NP-Completeness," Freeman, San Francisco, 1978.
5. M. R. GAREY, D. S. JOHNSON, AND L. STOCKMEYER, Some simplified NP-complete graph problems, *Theor. Comp. Sci.* 1 (1976), 237–267.
6. F. HARARY, "Graph Theory," Addison–Wesley, Reading, Mass., 1970.
7. J. HOPCROFT AND J. D. ULLMAN, "Formal Languages and Their Relation to Automata," Addison–Wesley, Reading, Mass., 1969.
8. R. M. KARP, *in* "Reducibility among Combinatorial Problems" (R. E. Miller and J. W. Thatcher, Eds.), pp. 85–103, Plenum, New York, 1972.
9. M. S. KIRSHNAMOORTHY AND N. DEO, "Node-Deletion NP-Complete Problems," Report, Computer Science Program, Indian Institute of Technology, 1977.
10. J. M. LEWIS, "On the Complexity of the Maximum Subgraph Problem," Ph. D. thesis, Yale University, 1978.
11. D. PLAISTED, Some polynomial and integer divisibility problems are NP-hard, *in* "Proceedings, 17th Annual IEEE Symposium on Foundations of Computer Science, 1976," pp. 264–267.
12. D. PLAISTED, New NP-hard and NP-complete polynomial and integer divisibility problems, *in* "Proceedings, 18th Annual IEEE Symposium on Foundations of Computer Science, 1977, pp. 241–253.
13. F. P. RAMSEY, On a problem of formal logic, *Proc. London Math. Soc., 2nd Ser.* 30 (1930), 264–286.
14. R. E. TARJAN, Depth-first-search and linear graph algorithms, *SIAM J. Comput.* 1 (1972), 146–160.
15. M. YANNAKAKIS, "Edge-Deletion Problems," TR 249, Computer Science Lab., Princeton University, 1978.