NODE-AND EDGE-DELETION NP-COMPLETE PROBLEMS†

Mihalis Yannakakis

PRINCETON UNIVERSITY
Princeton, New Jersey

ABSTRACT

If π is a graph property, the general node(edge) deletion problem can be stated as follows: Find the minimum number of nodes(edges), whose deletion results in a subgraph satisfying property π. In this paper we show that if π belongs to a rather broad class of properties (the class of properties that are hereditary on induced subgraphs) then the node-deletion problem is NP-complete, and the same is true for several restrictions of it. For the same class of properties, requiring the remaining graph to be connected does not change the NP-complete status of the problem; moreover for a certain subclass, finding any "reasonable" approximation is also NP-complete. Edge-deletion problems seem to be less amenable to such generalizations. We show however that for several common properties (e.g. planar, outer-planar, line-graph, transitive digraph) the edge-deletion problem is NP-complete.

KEYWORDS: approximation, computational complexity, edge-deletion, graph, graph-property, hereditary, maximum subgraph, node-deletion, NP-complete, polynomial hierarchy.

1. INTRODUCTION

The general node(edge) deletion problem can be stated as follows: Given a graph or digraph G find a set of nodes(edges) of minimum cardinality, whose deletion results in a subgraph or subdigraph satisfying the property π. (For the standard graph theory terminology the reader is referred to [H] or [B]).

Several of the well-studied polynomial graph-problems (such as the connectivity of a graph [Ev, T1], the arc-deletion [K], the maximum matching and the b-matching problems [Ed]), as well as NP-complete problems (such as the node cover, the max clique, the feedback-node set, the feedback-arc set [K], and the simple max-cut problem [GJS]) can be formulated in an obvious way as node or edge deletion problems, specifying appropriately the property π. Furthermore, Krishnamoorthy and Deo showed in a recent paper [KD] that the node-deletion problem for several other properties is also NP-complete. (For an exposition of NP-completeness, see [GJ]).

Since Cook's introduction of the concept of NP-completeness, the list of NP-complete problems has expanded rapidly, with more and more individual problems from various areas being added to it [GJ] Sections 2 and 3 are concerned with node-deletion problems. Our aim is to show, how this set of similar problems (with properties π drawn from a certain large class of properties) can be treated in a systematic fashion in order to prove the NP-completeness of all the members of the set. (A similar approach was taken also independently by Lewis, Dobkin and Lipton in [LDL] and [L]. However our results are significantly more general than theirs.). In this paper we consider properties that are <u>hereditary on induced subgraphs</u>, i.e. if G is a graph (or digraph) with property π, then deletion of any node does not produce a graph violating π. We call a property <u>nontrivial</u> if it is true for a single node and is not satisfied by all the graphs in a given input domain. Clearly the node-deletion problem makes sense only for nontrivial properties. We will require π to be easy (i.e. in P) at least to recognize (although our results are valid even if π cannot be recognized in nondeterministic polynomial time. In this case, the 'NP-complete' clause should be replaced by 'NP-hard').

Now suppose that π is such a property and that there is an upper bound k on the order of graphs satisfying π. Then the

†Work partially supported by NSF grant MCS-76-15255.

corresponding node-deletion problem is polynomial in a trivial way: Given a graph G, examine all the (induced) subgraphs of G of order up to k and find the one with the largest order that satisfies π. We call a property interesting (on some input domain) if there exists no such upper bound (for the graphs of the input domain), i.e. if there are arbitrarily large graphs satisfying π.

In Section 2 we show that for any nontrivial and interesting graph - or digraph-property that is hereditary on induced subgraphs, the node-deletion problem is NP-complete. Moreover the restriction to planar graphs (or digraphs) and to acyclic digraphs (in case of digraph-properties) is also NP-complete. For the restriction to bipartite graphs there are few (in a way that is defined more precisely in Section 2) exceptions. For example the node cover problem is polynomial (by Konig's theorem and the fact that the maximum matching is polynomial [Ed]). However we show that it is a unique exception among properties that are determined by the components. (We say that a property π is determined by the components - resp. by the blocks - of a graph, if whenever the components - resp. the blocks - of a graph G satisfy π, then G satisfies also π).

In these constructions the remaining graph after the deletion of a minimum number of nodes is often highly disconnected. One may wish to require the remaining graph to be connected and might even hope that this task could be easier: For example Krishnamoorthy and Deo proved [KD] that the node-deletion problem for π = 'nonseparable' is NP-complete. (If the resulting graph is disconnected they require that each of its components satisfies π). In this case the requirement for connectivity makes the problem very easy (linear): we can find the blocks of the graph [T2] and then determine the block with the maximum number of nodes.

In Section 3 we study the effect of the inclusion of a connectivity requirement. We show that for the same class of properties (hereditary, nontrivial and interesting on connected graphs) the NP-completeness of the node-deletion problem is not affected. (In case of digraphs we take connectivity to mean what is usually called weak connectivity, i.e. connectivity of the underlying undirected graph). Moreover for properties that are determined by the blocks (1) Any nontrivial approximation (with worst-case ratio $O(n^{1-\epsilon})$, for any $\epsilon>0$, with n the number of nodes) of the connected node-deletion problem is also NP-complete*, and (2) Determining whether the largest subgraphs satisfying π are connected or not is in Δ_2^P - [NP $\cup$ co-NP], provided of course that NP $\neq$ co-NP.**

The additional assumption, that π is determined by the blocks, is essential as exemplified by the property π = 'complete graph' (or π = 'star'). This node-deletion problem is equivalent to the node-cover problem, which has a polynomial 2-approximation.

In Section 4 we turn to edge-deletion problems, which tend to be easier to solve (or harder to show NP-complete) than their node-deletion versions. Note for example the difference for π = 'acyclic graph' (tree) or 'degree constrained'. We show the following problems to be NP-complete: (1) without cycles of specified length ℓ or of any length $\leq \ell$, (2) degree-constrained with maximum degree $r \geq 2$ and connected, (3) outerplanar, (4) planar, (5) line-invertible, (6) comparability graph, (7) bipartite (max-cut problem), (8) transitive digraph. For problems (5), (6), (7) we determine the best possible bounds on the node-degrees for which the problems remain NP-complete.

A few words on notation: we use $\gamma_\pi(G)$ (resp. $\gamma_\pi^c(G)$) to denote the minimum number of nodes whose deletion results in a subgraph (resp. connected subgraph) of G that satisfies property π. Usually when no ambiguity can arise, we drop the subscript π. By $\alpha_0(G)$ we denote the node-cover number of G, i.e. $\alpha_0(G) = \gamma_\pi(G)$, with π = ' independent set of nodes'.

2. THE NODE-DELETION PROBLEM FOR HEREDITARY PROPERTIES

Theorem 1. The node-deletion problem for nontrivial, interesting graph-properties that are hereditary on induced subgraphs is NP-complete.

* For definitions concerning approximation algorithms, see [J].

** Regarding the polynomial-time hierarchy and in particular Δ_2^P, see [S].

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 K_m or $\bar{K}_n$. 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 an interesting 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 to 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 $G_1, G_2, \ldots, G_t$. For each G_i take a cutpoint c_i and sort the components of G_i relative to c_i according to their orders. This gives a sequence

$\alpha_i \triangleq < n_{i1}, n_{i2}, \ldots, n_{ij_i} >$, with $n_{i1} \geq \ldots \geq n_{ij_i}$, 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 $\alpha_i = < n_i >$, where $n_i = |G_i|$). Sort the sequences of α_i's according to the lexicographic ordering and let $\beta_G = < \alpha_1, \alpha_2, \ldots, \alpha_t >$, where $\alpha_1 \underset{L}{\geq} \alpha_2 \underset{L}{\geq} \alpha_3 \ldots \underset{L}{\geq} \alpha_t$.

The sequences 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 $k \geq 1$, 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 $\beta_H \underset{R}{<} \beta_J$ satisfies π. (For example if π = complete p-partite graph, then J consists of a single edge and k = 2.) By nontriviallity 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 order no less than 2.

Let $J_1, \ldots, J_t$ 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 2 nodes).

Now, given a graph G, input to the node cover problem, let G* consist of n.k independent copies of G, where n is the order 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. 1). (It does not matter how we identify the nodes c_1,d with the nodes u and v).

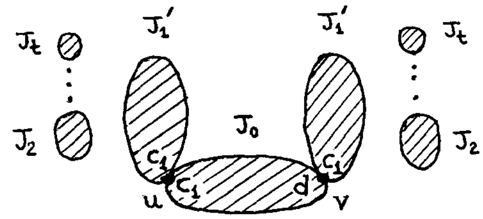
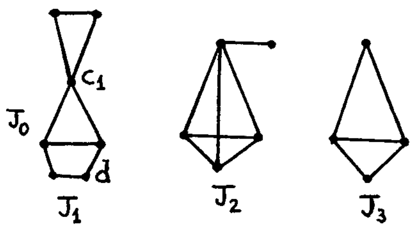


Fig. 1.

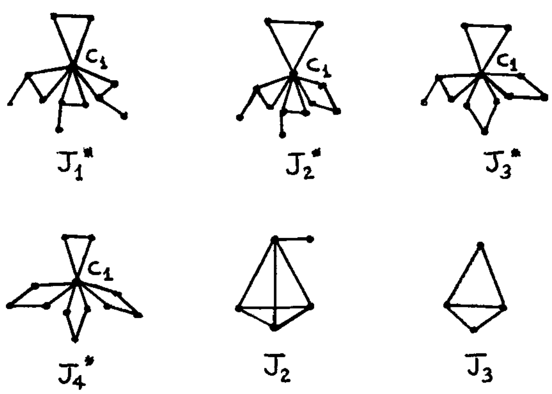
Let G' be the resulting graph.
We will show that $\alpha_o(G) \leq \ell \iff \nu(G') \leq nk\ell$.

1). Let V be a node cover for G, $|V| \leq \ell$. 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. 2 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. 2a, and the maximum degree of G is 3, then J^* is as shown in Fig. 2b.



A graph J
Fig. 2a.

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 order 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 concequently $\alpha_{J_i^*} \lneq_L \alpha_{J_1}$, for all i.



The corresponding graph J^*
Fig. 2b.

Therefore $\beta_{J^*} \lneq_R \beta_J$. (In our example $\beta_J = \langle\langle 5,3\rangle,\langle 4,1\rangle,\langle 4\rangle\rangle$ and $\beta_{J^*} = \langle\langle 4,4,4,3\rangle, \langle 4,4,4,3\rangle, \langle 4,4,4,3\rangle, \langle 4,4,4,3\rangle, \langle 4,1\rangle, \langle 4\rangle\rangle$.)

By our choice of J, any number of independent copies of J^* satisfies π, and by hereditariness the remaining graph does so too. Therefore $\nu(G') \leq n.k.\ell$.

2). Suppose that $\alpha_0(G) \geq \ell+1$, 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 and edge (v,u) with v < u (the ordering of nodes is arbitrary)$\}$. Clearly $|V_i'| \leq |V \cap N_i|$. 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. 1) $G_i' - [V \cap N_i]$ 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 $\ell+1$ nodes from each of $(n-1)k+1$ copies of G, i.e. $|V| \geq [(n-1)k+1](\ell+1) = nk\ell+\ell+2+k(n-1-\ell)$ $\Longrightarrow \nu(G') > nk\ell$, since $n > \ell+1$. □

Corollary 1. The node deletion problem restricted to planar graphs for graph-properties that are hereditary on induced subgraphs, nontrivial and interesting 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 interesting on planar graphs, all independent sets of nodes satisfy π. The node cover problem restricted to planar graphs is NP-complete [GJS]. Now note, that if the original graph G and the graph J defined in the proof of Theorem 1 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 1 and find such a planar graph J. Moreover we can choose node d to lie on a common face with c_1.

□

Theorem 2. The node-deletion problem restricted to bipartite graphs for graph-properties that are hereditary on induced subgraphs, nontrivial on bipartite graphs, and are satisfied by any independent set of edges is NP-complete.

Proof:

There are two cases to be considered.

Case 1. All graphs whose components are stars satisfy π. Since π is nontrivial on bipartite graphs we can carry through the proof of Theorem 1 using as J the R-least bipartite graph satisfying the conditions stated there. Since all graphs whose components are stars satisfy π, J_1 is not a star. (Note that the star S_ℓ has α-sequence $\alpha_{S_\ell} = <1,1,\ldots,1>$, and any connected graph H, with $\alpha_H \underset{L}{\leq} \alpha_{S_\ell}$ is itself a star S_r, with $r \leq \ell$). Thus there is at least one more node in the same set with c_1 in a bipartition of J_0. If we choose as d any such node, the graph G' constructed in the proof of Theorem 1 is bipartite.

Case 2. Some graph all of whose components are stars does not satisfy π. Then J is connected, i.e. has only one component J_1 and this is a star S_ℓ. Since all independent sets of edges satisfy π, $\ell \geq 2$. We distinguish two subcases depending on whether any number of independent copies of the graph shown in Fig. 3 satisfies π or not. For these two subcases we apply two different reductions from the SAT-3 problem. (For a description of the reductions see [Y1]).

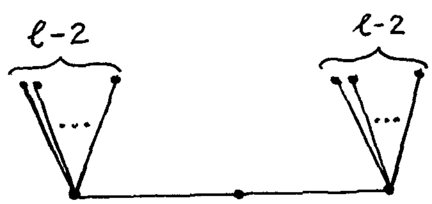


Fig. 3.

Corollary 2. With the exception of the node cover problem, the node-deletion problem restricted to bipartite graphs for graph-properties that are hereditary on induced subgraphs, nontrivial on bipartite graphs and determined by the components is NP-complete.

If some independent set of edges does not satisfy π, there are properties (besides π_0 = 'independent set of nodes') for which the node-deletion problem becomes polynomial when restricted to bipartite graphs, and properties for which it remains NP-complete. For example consider π = 'complete bipartite'. If $G = (N,E)$ is a bipartite graph with $N = S \cup T$ a bipartition of the node set, let $E' = \{(u,v) \mid u \in S, v \in T, (u,v) \notin E\}$. Then it is easy to see that $\gamma_\pi(G) = \min \{\alpha_0(G), \alpha_0(G')\}$. However it can be shown [Y1] that if property π_k has as its only forbidden subgraph k+1 independent edges, with $k \geq 2$, then the corresponding node-deletion problem remains NP-complete even when restricted to bipartite graphs.

Corollary 3 [KD]. The node-deletion π problem for the following properties π is NP-complete: π = 1) planar, 2) outer-planar, 3) line-graph, 4) chordal, 5) interval, 6) without cycles of specified length ℓ, 7) without cycles of length $\leq \ell$, 8) degree-contrained with maximum degree $r \geq 1$, 9) acyclic (forest), 10) bipartite, 11) comparability graph, 12) complete bipartite.

Furthermore the restriction to planar graphs for properties (2)-(12) and to bipartite graphs for properties (1)-(9) is also NP-complete.

Regarding now digraph-properties note that the first argument used in the proof of Theorem 1 does not hold in the case of digraphs, i.e. it may be the case that neither π nor $\bar{\pi}$ is satisfied by an independent set of nodes.

Theorem 3. The node-deletion problem for nontrivial, interesting digraph-properties that are hereditary on induced subgraphs is NP-complete.

Proof:

Using Ramsey's theorem we can show that for all P_1, P_2, P_3 there is a number $r(P_1, P_2, P_3)$ such that all digraphs of order at least $r(P_1, P_2, P_3)$ contain as an induced subgraph either an independent set of P_1 nodes, or a complete symmetric (abbreviated c.s.) digraph on P_2 nodes, or a complete antisymmetric transitive (abbreviated c.a.t.) digraph on P_3 nodes. Since π is interesting 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 1 works for cases (i) and (ii) (in case (ii) the construction is carried out for $\bar{\pi}$). It remains therefore to show the result for case (iii).

Let s be the largest number such that s independent c.a.t. digraphs of any order 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 order $k_1, \ldots, k_{s+1}$ violate π. (There exists such a number s if π is not satisfied by all independent sets of nodes, and $s \geq 1$). Since π is hereditary on induced subgraphs there exists a number k such that s+1 independent c.a.t. digraphs of order 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:
$N' = \{u_{ij} \mid 1 \leq i \leq n,\ 1 \leq j \leq r\} \cup$
$\{e_{ijh} \mid 1 \leq i \leq m,\ 1 \leq j \leq r.\ 1 \leq h \leq s-1\}$
(If s = 1 there are no e_{ijk} nodes) and
$E' = \{(u_{ij}, u_{gh}) \mid j < h$, or ($j = h$ and 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:
$\{u_{ig}, u_{jg}\} \cup \{e_{fgh} \mid 1 \leq h \leq s-1\}$ and the addition of some interconnecting edges.

We claim that $\gamma(D') \leq r.\ell \iff \alpha_0(G) \leq \ell$.

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 anti-symmetric transitive digraphs. The s-1 of them have node set $N_h = \{e_{ijh} \mid 1 \leq i \leq m,\ 1 \leq j \leq r\}$ and the s-th has node set $N_s = \{u_{ij} \mid 1 \leq i \leq n,\ 1 \leq j \leq r,\ u_i \notin V\}$. 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, $\alpha_0(G) \leq \ell \implies \gamma(D') \leq r.\ell$.

2). Suppose that $\alpha_0(G) \geq \ell+1$, 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 order $|K_f|$. By our choice of s and k and since π is hereditary on induced subgraphs, we must have $|K_f| \leq k-1$, if D'-V is to satisfy π. Therefore from at least $r-m.(k-1) = (n-1)m(k-1)$ 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 1 V must contain at least $\ell+1$ nodes from each of these copies, i.e. $|V| \geq (n-1)m(k-1).(\ell+1) = m.(k-1)\ [n.\ell + n-(\ell+1)] \implies$

$$\implies \gamma(G) > r.\ell, \text{ since } n > \ell+1. \quad \square$$

The proof of Corollary 1 is valid for digraph-properties too. Also the result of Theorem 2 can be extended to digraph-properties as well, although there are more subcases to be considered in case 2 according to the orientations of the edges.

Corollary 4. The node-deletion problem restricted to acyclic digraphs for digraph-properties that are hereditary on induced subgraphs, nontrivial and interesting on acyclic digraphs is NP-complete.

Proof:

Since π is interesting on acyclic digraphs we have to consider only cases (i) and (iii). For case (iii) the digraph D' constructed in the proof of Theorem 3 is clearly acyclic. For case (i), if in the construction of the proof of Theorem 1 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 irrelevent.). 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 1. $\square$

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

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.

3. INCLUSION OF A CONNECTIVITY REQUIREMENT.

Theorem 4. The connected node-deletion

problem for graph-properties that are hereditary on induced (connected) graphs, nontrivial and interesting on connected graphs is NP-complete.

Proof:

It is easy to show that for all ℓ,n,m, there is a number $r(\ell,n,m)$ such that all connected graphs of order at least $r(\ell,n,m)$ contain as an induced subgraph either a star S_ℓ or a clique K_n or a path P_m of length m. Since π is interesting on connected graphs and hereditary either all cliques, or all stars or all paths satisfy π.

Case 1. All cliques satisfy π. Then the construction of G' in Theorem 1 is carried out for $\bar{\pi}$. If V* is a node cover for G*, G'-V* is disconnected and consequently $\bar{G}$'-V* is connected.

Case 2. All stars satisfy π. Define a property π' as follows: A (not necessarily connected) graph G satisfies π' iff the graph G_1 formed by taking a new node and connecting it to all nodes of G satisfies π. Clearly π' is nontrivial, interesting, hereditary on induced subgraphs and is satisfied by any independent set of nodes. Apply the construction of Theorem 1 for π'. From the resulting graph G' construct G" by taking a new node v and connecting it with an edge to all nodes of G'. We claim that $\gamma^c_\pi(G") = \gamma_{\pi'}(G')$. Obviously the graph formed by node v and any subgraph of G' satisfying π' is connected and satisfies π, thus $\gamma^c_\pi(G") \le \gamma_{\pi'}(G')$. For the other direction suppose that the optimal solution V to the (connected) node-deletion π problem contains node v. Then V must contain all the nodes from nk-1 copies of G, if G"-V is to be connected. Thus $|V| \ge (nk-1)\, n+1$. Since $\gamma_{\pi'}(G') \le nk\alpha_0(G) \le nk(n-1)$ this is impossible. Therefore V does not contain node v and from the definition of π', $\gamma_{\pi'}(G') \ge \gamma^c_\pi(G")$.

Case 3. Some clique and some star do not satisfy π. Then all paths have to satisy π. In this case we use a reduction from the SAT-3 problem. (see [Y2]). □

Corollary 1 (regarding the planar restriction) is no more true if we include a connectivity requirement. As an example consider π = 'star' (π is clearly hereditary on connected induced subgraphs, nontrivial and interesting on planar connected graphs). Given a planar graph G a maximum (induced) subgraph of G with π consists of a node v and a maximum independent set of its neighborhood $\Gamma(v)$. Since G is planar, $\Gamma(v)$ is outerplanar for each node v. Since the maximum independent set of an outerplanar graph can be found in polymonial time, the same is true for $\gamma^c_\pi(G)$.

Theorem 4 can be extended to digraph-properties as well. To case 1 (all cliques satisfy π) there correspond two cases: (1i) all complete symmetric digraphs satisfy π, and (1ii) all complete antisymmetric transitive digraphs satisfy π. In case (1i) the proof is as in case 1 of Theorem 4. In case (1ii) the proof is as in case 2. To case 2 there correspond three cases according to the orientations of the edges of the stars (**Fig. 4** on last **pg.**) In all 3 cases the proof goes as in case 2 of Theorem 4. Finally if none of the previous cases is true then an infinite number of semipaths satisfies π. (A **semipath** is a digraph whose underlying graph is a path.) In this last case we need to know for each n at least one graph (for example a semipath) of order n satisfying π (or at least be able to generate such a graph in polynomial time). Under this last assumption we have:

Theorem 5. The connected node-deletion problem for digraph-properties that are hereditary on induced (connected) subgraphs, nontrivial and interesting on connected digraphs is NP-complete.

Corollary 6. The connected node-deletion problem restricted to acyclic digraphs for digraph-properties that are hereditary on induced (connected) subgraphs, nontrivial and interesting on connected acyclic digraphs is NP-complete.

From now on we will concentrate on properties that are determined by the blocks of a graph and will not distinguish between graph-and digraph-properties. For such a property π that is hereditary on induced subgraphs, the following are equivalent:

(1) π is interesting on connected graphs
(2) a single edge satisfies π.

Lemma 1 If π is determined by the blocks and is hereditary on induced subgraphs and there exists a forbidden biconnected subgraph H_1 for π with an edge e, whose deletion results in a singly connected graph that satisfies π, then (1) The approximation of the connected node-deletion problem π with worst-case ratio $O(n^{1-\epsilon})$, for any $\epsilon > 0$, with n the number of nodes is NP-complete, (2) It is NP-hard to decide whether all largest induced subgraphs with property π, of a given graph are connected.

Proof:

(1) The reduction is from the SAT-3 (satisfiability with 3 literals per clause) problem. For each clause we form a part of the graph with one node corresponding to each literal, so that for each clause at most one such node can remain if the graph is to satisfy π. Then we take a cutpoint c of H_1-e and connect it to all nodes corresponding to variables through copies of the one component of H_1-e relative to c and to the nodes corresponding to negations of variables through copies of the other component of H_1-e relative to c. We add an edge between any two nodes that correspond to complemented literals. Let G_1 be a graph with m nodes satisfying property π. Attach a copy of G_1 to all nodes of the previous construction that do not correspond to literals. For each clause connect a copy of G_1 to the rest of the graph by the 3 nodes corresponding to the literals of the clause and let G be the resulting graph. If the input set of clauses is satisfiable $\gamma_\pi^c(G) = 2P$ (with P the number of clauses) whereas if it is not satisfiable $\gamma_\pi^c > m$. The result follows by taking m an appropriately high (but constant for fixed ϵ) power of P.

(2) If the set of clauses is satisfiable $\gamma_\pi^c(G) = \gamma_\pi(G) = 2P$ and all subgraphs with property π of G obtained by deleting that few nodes are connected. If the set is not satisfiable $\gamma_\pi(G) < \gamma_\pi^c(G)$. □

If a forbidden subgraph H_1 as above does not exist, then we will have to connect complemented literals by a forbidden subgraph. But then to get connectivity we will have to keep exactly one of these two nodes.

Lemma 2. Given a set of clauses $S = \{C_1,..,C_p\}$ with variables $x_1,...,x_n$ and exactly 3 literals per clause, there is another set of clauses $S' = \{C_1',...,C_{p'}'\}$ with variables $x_1,...,x_n,x_{n+1},...,x_r$ with r and p' linear in p, and at most 3 literals per clause, such that: (1) S is satisfiable iff S' is, (2) Each variable, whose complement appears in S', occurs as many times as its complement, (3) If S is satisfiable, then every satisfying truth assignment for S' satisfies at most 2 literals in each clause. Furthermore there is such a truth assignment that satisfies all noncomplemented variables. (4) If S is not satisfiable, then every truth assignment for S' that satisfies all noncomplemented variables, satisfies 3 literals in some clause.

Lemma 3: If π is nontrivial on connected graphs, determined by the blocks and hereditary on induced subgraphs and there exists a graph H_2 with at least 5 nodes, three of which are distinguished, such that (i) deletion of any distinguished node results in a graph satisfying π, (ii) deletion of at most 2 distinguished nodes does not disconnect the graph, (iii) deletion of all 3 distinguished nodes disconnects the graph, then the approximation of the connected node-deletion problem π with worst-case ratio $O(n^{1-\epsilon})$ is NP-complete.

Theorem 6. If π is nontrivial and interesting on connected graphs, determined by the blocks and hereditary on induced subgraphs, then the approximation of the connected node-deletion problem π with worst-case ratio $O(n^{1-\epsilon})$ is NP-complete.

Proof:

By showing that for each property π either Lemma 1 or Lemma 3 can be applied. □

Lemma 4: If π is determined by the blocks of a graph and hereditary on induced subgraphs and there exists a graph H_3 with at least 4 nodes, three of which are distinguished, satisfying assumptions (i) and (ii) of Lemma 3 and (iii) H_3 does not satisfy π, then it is NP- and co-NP-hard to decide whether all largest induced subgraphs with property π are connected.

Proof:

We use reductions from the SAT-3 problem, where we assume that the input set of clauses consists of a clause containing a single variable x and a set S of clauses that is satisfiable. (Note that Cook's original reduction for the satisfiability problem has exactly this form [C], if we take for example as x the variable that asserts that at the final time the Turing machine is in an accepting state). We apply the transformation of Lemma 2 to S and construct a graph J from the resulting set S' such that all maximum subgraphs with π of J are connected, correspond to a satisfying truth assignment for

S' and keep all nodes corresponding to true literals.

For the NP-hardness proof we form a graph as in Fig. 5 where c is a node con tained in all maximum subgraphs with π of J and the node b is connected to all nodes of the 2 copies of J corresponding to $\bar{x}$ by copies of H_3.

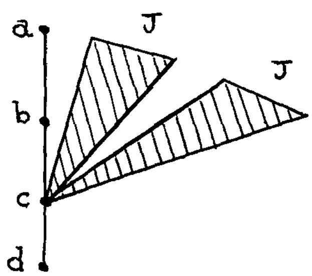


Fig. 5.

For the co-NP-hardness proof we form the graph of Fig. 6, where a, and b are distinguished nodes of H_3, and b is connected to all nodes of J corresponding to $\bar{x}$ by copies of H_3. □

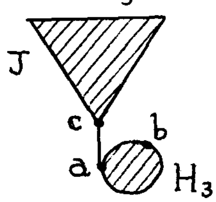


Fig. 6.

Theorem 7. If π is nontrivial and interesting on connected graphs, determined by the blocks and hereditary on induced subgraphs then it is NP-and co-NP hard to decide whether all largest induced subgraphs satisfying property π are connected.

Proof:

We show that Lemma 4 is applicable, unless the triangle does not satisfy property π. In this case the NP-hard part is covered by Lemma 1, and we shall give a separate proof for the co-NP hard part. □

The above problem is easily seen to be in Δ_2^p (the set of languages recognizable in polynomial time deterministically by a query machine with oracle from NP; for the polynomial-time hierarchy see [S]) provided that π is in NP. Thus Theorem 7 shows that NP $\neq$ co-NP $\Longrightarrow \Delta_2^p \supsetneq$ NP $\cup$ co-NP. However this is not a peculiarity of the unrelativised polynomial hierarchy, i.e.:

Proposition Relative to any set X,

$$\Sigma_1^{p,X} \neq \Pi_1^{p,X} \Longleftrightarrow$$

$$\Delta_2^{p,X} \supsetneq \Sigma_1^{p,X} \cup \Pi_1^{p,X}$$

Thus Theorem 7 gives a class of natural graph problems that testify to $\Delta_2^p \supsetneq$ NP $\cup$ co-NP, in case that NP- $\neq$ co-NP. Another such problem is reported in [Le].

Corollary 7: The conclusions of Theorems 6,7 hold for the following node-deletion problems: planar, outerplanar, bipartite, chordal, acyclic graph (tree), cactus, acyclic digraph, symmetric, antisymmetric digraph, without cycles of specified length ℓ, of any length $\leq \ell$.

4. EDGE-DELETION PROBLEMS

Theorem 8. The following edge-deletion problems are NP-complete.

(i) "without cycles of specified length ℓ", for any fixed $\ell \geq 3$,

(ii) for even ℓ, the same problem restricted to bipartite graphs,

(iii) "without any cycles of length $\leq \ell$" restricted to bipartite graphs, for fixed $\ell \geq 4$.

Theorem 9. The edge-deletion "connected, with maximum degreer" problem is NP-complete, for any fixed $r \geq 2$.

Theorem 10. The edge-deletion "outerplanar" problem is NP-complete.

Theorem 11. The edge-deletion "planar" problem is NP-complete.

Proofs:

The reductions for the two last theorems are from the Hamiltonian path problem (with maximum degree 3 [GJS] for the planar case) and are based on counting arguments. We take two copies of the original graph and two new nodes that we connect to all the nodes of the original graphs. We show that if the original graph has a Hamiltonian path then the new graph contains a maximal outerplanar (resp. maximal planar minus one edge) subgraph. Conversely if there is such a subgraph then embedding it on the plane andusing properties of maximal outerplanar (resp. planar) graphs we can exhibit a Hamiltonian path of the original graph. □

Theorem 11 has been independently shown by Geldmacher and Liu.

In the next theorems we use a restricted version of the SAT-3 problem.

Lemma 5: The SAT-3 problem is NP-complete even when each variable appears 3 times. The requirements of the lemma are in a sense the best possible, since if each variable appears at most twice then the clauses are trivially satisfiable (assuming that each clause contains 2 or more literals). Lemma 5 appears to be useful in proving the NP-completeness of restricted problems. For example from it (rather from the transformation used) and Karp's reduction of the SAT-3 problem to the node cover problem [K], follows a result of [GJS]: that the node cover problem on graphs with maximum degree 3 is NP-complete. We use it to determine the best possible bounds on the node degrees for the next three theorems.

Theorem 12: The edge-deletion 'line invertible' graph problem on graphs with maximum degree 4 is NP-complete.

Proof:

Given a set of clauses $C_1,\ldots,C_p$ with variables $x_1,\ldots,x_n$ as in Lemma 5 construct the following graph $G = (V,E)$.

$V = \{a_i, b_i \mid 1 \le i \le n\} \cup \{d_{ij} \mid x_i$ occurs in $C_j\} \cup \{e_{ij} \mid \bar{x}_i$ occurs in $C_j\} \cup$ $\{d'_i,\ e'_i \mid 1 \le i \le n\} \cup \{c_j \mid 1 \le j \le p\} \cup$ $\{c'_j \mid C_j$ has 2 literals$\}$

$E = \{(a_i,b_i) \mid 1 \le i \le n\} \cup \{(a_i,d_{ij}),$ $(a_i,e_{ik}),\ (d'_i,\ d_{ij}),\ (e'_i,\ e_{ik}) \mid$ $d_{ij}, e_{ik} \in V\} \cup \{(d_{ij},\ c_j),\ (e_{ij},c_j),$ $(d_{ij},d_{ie}),\ (e_{ij},\ e_{ie}) \mid j \ne e;\ d_{ij},\ d_{ie},$ $e_{ij},\ e_{ie} \in V\} \cup \{(c_j,\ c'_j) \mid c'_j \in V\}.$

For example if $C_1 = x_1 \vee x_2 \vee \bar{x}_3$, $C_2 = \bar{x}_2 \vee x_3$, $C_3 = \bar{x}_1 \vee x_2 \vee x_3$. the graph G is an in Fig. 7. We show that G has a line-invertible subgraph obtained by deleting r edges, where r is the total number of literal-occurences iff the clauses are satisfiable. In our example the set of edges that are deleted corresponding to the truth assignment $x_1 = 1$, $x_2 = 0$, $x_3 = 1$, is shown with heavy lines in the figure.

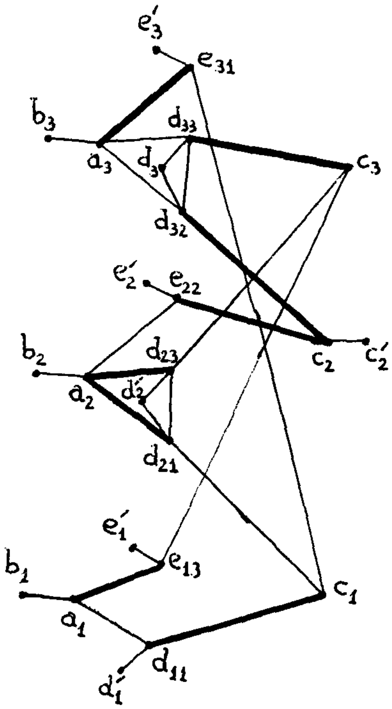


Fig. 7

Remarks:

(1) The restriction of the maximum degree to 4 in the Theorem 12 is best possible, i.e. for graphs with maximum degree 3 the problem can be solved in polynomial time. The algorithm consists in applying successive transformations to the input graph (keeping the maximum degree 3) until a graph without triagles results. Then the problem is reducible to the line-cover problem.

(2) If a connectivity requirement is included, then the best bound can be brought down to 3.

Theorem 13. The edge-deletion "bipartite" (simple max-cut) problem on cubic graphs is NP-complete. The proof uses Lemmas 2 and 5 (rather the transformations employed there). The NP-completeness of the simple max-cut problem (without the restriction on the degrees) was shown in [GJS], where there was also mentioned the open status of the problem on graphs with restricted maximum degree. The NP-completeness of the weighted version (i.e. with the edges having weights) was first shown in Karp's paper [K].

Theorem 14. The edge-deletion 'comparability graph' problem is NP-complete, even on cubic graphs.

Theorem 15. The edge-deletion 'transitive-digraph' problem is NP-complete. For the proof of the two last theorems we first modity the construction of Theorem 13 to get a graph without triangles and then reduce the simple max-cut problem to them.

Finally we note that for all the above properties the edge deletion problem remains NP-complete if we conclude a connectivity requirement.

5. CONCLUSIONS

In Sections 2 and 3 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-the reductions we found for the properties considered in Section 4 do not seem to fall into a pattern. A class of problems which seems more likely to be amenable to such a treatment is the class of polynomial and integer divisibility problems [P1,P2], 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 Section 2 show that the node-deletion problem (without the connectivity requirement) has at least as rich a structure (in the combinatorial sense-see also [ADP]) 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 behaviour with respect to approximation algorithms.

ACKNOWLEDGEMENT I wish to thank J. D. Ullman for his encourangement, support, the patient and careful reading of various forms of the manuscript, as well as many suggestions that improved the presentation of the material.

REFERENCES

[ADP] G. Ausiello, A. D'Atri, and M. Protasi, "On the structure of combinatorial problems and structure preserving reductions", Proc. of Fourth Coll. on Automata, Languages, and Programming, Lecture Notes in Computer Science 52, Springer-Verlag (1977) 45-57.

[B] C. Berge, Graphs and Hupergraphs, North-Holland, Amsterdam, 1973.

[C] S. A. Cook, "The Complexity of theorem-proving procedures", Proc. of Third Annual ACM Symp. on Theory of Computing (1971), 151-158.

[Ev] S. Even, "Maximal flow in a network and the connectivity of a graph," Proc. of Eighth Annual Princeton Conf. on Info. Sci. and Systems (1974), 470-472.

[Ed] J. Edmonds, "Paths. Trees and Flowers", Can. J. Math. 17 (1965), 449-467.

[GJ] M. R. Garey, and D. S. Johnson, Computers and Intractability: A Guide to the Theory of NP-completeness, H. Freeman & Sons, San Francisco, 1978.

[GJS] M. R. Garey, D. S. Johnson, and L Stockmeyer, "Some simplified NP-complete graph problems", Theoretical Comp. Science 1 (1976), 237-267.

[H] F. Harary, Graph Theory, Addison-Wesley, Reading, Mass., 1970.

[J] D. J. Johnson, "Approximation algorithms for combinatorial problems", J. Comp. System Sci. 9(1974), 256-278.

[K] R. M. Karp, "Reducibility among combinatorial problems", in Complexity of Computer Computations, R. E. Miller and J. W. Thatcher, Eds., Plenum Press, NY, 1972, 85-103.

[KD] M. S. Kirshnamoorthy and N. Deo, "Node-deletion NP-complete problems", Computer Science Program, Indian Institute of Technology, report, (1977).

[Le] E.W. Legged Jr., "Tools and techniques for classifying problems", Ph.D. Thesis, Dept. of Comp. and Info. Sciences, Ohio State Univ. (1977).

[L] J. M. Lewis, "On the complexity of the maximum subgraph problem", these Proceedings.

[LDL] J. M. Lewis, D. P. Dobkin, and R. J. Lipton, "Graph properties defined by a forbidden subgraph", Proc. of the 1977 Conference on Info. Sci. and Systems, 108-112.

[P1] D. Plaisted, "Some polynomial and integer divisibility problems are NP-hard", 17th Annual IEEE Symp. on Foundations of Computer Science (1976), 264-267.

[P2] D. Plaisted, "New NP-hard and NP-complete polynomial and integer divisibility problems", 18th Annual IEEE Symp. on Foundations of Computer Science (1977), 241-253.

[S] L. Stockmeyer, "The polynomial time hierarchy", Theoretical Comp. Sci. 3(1977), 1-22.

[T1] R. E. Tarjan, "Testing Graph connectivity", Proc. of the Sixth Annual ACM Symp. on Theory of Computing, (1974), 185-193.

[T2] R. E. Tarjan, "Depth-first-search and linear graph algorithms", SIAM J. on Computing 1(1972), 146-160.

[Y1] M. Yannakakis, "Node-deletion problems on bipartite graphs", Technical report, Computer Science Lab., Princeton University, 1978

[Y2] M. Yannakakis, "The effect of a connectivity requirement on the complexity of maximum subgraph problems", Technical report, Computer Science Lab., Princeton University, 1978.

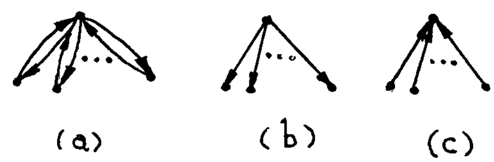


Fig. 4.