

Orienting graphs to optimize reachability¹

S. Louis Hakimi^a, Edward F. Schmeichel^b, Neal E. Young^{c,*}

^a Department of Electrical and Computer Engineering, University of California, Davis, CA 95616, USA

^b Department of Mathematics and Computer Science, San Jose State University, San Jose, CA 95192, USA

^c Department of Computer Science, Dartmouth College, Hanover, NH 03755, USA

Received 10 October 1996; revised 9 June 1997

Communicated by V. Ramachandran

Abstract

It is well known that every 2-edge-connected graph can be oriented so that the resulting digraph is strongly connected. Here we study the problem of orienting a connected graph with cut edges in order to maximize the number of ordered vertex pairs (x, y) such that there is a directed path from x to y . After transforming this problem, we prove a key theorem about the transformed problem that allows us to obtain a quadratic algorithm for the original orientation problem. We also consider how to orient graphs to minimize the number of ordered vertex pairs joined by a directed path. After showing this problem is equivalent to the comparability graph completion problem, we show both problems are NP-hard, and even NP-hard to approximate to within a factor of $1 + \epsilon$, for some $\epsilon > 0$. © 1997 Published by Elsevier Science B.V.

Keywords: Algorithms; Computational complexity; Orientation

1. Introduction

Our terminology and notation is standard except as indicated. We mention only that if X is a subset of the vertices in a graph, we use $\langle X \rangle$ to denote the subgraph induced by X . Good references for any other undefined terms are [1,2].

Let G be any connected graph. Given an orientation \mathcal{G} of G , we will use $R(\mathcal{G})$ to denote the number of ordered vertex pairs (x, y) such that there is a directed path from x to y in \mathcal{G} . We call $R(\mathcal{G})$ the *reachability* of \mathcal{G} .

Robbins [12] showed that G can be oriented so that G is strongly connected (i.e., $R(\mathcal{G}) = |V|(|V| - 1)$)

if and only if G is 2-edge-connected. In the following section (Section 2), we study the problem of how to orient an arbitrary graph G to obtain a digraph \mathcal{G} with $R(\mathcal{G})$ as large as possible. We first transform the problem into an equivalent orientation problem on vertex-weighted trees. Although this transformed version is NP-hard, we prove a key theorem (Theorem 2) which allows us to get a quadratic algorithm for the original problem, as well as a fully polynomial approximation scheme for the transformed problem.

In Section 3, we consider the analogous problem of how to orient G so as to minimize $R(\mathcal{G})$ for the resulting digraph \mathcal{G} . We show this problem is equivalent to comparability graph completion (adding the fewest edges so the resulting graph can be transitively oriented), and then show that both problems are NP-hard, and even NP-hard to approximate within a fac-

* Corresponding author.

¹ This work was supported by the National Science Foundation under Grants NCR-95-05551 and DMS-92-06991.

tor of $1 + \epsilon$, for some $\epsilon > 0$. Related hardness results appear in [6,10].

2. Orienting graphs to maximize reachability

Suppose we are given a connected graph G on n vertices with cut edges. Our goal is to orient G to obtain a digraph G with $R(G)$ as large as possible.

It will be useful to first transform this basic problem. Let C_1, C_2, \dots, C_b denote the components left when the cut edges of G are removed from G . By Robbins' Theorem [12] each C_i can be oriented into a strongly connected digraph. Let us contract each non-trivial C_i into a single vertex x_i , giving x_i weight $wt(x_i) = |V(C_i)|$. The resulting contracted graph is, of course, a b -vertex tree $T = T(G)$ with integer weights on the vertices. Our original problem is now easily seen to be equivalent to the following problem: How should we orient T to maximize $\sum wt(x_i) \cdot wt(x_j)$, the sum being taken over all vertex pairs (x_i, x_j) in T such that there is a directed path from x_i to x_j in T ? Note that the input size of the transformed problem may be exponentially smaller than the input size of the original problem (roughly, $b(1 + \log n)$ versus n).

We begin by showing that this transformed problem is NP-hard. In particular, consider the following decision problem:

Weighted Tree Orientation (WTO).

Instance: Tree T , weight function $wt : V(T) \rightarrow Z^+$, integer $B > 0$.

Question: Is there an orientation T such that $\mu(T) \geq B$, the sum taken over all pairs (v, w) with a directed path from v to w in T ?

Theorem 1. WTO is NP-complete.

Proof. WTO is obviously in NP, and so it suffices to show it is NP-hard. For this we reduce PARTITION [4, p. 233] to WTO. Given positive integers a_1, a_2, \dots, a_m with even sum S , consider the weighted tree T in Fig. 1 and set $B = 5(S/2)^2$. Given any $I \subseteq \{1, 2, \dots, m\}$, consider the orientation T obtained by orienting toward (respectively, away from) the vertex with weight S each edge whose other end vertex has weight a_i for $i \in I$ (respectively, $i \in \{1, 2, \dots, m\} - I = \bar{I}$). We then find

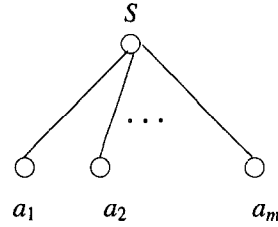


Fig. 1. Tree $T(a_1, \dots, a_m)$.

$$\begin{aligned} \mu(T) &= S \left(\sum_{i \in I} a_i + \sum_{i \in \bar{I}} a_i \right) + \sum_{i \in I} a_i \cdot \sum_{i \in \bar{I}} a_i \\ &\leq S^2 + \left(\frac{S}{2} \right)^2 = 5 \left(\frac{S}{2} \right)^2 \end{aligned}$$

with equality precisely if $\sum_{i \in I} a_i = \sum_{i \in \bar{I}} a_i = S/2$. \square

We now wish to develop a key result which will allow us to obtain a quadratic algorithm for the original orientation problem, as well as a fully polynomial approximation scheme for WTO. First, however, we need some terminology and notation. Let T be any tree with positive integer vertex weights. An *optimal orientation* T is one which maximizes $\mu(T)$. Given an orientation T and a vertex $w \in V(T)$, define

$In_T(w) \doteq \{x \in V(T) \mid \text{there exists a directed path from } x \text{ to } w \text{ in } T\}$,

$Out_T(w) \doteq \{x \in V(T) \mid \text{there exists a directed path from } w \text{ to } x \text{ in } T\}$.

In particular, $w \in In_T(w) \cap Out_T(w)$. Define

$In'_T(w) \doteq In_T(w) - \{w\}$ and

$Out'_T(w) \doteq Out_T(w) - \{w\}$.

We will usually drop the subscript T if it is clear from the context. Finally, given any subgraph with vertex set $X \subseteq V(T)$, let $\|X\| \doteq \sum_{x \in X} wt(x)$ be called the *total weight* of that subgraph. A *centroid* in a vertex weighted tree is any vertex c whose removal minimizes the maximum total weight of any component in $T - c$.

Our goal now is to prove the following.

Theorem 2. *Let c be a centroid of T . In every optimal orientation T , we have $In_T(c) \cup Out_T(c) = V(T)$.*

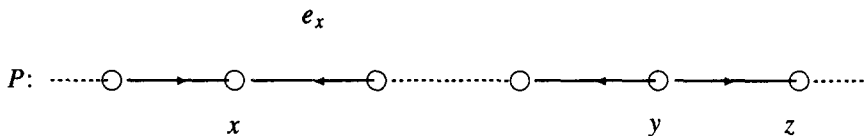


Fig. 2. Double reversal on P .

Before proving this, we require the following result.

Lemma. *Let T be any optimal orientation of T , and let P be any undirected path in T . Then there are no two vertices x, y on P such that the edges of P incident to x (respectively, y) are both directed toward x (respectively, away from y) in T .*

Briefly, there are no “double reversals” on any path in T (see Fig. 2).

Proof. Otherwise, let x, y be a closest such pair on P , so the edges of P between x and y form a directed path from y to x . Let e_x denote the edge at x on the path from y to x . Let $X = \{w \in V(T) \mid \text{there is an undirected path in } T \text{ between } w \text{ and } x \text{ which does not contain } e_x\}$ and let $X' = X \cap \text{In}(x)$.

Let z denote the neighbor of y on P which does not occur on the directed path from y to x (see Fig. 2). Thinking of T as rooted at z , consider the subtree T_y rooted at y . Note that $\|\text{In}'(y)\| \geq \|\text{Out}'(y) \cap T_y\|$ (else we could reverse the orientations of all edges in T_y to obtain an orientation better than T). Reverse all edges in (X') , and consider the gain and loss in $\mu(T)$ in doing so, where the loss in $\mu(T)$ is the sum of the terms in $\mu(T)$ which no longer exist under the new orientation, and the gain in $\mu(T)$ is defined analogously. Since $\text{Out}'(x) \subset \text{Out}'(y) \cap T_y$, we find

$$\begin{aligned} \text{Loss in } \mu(T) &\leq \|X'\| \cdot \|\text{Out}'(x)\| \\ &< \|X'\| \cdot \|\text{Out}'(y) \cap T_y\| \\ &\leq \|X'\| \cdot \|\text{In}'(y)\| \\ &\leq \text{Gain in } \mu(T) \end{aligned}$$

and thus we would have a better than optimal orientation. This proves the lemma. \square

Proof of Theorem 2. Throughout, think of T as rooted at centroid c . If the theorem fails for some optimal orientation T , there must be a ver-

tex $x \notin \text{In}(c) \cup \text{Out}(c)$ adjacent to a vertex $v \in \text{In}'(c) \cup \text{Out}'(c)$. (We will call (v, x) a *dangling edge* at v .) We now consider two cases, assuming $\|\text{Out}(c)\| \geq \|\text{In}(c)\|$ (else reverse the orientation of all edges of T).

Case 1: There is a dangling edge (v, x) at $v \in \text{In}'(c)$.

Note that since $c \in \text{In}(c) - (T_v \cap \text{In}(c))$, we have

$$\|\text{Out}(c)\| \geq \|\text{In}(c)\| > \|T_v \cap \text{In}(c)\|.$$

Reverse (v, x) and the edges in T_x . We find

$$\begin{aligned} \text{Gain in } \mu(T) &\geq \|\text{Out}(c)\| \cdot \|\text{Out}(x)\| \\ &> \|T_v \cap \text{In}(c)\| \cdot \|\text{Out}(x)\| \\ &\geq \text{Loss in } \mu(T) \end{aligned}$$

contradicting the optimality of T .

Case 2: There are no dangling edges at any vertex of $\text{In}'(c)$.

Let (x, v) be a dangling edge at $v \in \text{Out}'(c)$. If $\text{outdeg}(c) \geq 2$, then T contains a path which violates the Lemma (with v, c playing the roles of x, y in the Lemma). Hence we may assume $\text{outdeg}(c) = 1$. But then, since c is a centroid and there are no dangling edges off vertices in $\text{In}'(c)$, and since $\{x\} \cup (T_v \cap \text{Out}(c)) \subset V(T) - \text{In}(c)$, we find

$$\|\text{In}(c)\| \geq \|V(T) - \text{In}(c)\| > \|T_v \cap \text{Out}(c)\|.$$

Reverse (x, v) and the edges in T_x . We find

$$\begin{aligned} \text{Gain in } \mu(T) &\geq \|\text{In}(c)\| \cdot \|\text{In}(x)\| \\ &> \|T_v \cap \text{Out}(c)\| \cdot \|\text{In}(x)\| \\ &\geq \text{Loss in } \mu(T), \end{aligned}$$

contradicting the optimality of T . \square

It is well known that the centroid of a vertex-weighted tree can be found in linear time, and that the centroid consists of either a single vertex or two adjacent vertices [9]. In the latter case, Theorem 2

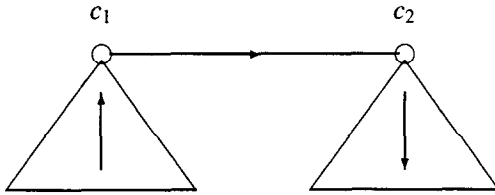


Fig. 3.

implies that the optimal orientation may be represented schematically as in Fig. 3, where c_1 and c_2 denote the adjacent centroids.

In the former case, we have essentially an instance of PARTITION. Let T be rooted at c , and let v_1, \dots, v_k denote the neighbors of c . Consider an optimal orientation of T of T . By Theorem 2, the subtree T_{v_i} must be oriented entirely toward (respectively, away from) c if and only if the edge between v_i and c is oriented toward (respectively, away from) c . Thus,

$$\mu(T) = \|c\| \cdot \|T - c\| + \left(\sum_{i \in I} \|T_{v_i}\| \right) \left(\sum_{i \in \bar{I}} \|T_{v_i}\| \right) + \sum_{i=1}^k \mu(T_{v_i}^*),$$

where $I = \{i \mid T_{v_i} \text{ is oriented toward } c \text{ in } T\}$ and $T_{v_i}^*$ denotes T_{v_i} oriented entirely towards (respectively, away from) c if $i \in I$ (respectively, $i \in \bar{I}$). To maximize $\mu(T)$, we need to find a partition $I \cup \bar{I}$ of $\{1, 2, \dots, k\}$ such that $\sum_{i \in I} \|T_{v_i}\|$ and $\sum_{i \in \bar{I}} \|T_{v_i}\|$ are as equal as possible. There is, of course, a well-known dynamic programming algorithm to find such a partition [4, Section 4.2]. Since $\sum_{i=1}^k \|T_{v_i}\| \leq n$, the running time of this dynamic programming algorithm is $O(kn) = O(n^2)$. Since all other tasks in the original orientation algorithm (e.g., finding the 2-edge-connected components in G , giving these components strongly-connected orientations, etc.) can easily be done in $O(|E|)$ time using standard depth-first search techniques, we see that the original orientation problem can be solved in $O(n^2)$ time; i.e., quadratic in the original input size n . On the other hand, there is also a well-known fully polynomial approximation scheme for the above partition problem [7], which in turn provides a fully polynomial approximation scheme for the NP-complete problem WTO.

3. Orienting graphs to minimize reachability

Let G be a connected graph, and let $r(G) = \min R(G)$, the minimum being taken over all orientations of G . We call an orientation G a *minimal orientation* if $R(G) = r(G)$. We begin with the following result.

Theorem 3. *Every minimal orientation of G is acyclic.*

Proof. Suppose to the contrary there exists a minimal orientation G which is not acyclic. Since G is not acyclic, at least one of the strongly-connected components of G , say C , is not a single vertex. Let E' denote a set of edges in C whose reversal renders $\langle V(C) \rangle$ not strongly connected, and let G' denote G with the edges in E' reversed. It is easy to see every ordered pair of vertices joined by a directed path in G' is joined by a directed path in G , while clearly there is a pair of vertices in $V(C)$ which are joined by a directed path in G but not in G' . Thus, $R(G') < R(G)$, which violates the minimality of G . \square

Obviously $r(G) \geq |E(G)|$, and it is known that $r(G) = |E(G)|$ if and only if G is a comparability graph [1,5], or, equivalently, if G is transitively orientable [3]. It is known that comparability graphs can be recognized in polynomial time [3].

Given a connected graph G , let $\underline{c}(G)$ (respectively, $\bar{c}(G)$) denote the minimum number of edges which must be deleted from (respectively, added to) G to obtain a comparability graph. We now apply Theorem 3 to establish a connection between $r(G)$ and $\bar{c}(G)$.

Theorem 4. $r(G) = |E(G)| + \bar{c}(G)$.

Proof. Suppose we add $\bar{c}(G)$ edges to G to obtain a comparability graph G' , and then orient G' so that $R(G') = |E(G')| = |E(G)| + \bar{c}(G)$. But, of course $r(G) \leq R(G')$, and so $r(G) \leq |E(G)| + \bar{c}(G)$.

On the other hand, consider a minimal orientation G of G , so that $r(G) = R(G)$. By Theorem 3, G is acyclic. Consider the transitive closure $\text{cl}(G)$ of G . We then find

$$\begin{aligned}
r(G) &= R(G) = R(\text{cl}(G)) \\
&= |E(G)| + (\text{no. of edges added to } G \text{ to} \\
&\quad \text{obtain } \text{cl}(G)) \\
&\geq |E(G)| + (\text{min. no. of edges which need to} \\
&\quad \text{be added to } G \text{ to obtain a transitively} \\
&\quad \text{orientable graph}) \\
&= |E(G)| + \bar{c}(G).
\end{aligned}$$

Thus, $r(G) = |E(G)| + \bar{c}(G)$, as asserted. \square

We now turn to the complexity of computing $r(G)$. Consider the following decision problem.

Minimum Reachability Orientation (MRO).

Instance: Graph G , integer $k \geq |E(G)|$.

Question: Is $r(G) \leq k$?

In a moment we will prove

Theorem 5. *MRO is NP-complete.*

It follows immediately from Theorems 4 and 5 that the following problem is also NP-complete.

Comparability Graph Completion (CGC).

Instance: Graph G , integer $k \geq 0$.

Question: Is there a superset E' of E such that $|E' - E| \leq k$ and $G = (\bigvee E')$ is a comparability graph (i.e., is $\bar{c}(G) \leq k$)?

(Previously, it was known only that COMPARABILITY SUBGRAPH (i.e., deciding if $\bar{c}(G) \leq k$) is NP-complete [4, p. 197]. For the optimization versions of MRO and CGC, the goal is to compute (or approximate) the minimum k such that (G, k) is a positive instance of the decision problem.)

Proof of Theorem 5. Clearly $\text{MRO} \in \text{NP}$. To show MRO is NP-hard, we will reduce NOT-ALL-EQUAL 3SAT [4, p. 259].

Let I be an instance of NAE3SAT with m clauses. Construct a graph G_I as follows. Each variable x will be represented by an edge (xT, xF) in G_I , and orienting this edge towards xT (respectively, xF) will correspond to setting x to T (respectively, F). Each clause C will be represented by a 9-cycle in G_I . The three literals in C will be assigned to three equally

spaced edges $(1T, 1F)$, $(2T, 2F)$, $(3T, 3F)$ on C 's 9-cycle, as shown in Fig. 3. If the literal assigned to the edge $(1T, 1F)$ is X (respectively, \bar{X}), add two 2-paths joining $1T$ to xF and $1F$ to xT (respectively, joining $1T$ to xT and $1F$ to xF) to G_I . Make analogous connections for the other two literals of C , as well as for the remaining clauses of I , to complete G_I .

Note that however we orient G_I , each of the m 9-cycles will contain a directed 2-path, and thus $r(G_I) \geq |E(G_I)| + m$. In fact, $r(G_I) = |E(G_I)| + m$ if and only if I has a satisfying truth assignment in the not-all-equal sense. Indeed, given a satisfying truth assignment for I , we can obtain such an orientation for G_I as follows: Orient the edges (xT, xF) to correspond to the truth assignment, and orient all the 6-cycles in G_I containing these edges so that none contains a directed 2-path. Note that the edges in each 9-cycle to which a literal was assigned are now oriented to reflect the truth value of that literal under the truth assignment. Since each clause contains both a true and a false literal, it is trivial to complete the orientation of the 9-cycles so each contains exactly one directed 2-path passing through a darkened vertex on the 9-cycle. Conversely, any orientation of G_I with only $|E(G_I)| + m$ reachable pairs must have exactly one directed 2-path per 9-cycle, and so corresponds to a satisfying truth assignment. \square

We also observe that it remains NP-hard even to approximate $r(G)$ to within a factor of $1 + \epsilon$, for some $\epsilon > 0$. Consider the following optimization problem.

Max Not-All-Equal 3SAT.

Instance: Boolean formula in 3CNF.

Question: What is the maximum number of clauses that can be satisfied (in a not-all-equal sense) by a truth assignment?

It was established recently [8,11] that it is NP-hard to approximate MAX NAE3SAT to within a factor of 1.013 (that is, finding an assignment that satisfies, in the not-all-equal sense, $1/1.013$ of the maximum possible number of clauses is NP-hard). Using this, we can strengthen Theorem 5 as follows:

Theorem 6. *There exists a constant $\epsilon > 0$ such that approximating the optimization versions of MRO or CGC to within a factor of $1 + \epsilon$ is NP-hard.*

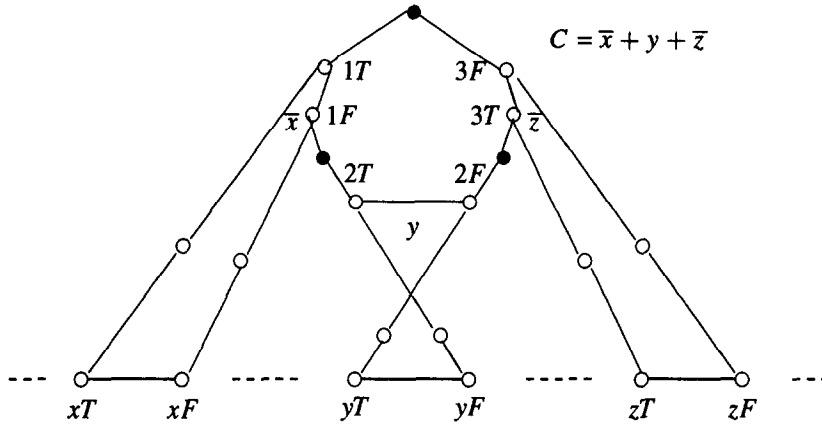


Fig. 4. The graph G_I .

Proof (Sketch). Let I be an instance of NAE3SAT with m clauses. In the following, “satisfying” a clause of I refers to making at least one of its literals true and at least one false. Recall that $R(G_I)$ denotes the number of ordered pairs of vertices (x, y) with a directed path from x to y in an orientation G_I of G_I .

Claim 1: Given any assignment satisfying x of the clauses of I , there is an orientation G_I of G_I such that $R(G_I) = |E(G_I)| + 3m - 2x$.

Claim 1 follows from the proof of Theorem 3.3.

Claim 2: Given any orientation G_I , there is an assignment satisfying at least $\frac{1}{2}(|E(G_I)| + 3m - R(G_I))$ clauses of I .

Claim 2 holds because if any clause-subgraph (see Fig. 4) in G_I has more than one directed 2-path, then it has at least three. Thus, any orientation can be converted into an equally good orientation corresponding to a truth assignment (where the only directed 2-paths are 2-paths through the darkened vertices on the 9-cycles). By simple algebra, the claimed bound holds for this assignment.

Suppose one could find an orientation G_I of G_I with $R(G_I) \leq (1 + \epsilon)r(G_I)$. By Claim 2, this would yield an assignment satisfying at least

$$\frac{1}{2}[|E(G_I)| + 3m - (1 + \epsilon)r(G_I)] \tag{1}$$

clauses of I . Let $\max(I)$ denote the maximum number of clauses in I which can be simultaneously satisfied in a not-all-equal sense. By Claim 1, $r(G_I) \leq |E(G_I)| + 3m - 2\max(I)$. Thus expression (1) is

at least $\frac{1}{2}[|E(G_I)| + 3m - (1 + \epsilon)(|E(G_I)| + 3m - 2\max(I))]$, which equals

$$\max(I) - \frac{1}{2}\epsilon[|E(G_I)| + 3m - 2\max(I)]. \tag{2}$$

Using $|E(G_I)| \leq 24m$ and $m \leq 2\max(I)$ (assuming without loss of generality that every clause has at least two literals, a random assignment satisfies at least half the clauses on average), expression (2) is at least $\max(I) - \frac{1}{2}\epsilon[48\max(I) + 6\max(I) - 2\max(I)] = (1 - 26\epsilon)\max(I)$. From this the claimed hardness of approximating MRO follows, with $\epsilon > 0.00049$.

A similar argument, using Theorem 4 and omitting the $|E(G_I)|$ term in expression (1) (and subsequent expressions), establishes the claimed hardness of approximating CGC, with $\epsilon > 0.0064$. \square

4. Concluding remarks

It would be interesting to determine the algorithmic complexity of completing the orientation of a partially oriented graph to maximize reachability. Thus far we have made little progress on this problem. The analogous completion problem to minimize reachability is, of course, NP-hard.

On the other hand, it is very easy to characterize partial orientations which can be completed into strong orientations. An edge-cut (X, \bar{X}) in a partially oriented graph is called *one-way* if all the edges in the cut are

already oriented, and all are oriented from X to \bar{X} or all from \bar{X} to X . We have

Theorem 7. *A partial orientation of a graph G can be completed into a strong orientation if and only if G is 2-edge-connected and there are no one-way edge cuts in the partial orientation.*

References

- [1] C. Berge, *Graphs and Hypergraphs* (North-Holland, Amsterdam, 1973).
- [2] J.A. Bondy and U.S.R. Murty, *Graph Theory with Applications* (North-Holland, Amsterdam, 1976).
- [3] S. Even, A. Pnueli and A. Lempel, Permutation graphs and transitive graphs, *J. ACM* 19 (1972) 400–410.
- [4] M.R. Garey and D.S. Johnson, *Computers and Intractability: A Guide to NP-Completeness* (Freeman, San Francisco, 1979).
- [5] P.C. Gilmore and A.J. Hoffman, A characterization of comparability graphs and of interval graphs, *Canad. J. Math.* 16 (1964) 539–548.
- [6] M.C. Golumbic, H. Kaplan and R. Shamir, Graph sandwich problems, *J. Algorithms* 19 (1995) 449–473.
- [7] O.H. Ibarra and C.E. Kim, Fast approximation algorithms for the knapsack and sum of subset problems, *J. ACM* 22 (1975) 463–468.
- [8] V. Kann, J. Lagergren and A. Panconesi, Approximability of maximum splitting of k -sets and some other APX-complete problems, Tech. Rept. TRITA-NA-P9509, Dept. of Numerical Analysis and Computer Science, Royal Institute of Technology, Stockholm.
- [9] O. Kariv and S.L. Hakimi, An algorithmic approach to network location problems, II: The p -medians, *SIAM J. Appl. Math.* 37 (1979) 539–560.
- [10] P. Dell’Olmo, M. Grazia Speranza and Z. Tuza, Comparability graph augmentation for some multiprocessor scheduling problems, *Discrete Appl. Math.* 72 (1997) 71–84.
- [11] C.H. Papadimitriou and M. Yannakakis, Optimization, approximation and complexity classes, *J. Comput. Systems Sci.* 43 (1991) 425–440.
- [12] H.E. Robbins, A theorem on graphs with an application to a problem of traffic control, *Amer. Math. Monthly* 46 (1939) 281–283.